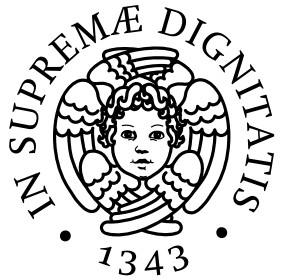

UNIVERSITÀ DI PISA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

CORSO DI LAUREA MAGISTRALE IN
CONSERVAZIONE ED EVOLUZIONE

TESI DI LAUREA MAGISTRALE

Multi-marker molecular characterization of ciliates and their endosymbionts

CANDIDATO

Vittorio Boscaro

RELATORE

Dott. Giulio Petroni

CORRELATORI

Prof. Alberto Castelli
Prof. Sergei Fokin

Anno Accademico 2010/2011

1. Introduction

This Master Thesis is the product of a yearlong period of work in the laboratory of Protistology/Zoology of the University of Pisa, Department of Biology, under the supervision of dr. Giulio Petroni. It is actually a composition of three different subprojects that were developed during this internship, united by the type of biological system studied and the approach employed. This introduction will serve the purpose of presenting both.

The study of ciliated protozoa and their bacterial symbionts is one of the main research lines in the lab, introduced in the '70s by dr. Giovanna Rosati and Prof. Franco Verni and still carried on, along with other research topics, by dr. Giulio Petroni, Prof. Sergei Fokin, dr. Claudia Vannini and dr. Letizia Modeo. Handling and studying ciliates was the main feature of my daily activities; a brief introduction to ciliate biology and their symbiotic relationships with bacteria is provided in section 1.1.

As for methods, I focused on the molecular characterization using multiple markers (when possible) both for the ciliate hosts and the bacterial symbionts. The methodology aspects common to all projects are described in section 1.2. Section 1.3 contains an overview of the systems studied, which may be thought of as different applications of the same approach.

Chapter 2: Methods details the procedures and the kinds of analysis performed. It gives just a summarized survey of the standard molecular characterization and fluorescence *in situ* hybridization (FISH) techniques, which are fairly common, and is more focused on the phylogenetic analysis methods.

Chapters 3, 4 and 5 are dedicated to the results obtained for each separate project and their discussion. Chapter 6 considers again the work as a whole, summarizing the conclusions and the further projects that can be planned as a continuation of mine. It also compares the three applications in order to emphasize the flexibility of the approach used.

This thesis ends with the acknowledgements, an appendix detailing the exact protocols employed in a schematic fashion and some tables, and a list of all the research papers and books cited in the other sections.

1.1. Ciliates and their endosymbionts

1.1.1. Some words on symbiosis

The study of symbiotic relationships is one of the promising fields in the post-Synthesis era of evolutionary biology (e.g. Margulis, 1981; Sapp, 2004). Symbiotic processes are ubiquitous and extremely varied, and are very suited for a multidisciplinary approach. Ciliated protozoa (phylum Ciliophora) are good models for the study of different kinds of symbioses (e.g. Görtz, 1996); probably the most common is the harbouring of bacterial populations or communities inside one of the protozoan's subcellular compartments (e.g. Görtz, 2006).

The meaning of "symbiosis" changed through time; in what follows, I will always refer to the less onerous definition of "the living together of unlike organisms" (DeBary, 1879). It is however important to keep in mind also more modern definitions (e.g. ISS, 2003), which put the emphasis on the evolutionary novelties that the symbiotic process promotes in the organisms involved. Because the kind of symbiosis I'm referring to concerns ciliated protozoa and bacteria living inside them, I will use the terms "host" and "(endo)symbiont" for the eukaryotic and prokaryotic organism respectively. All of these terms are neutral with respect to ecological concepts like "parasitism" or "mutualism".

Empirically, ciliatologists usually employ the term "symbiont" for any organism that lives, survives and reproduces inside one of the cellular compartment of the ciliate.

It is worth spending some words on the value and importance of symbiosis in evolutionary biology. First of all, symbiotic processes (in the meaning of close spatial and prolonged temporal relationships between members of different species) are literally everywhere. Virtually all metazoans harbour – and require for surviving – a disparate community of

microbial organisms; many plant taxa require fungal mycorrhizae in order to obtain nutrients; nearly every multicellular organisms is the host of at least one parasite; and so on… Thus, symbiosis is not only a widespread phenomenon, but also a heterogeneous one, supplying a conceptual unity to apparently different phenomena like mutualism and phatogenesis (e.g. Ewald, 1987; Hentschel et al., 2000).

As defined above, symbiotic relationships are bound to boost known evolutionary processes, like coevolution of different species. But there is more: in many cases, symbiosis brings to rapid establishment of evolutionary novelties (e.g. Werren et al., 2008). This happens because different genomes, each one a product of a different evolutionary history, comes into contact and may act in concert. The two organisms can indeed end up being inseparable (we speak then of obligate symbionts), unable to survive if detached. In this case, they actually become a single unit of selection. In one of the most famous cases, it happened to eukaryotic cells and their bacterial endosymbionts that eventually evolved into mitochondria.

Summarizing, it is true that the impact of symbiosis on evolution is not detached from classical processes like natural selection and genetic drift (indeed, they are always needed). But its still underestimated importance lies in the amount of variability that it can offer to those mechanisms. Major evolutionary jumps couldn't have been reached without symbiosis, simply because they couldn't be obtained by the slow accumulation of mutations in a single, constrained genome.

1.1.2. General features and systematics of ciliates

Ciliates are unicellular or colonial eukariotic organisms generally labelled as "protists" or "protozoa" – very general and quite uninformative terms nowadays. They are known since a very long time (indeed, since the first microscope was invented; Leeuwenhoeck, 1674) because they are among the largest of microbes, ubiquitous and fairly common wherever liquid water is available. The phylum is, according to all modern authors, monophyletic. Three apomorphies unites it: the presence of cilia in at least one stage of life cycle, nuclear dimorphism (they always possess one or more macronuclei constitutively active for vegetative functions and one or more micronuclei specialized for sexual processes) and conjugation (a

unique type of sexual recombination that is known for at least some species in nearly all classes). Phylogenetically, they are related to two other well-known phyla: the common flagellate Dinoflagellata and the parasitic Apicomplexa. The three of them form a monophyletic clade called Alveolata.

Ciliate biology is pretty variable and interesting from many points of view. Molecular biologists use some taxa – like *Paramecium* and *Tetrahymena* – as model organisms. Ecologists, cytologists and evolutionary biologists are just some of the major categories of researcher very actively studying this group of organisms.

I will not even attempt to summarize ciliate biology, referring to the appropriate textbooks for most of this topic (especially Lynn, 2008 and Hausmann & Bradbury 1996). In order to understand my work, it will be more than enough to overview three aspects: ciliate general systematics, ecological niche and symbiotic relationships.

The phylum contains about 3000 known species, but estimates on the total undiscovered biodiversity are very different and objects of heated debates (e.g. Finlay & Fenchel, 1999; Foissner, 1999). There is more than one classification scheme, but aside from names they generally agree on a broader scale. I will refer to that of Denis Lynn monography (Lynn, 2008).

Phylum Ciliophora is subdivided into two subphyla, whose monophyly has been extensively proven: Postciliodesmatophora and Intramacronucleata. Postciliodesmatophora comprises two classes united by the presence of a peculiar ciliary-associated structure, the postciliodesma, resulting from the overlapping of many microtubular ribbons called postciliary microtubules (and typical of all ciliates' cilia).

Postciliodesmatophoreans have no microtubular structures inside their macronuclei aiding the division process; this is an apomorphy of the other subphylum, Intramacronucleata, which is much more diverse and comprises 9 ciliate classes.

The organisms actually studied during my thesis belong to the genera *Paramecium* and *Condylostoma* (Fig. 1.1), so I will give more details only on the phylogenetic positions of these organisms.

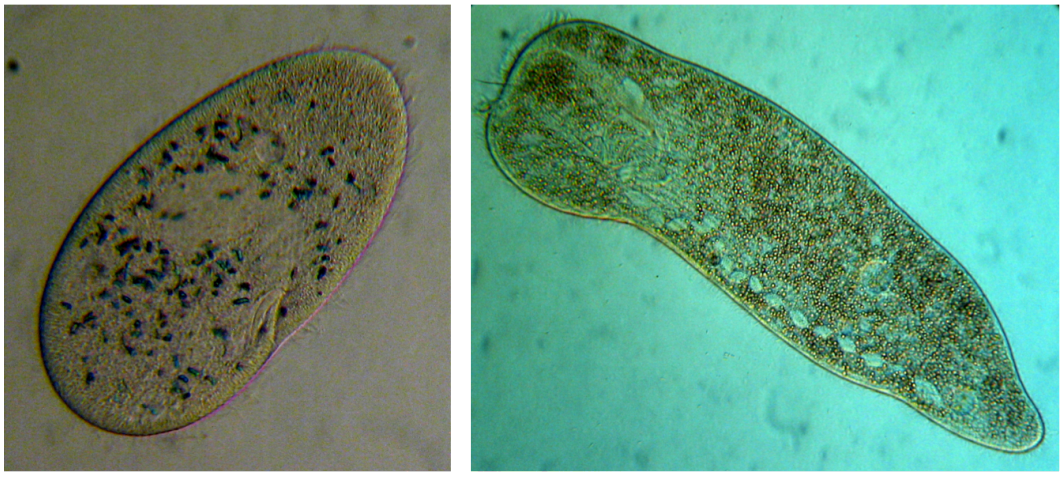

Fig 1.1. Single cells from *Paramecium duboscqui* strain Tub2 (left, photograph by the author) and *Condylostoma* sp. strain COL2 (right, photograph by Gabriele Tomei)

Paramecium O.F. Müller, 1773 is probably the most famous of all ciliate genera, often cited even in high school texts and a commonly used model in molecular biology labs. It is a fresh- or brackish water free-living ciliate of medium size (100-150 μm), generally bacterivorous, which under laboratory conditions usually divides at a fast rate. The genus comprises about 20 species (Fokin et al., 2004), just a few of which are extensively studied. *P. aurelia* is an old and classical example of a complex of sibling (biological) species. The single morphospecies is actually subdivided in 15 subgroups sexually isolated from one another (syngens; Sonneborn, 1975; Aufderheide, 1983).

Paramecium belongs to family Parameciidae, order Peniculida, subclass Peniculia and class Oligohymenophorea. The oligohymenophoreans form the most specious, various and derived class of living ciliates. They are united by the oligohymenium, a specialised form of oral apparatus comprising three adoral kineties (row of cilia functionally coupled on the left side of the oral cavity) and one paroral (right-side) kinety. To this class, but in different subclasses, belong also the other most exploited ciliate model, *Tetrahymena*, and the peritrich *Vorticella*, probably the first ciliate discovered. Oligohymenophorea belongs to subphylum Intramacronucleata.

Condylostoma Bory de St. Vincent, 1824 is a much less glorified ciliate, but is actually a common sight in marine- and brackish water interstitial community. A large ciliate (averaging 200-800 μm) easy to

identify thanks to its distinctive oral apparatus, it can often be cultured in laboratory for long periods. The number of different (morpho)species inside this genus is hard to evaluate, because many are ill-defined or poorly described.

Condylostoma belongs to family Condylostomatidae, order Heterotrichida and class Heterotrichea. The heterotrichs are a small but distinctive class of medium to very large ciliate from fresh- and saltwater. They belong to Postciliodesmatophora, so *Condylostoma* is as distantly related to *Paramecium* as a ciliate can be. As postciliodesmatophors, heterotrichs have no microtubular structure inside the big, often moniliform macronuclei, but have independently developed external macrotubular structures that aid the nuclear division (which is not a mitosis nor a meiosis). This differentiates them from their sister-group class, the Karyorelictea, which are totally incapable of dividing their macronuclei (they need to form it anew from the micronuclei after every cellular division). Molecular phylogenetic analyses have been published for this class, always showing the same robust clades as well as other much less resolved relationships. Because an attempt in this way was performed also in this work, I will detail this issue in the appropriate section.

1.1.3. Ecology of ciliates

Ciliates are as variable in their ecology as they are in their morphology, which is quite complex for unicellular organisms like them. They live nearly everywhere there is water, from the pole to the equator, from hypersaline to freshwater ponds, from running rivers to the film of dew on mosses and bromeliads. Of course, this variability is not shared by all ciliate taxa. Most of them have their preferential niche and habitat. Among the principal different categories of free-living ciliate habitats there are the column of water, the coarse sediments on the bottom of water bodies and the rhizosphere connected to aquatic plants. Most ciliates are free-living, but there are many exceptions (see below).

One of the few unifying feature of their life-style is that all Ciliophora are heterotrophs, without known traces of plastid relics (although some argue that their ancestors should have had one). The only known exceptions are those species that harbour photosynthetic algae (see below).

Ciliates are generally phagotroph, and this is probably their ecological feature most strictly related with the establishment of symbiotic relationships with bacteria. Many of them are indeed bacterivorous and ingest bacteria in a specialized portion of the **cortex**, the outer layer of the cell. This area is called the cytostome, and is surrounded by a very variable oral area. Small preys are captured through active filtration of vortexes generated by the oral ciliature, fagocytosed and enveloped in a phagosome. The digestion process then starts with the acidification of the phagosome medium and the activation of digestive enzymes released through the fusion with lisosome-like vesicles.

Thus, bacteria always have a way to get inside the ciliate cells. Even when the elected prey is a eukaryote (typically algae, flagellates or other ciliates), during the formation of the phagosome it is quite common that free-living bacteria from the environment got included in the digestive vacuole. The digestive process of course destroys most of them. But some, as detailed below, sometimes manage to escape and colonize some compartment of the ciliate.

1.1.4. Bacterial endosymbionts of ciliates

As a preliminary remark, it is important to say that the field of ciliate symbiosis is potentially and actually wider than the one hinted here, which focuses on the ciliate as a host and on bacteria as endosymbionts. Ciliate can be symbionts themselves (recall that I use the term "host" and "symbiont" for the bigger and smaller organisms respectively). There are many known cases for arthropods and vertebrates, and the relationships can go from a commensal to a mutualistic or parasitic one (e.g. Görtz, 1996). One common situation is the finding of anaerobic ciliate inside the gut of large metazoans (like the rumen of cattle or the hindgut of insects), where they feed on the local bacterial flora. There is just one known human pathogen (*Balantidium coli*), but in recent years a fish parasite with the incredibly cacophonous name *Ichthyophthyrius* has emerged as an economic important plague of farmed fishes.

Moreover, there are some very well studied examples of eukaryotic endosymbionts of ciliates, like yeasts, algae (e.g. the classical *Paramecium bursaria – Chlorella* symbiosis; Kodama & Fujishima, 2009) or even other parasitic ciliates (e.g. those of the subclass Suctoria). Another category is

that of endosymbiotic archaea, like the methanogens associated with many anaerobic species. And finally, there are also intriguing cases of bacterial ectosymbionts (e.g. Rosati, 1999), which I cannot cover here for reasons of space and coherence.

Bacterial endosymbionts constitutes however a large part of the researches, are probably the most common type of symbionts of ciliates, and the scope of my projects. From this point on, I will refer only to this kind of relationship.

From a theoretical point of view, there are many reasons to explain the establishment and advantages of a symbiotic process between ciliates and bacteria. One was explained in the previous section: ciliates are phagotrophs that almost always ingest bacteria, opening a path for the infection. Another one is that ciliates are both quite large (from a bacterial point of view) and internally complex, with many different compartments that can be targeted by bacterial invasion. As in all symbiotic processes, there are then complex ecological and evolutionary forces that drive the development of the relationship as a mutualism or parasitism or else.

When we actually look at ciliates collected in nature, we can see that the presence of bacterial symbionts is indeed the rule (e.g. Vannini et al., 2003). But the history of the study on ciliate symbionts was long and a bit difficult to reconstruct, mainly because it was carried on by groups and schools with different backgrounds and purposes, in different times and places.

In extreme synthesis, and only for the scope of this introduction, we can subdivide in two groups the papers published on this field. I will call the first one the "classic" set. This comprises the "famous" symbionts chiefly found in the *Paramecium* genus. Here we found the two most studied bacterial symbionts of ciliates, belonging to the genera *Caedibacter* and *Holospora*.

Holospora is the main focus of the first project, so I will refer to chapter 4 for the details on this intriguing organism, whose principal feature is the adaptation to horizontal transmission between hosts. *Caedibacter* bacteria are perhaps even more fascinating, but I didn't study them during this thesis, so I must limit myself to few sentences (see for a review Schrallhammer & Schweikert, 2009). They are responsible for the so-called "killer effect" of some *Paramecium* strains. "Sensitive" strains of

Paramecium put in the same medium of "killer" strains die within hours. Agent of the death is the release from the *Caedibacter*-bearing "killers" of a peculiar form of the bacterium, called "bright". Brights are ingested by "sensitive" paramecia and in the acidic environment of the digestive vacuole unroll a peculiar structure, the refractile body (R-body), and probably release one or more toxins thus killing the sensitives. According to phylogenetic analyses performed on the SSU rRNA gene sequence, *Caedibacter* is a polyphyletic genus that contains organisms united by the common feature of inducing the killer trait.

Less studied, but still included in the "classic" set (Fig. 1.2), are bacteria from genus *Polynucleobacter*, obligate symbionts of a clade of species of the *Euplotes* genus (e.g. Heckmann & Schmidt, 1987; Vannini et al., 2007).

The infectious *Holospora*, killer *Caedibacter* and mutualistic *Polynucleobacter* were and still are studied by the ciliatology community, especially by those researcher specialized in *Paramecium*. There are many publications regarding various topics of their biology, they are a common presence in protists- and ciliates-focused congresses, and they were known from many decades.

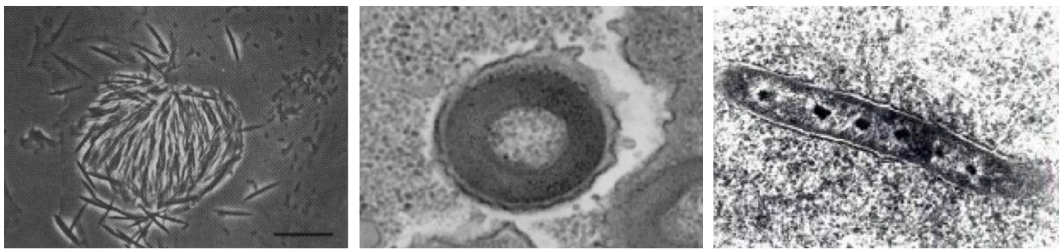

Fig 1.2. From left to right: *Holospora elegans* from the micronucleus of *Paramecium caudatum* hybridized with a specific tetramethylrhodamine-labelled probe (Amann et al., 1991); ultrastructure of *Caedibacter taeniospiralis* from *P. tetraurelia* harbouring R-bodies (Beier et al., 2002); ultrastructure of *Polynucleobacter necessarius* inside *Euplotes aediculatus* (Heckmann, 1975)

There is then a second set of studies, much more scattered and less glorified, that concerns a vast number of poorly described and rarely found bacteria discovered inside a great variety of ciliates. For many, we have just some pictures at the optical or electron microscope. Most lack any

kind of molecular characterization. For all of them, we knew practically nothing of their biology.

And still, at least we know that they are there. The information we have allow us to suspect that there are many more things to discover in this still unexplored world. Few works cover this issue in its entirety, looking at the whole picture. I will try to review something about the taxonomy of these bacteria in chapter 4.

Just in order to illustrate the potentiality of this field, I will cite as an example the so-called *Rickettsia*-like organisms (RLO) previously discovered in our lab. *Rickettsia* belongs to order *Rickettsiales* (*Alphaproteobacteria*) and comprises known human pathogens transmitted by arthropods. In recent years, several bacteria belonging to genus *Rickettsia* or closely related to it (hence, RLO) were found inside the cytoplasm of different ciliate species (e.g. Vannini et al., 2005; Ferrantini et al., 2009). We still know very little about their biology, and especially about their pathogenicity (if any). But the thought of water-bound bacteria strictly related to pathogens that are known only for terrestrial environments is intriguing. Only further thorough studies can tell if these very organisms can constitute an unknown threat to humans or other vertebrates. This is just one example of the possibilities opened by looking at the organisms living inside ciliates. From a more ecological point of view, there is still the unresolved question of the role they play in the host's biology. Because very few research groups work on this field, and the proportion of ciliates studied is tiny, many more surprises can be expected in the future.

1.2. Molecular markers

1.2.1. The multi-marker characterization

Before briefly reviewing the actual molecular approach I used, it is necessary to make a brief introduction on why to use molecular techniques at all in a characterization process.

First of all, for many organisms the classical, morphological data are simply too few to be useful. This is particularly true for bacteria and many other microorganisms, whose morphology is too simple (and sometimes too variable) to be informative enough to discriminate between taxa. This is also true for many protists. Most ciliates are indeed the exception, because their morphology is very distinctive and features-rich. Nevertheless, it was repeatedly proven that, especially in some order of ciliates, even the most careful morphometric measures are not able to separate different species, or are at odds with other data (e.g. Strüder-Kypke & Lynn, 2003).

Even when we have abundant morphological characters, molecular data supply a different and largely independent set of information. Hypotheses supported by both kinds of data are among the most robust in the field of systematics, and they are not rare. But in other circumstances molecular analyses overturned previous, classical schemes, often based on subjective and arbitrary premises. While morphological characters themselves can be sound and unambiguously measured, the choice of the ones to be relied upon for phylogenetic inferences and evolutionary considerations can be difficult.

Molecular data are generally easier to define and circumscribe. As said below, they can be very different from each other. Finally, methodologies of phylogenetic analysis that employ them are less flawed (but far from perfect, see section 2.4) than those used for the analysis of morphological characters. As a conclusive note, it is however fair to say that both methods have their critics, and that no one single approach can be "the correct one" on its own. Multidisciplinary approaches are always the preferable choice, when there is availability of time, money and expertise.

Molecular markers can be of many types. Their common feature is that they are related to the genetic material of the organism involved – its DNA. They can be microsatellite, SNP, measure of whole genomic similarities, codon usage, etc.

In this work, I always employed as molecular markers the nucleotidic base sequences of fragments of DNA – genes, gene portions or other transcribed sequences. These sequences are the results of the process of sequencing, that itself requires a previous work of selection and "amplification" (increase of copy number) of the sequences of interest. The most common method to obtain them, as briefly reviewed in section 2.2, is

the PCR, an easy and quick procedure with one constraint: at least small portions at both ends of the targeted sequence must be already known.

Nowadays, molecular sequences are among the most exploited markers. They can be a bit more tricky and expensive to obtain than other characters, but are usually quite information-rich and flexible. They can be used for many different purposes and are extremely easy to confront between different works. They also supply the highest informational content: a sequence of n bases theoretically consists of n different (even if not always independent) characters.

Homologous sequences comparison can give a quick measure of the difference between the organisms from which they were obtained. As a consequence, they can be used in order to define taxa, although this is still a very controversial topic (see next sections). On a more complex level, they can be used for inferring phylogenies (see section 2.4).

Different sequence markers – sequences of different genes – have different properties. One categorization attempt can distinguish between protein-coding and non protein-coding genes. A classification that is also useful for the purposes of this work is between slow-evolving and fast-evolving sequences.

Fast-evolving sequences accumulate substitutions at a faster rate and quickly diverge during evolution of isolated genetic pools. This makes them perfect for discriminating between closely related taxa, down to intraspecific lineages. On the other side, because they quickly become saturated by substitutions (the degree of differences ceases to increase linearly and reaches a plateau), they are not very informative on lineages that split deep in time. Slow-evolving sequences are quite the opposite. They accumulate substitutions too slowly to discriminate between taxa that split recently, but some of them can still contain enough information to allow us recovering the most ancient speciation events, and hence the phylogenetic relationships between distantly related organisms. It is important to note that these categories are also dependent on the group of organisms studied. Generalizations are always risky.

It is at this point clear why a multi-marker approach is so tempting. Obtaining several different sequences from the same organisms allows making conclusions on firmer grounds. If the markers are chosen carefully, they can indeed be largely independent (even if not in the mathematical

sense; here independency means somehow uncoupled during evolutionary processes). Accord between the results of such an analysis is a good sign of robustness of the theory. Moreover, different types of markers can give, as said, different information on the single biological objects. Many individuals of the same species can have similar or identical slow-evolving sequences, that can be used to inferri their position in the larger tree of life; but also supply slightly or largely different fast-evolving marker sequences that give us information on their degree of divergence, the presence of genetic flows, the phylogeography, the place of origin of the species and so on…

In the next sections, I will very briefly review the different markers I used in these projects. They were chosen in virtue of their different and complementary properties, and (with one exception) for the extensive amount of data available on each of them.

1.2.2. Nuclear rRNA gene sequences

rRNA genes have unique properties. They are transcribed into structural RNAs, and mutations can have drastic consequences on the organism's fitness. They do not encode for proteins, thus they are not subjected to the genetic code rules. They are not subdivided into codons, and there is no easy way to predict which site will be more conserved or prone to substitution events.

It can be empirically showed, however, that some of them – especially SSU rRNA gene sequences – are fairly conserved through all living organisms. This led to the pioneering papers of Woese (Woese et al., 1990) that, using this gene as a phylogenetic marker, obtained one of the first phylogenetic tree of life. We now know that that tree contained some mistakes due to long-branch attraction artefacts, but the major splitting of life in three domains (Bacteria, Archaea and Eukarya) is still considered valid.

SSU rRNA gene can thus be thought of as a slow-evolving marker useful in assessing distant relationships (generally above the species level). It has to be said, however, that some regions of this gene can be pretty variable, especially in some groups (e.g. Petroni et al., 2002). It is also common to find more or less long inserts, and generally a lot of indels, that

can be used in order to discriminate between closer taxa (but that also make homologies assessing more difficult).

The most conserved regions can be used to design universal PCR primers (see section 2.2.2). Indeed, also domain-specific and class-specific primers exist, are very effective and can be used under most circumstances during routine analyses. Thus, SSU rRNA gene sequences are usually easy to obtain.

Another great advantage of this gene is that it is commonly used. Online databases contain millions of sequences from the majority of known organisms. All living beings that we know of, in fact, have ribosomes. Thus the possibilities of comparisons and meta-analyses are endless. Moreover, there are so many studies exploiting them that their properties as markers are well established. For bacteria only, there are even formal threshold values of SSU rRNA gene sequence similarity that allow (alongside other data) establishing new species- or genus-level taxa.

These thresholds are 98.7% for species and 95% for genera (Stackebrandt et al. 2002; Stackebrandt & Ebers, 2006). Even if these values can be considered somewhat arbitrary, it is important to remember that they rest on a huge amount of empirical data showing their correlation with other parameters. In addition, it is also fair to say that bacterial "species" are a blurred concept anyway, and "genera" are always arbitrarily defined. Thus, until more knowledge will allow developing sounder rules for bacterial taxonomy, these thresholds are gladly welcomed.

A formal species characterization requires however additional data, especially on the metabolic properties of the organism obtained through culture-dependent characterizations. Because many symbionts aren't able to grow outside the host body, the category of "*Candidatus*" has been established for those species whose SSU rRNA sequence is known (and validated through FISH experiments, see below) and other data are missing (Murray & Stackebrandt, 1995; Tindall et al., 2010).

There is another fact that renders this marker so useful in characterization projects. The DNA sequence of this gene encodes for a structural RNA that is constitutively present in a huge number of copies. This makes particularly easy to design fluorescent probes for FISH experiments (see section 2.3). Thus, we possess a quick tool to prove beyond doubts that

the sequence we obtained through molecular protocols belongs to the organisms we are interested in.

Using SSU rRNA sequences has of course its inconveniences. They are generally longer than many protein genes (about 1500 bp in bacteria and 1800 bp in eukaryotes), thus requiring more sequencing reactions to be completed. As slow-evolving sequences, they are also nearly useless for the purposes of population genetics (that usually focus on sub-specific taxa down to the organisms).

Adjacent to SSU rRNA gene, in virtually all prokaryotes and eukaryotes other rRNA genes can be found. Using eukaryotic nomenclature, they are ITS1, 5.8S, ITS2 and 28S rRNA. ITS1 and ITS2 mean Internal Transcribed Sequences 1 and 2. They are transcribed together with the rest of the rRNA operon, then processed and eliminated. They are theoretically more prone to substitutions, because they have no structural role. This assumption notwithstanding, the actual degree of variability depends heavily on the group of organisms (they can be hypervariable to pretty conserved). 5.8S and 28S rRNA are located, together with 5S rRNA, in the LSU and follow the same evolutionary role of SSU rRNA; they are generally less used as molecular markers because they are either not informative enough (5.8S rRNA) or too long to sequence and with a limited reference database (28S rRNA).

In this work, I characterized nearly complete SSU rRNA gene sequences and the complete ITS1+5.8S+ITS2 sequence. As said below, I sometimes take advantage of their proximity in order to obtain a single, longer sequence.

1.2.3. *cox1* gene sequence

Cytochrome *c* oxydase subunit 1 (*cox1*) gene is encoded in mitochondria of most eukaryotes and usually on the chromosome of prokaryotes. As for all good markers, its phenotypic function is uncoupled to the information it supplies. Nevertheless, it is important to remember that the protein it encodes is one of the fundamental enzymes in the aerobic respiratory chain. Thus, it is usually lost or absent in anaerobic organisms, like many bacteria and also some eukaryotes (even many ciliates). It is then a less universal a marker than rRNA genes.

The gene is of course protein coding, and this makes less easy to develop non-specific amplification primers (see section 2.2.2). It is also difficult to make FISH experiments on the sequence itself, because *cox1* mRNA have a short half-life. On the other side, it usually contains much less indels, and is thus easier to align (see section 2.4.3).

cox1 gene sequence became much famous as a molecular marker in the last decade, especially due to the paper of Hebert and others (Hebert et al., 2003). These authors, working with metazoans, were searching for a barcode marker, a sequence reliable enough to be unambiguously assigned to each single species in order to identify it.

They thought that such a sequence should be protein-coding (thus with few indels and easy to align), of mitochondrial origin (immune from recombination, haploid), easy to obtain through PCR (with some at least class-specific conserved regions at both ends), with relatively slow rate of aminoacidic change (so that the aminoacidic sequence could be used for identifying higher taxa) and relatively fast rate of nucleotide change (in order to be variable enough to act as a good species barcode). They argued that *cox1* gene sequence satisfies all of these requirements.

Since then, *cox1* gene sequence (to be precise, a specific portion of it, called the barcoding portion and about 800 bp long) has been used by many authors working with metazoans. For our purposes, it can be considered a fast-evolving sequence, useful for lower taxa systematics. It shares with SSU rRNA the advantage of an extensive literature and a wide public database.

The effectiveness of *cox1* sequence outside the metazoan phyla is still to be thorough evaluated, and a generalization of its properties is still waiting. Papers on ciliates are still few (e.g. Barth et al., 2006; Lynn & Struder-Kypke, 2006; Gentekaki & Lynn, 2010), and most of them only provides data on *Paramecium*, *Tetrahymena* and a few other oligohymenophorean taxa. One of the topics addressed by these works was the detection of previously unnoticed molecular variability inside morphospecies suspected to be complex of sibling species.

Only one paper actually tried to generalize the approach (and the amplification primers) in order to adapt this marker to the entire Ciliophora phylum (Struder-Kypke & Lynn, 2010). The results of this paper dictate carefulness. Even after intense efforts, the authors were not

able to develop universal primers that could work for all the classes of the phylum. Moreover, sequences from multiple strains of single or closely related species outside class Oligohymenophorea are still to be obtained and evaluated. There is then the problem of the important groups of anaerobic ciliates that lack this enzyme, and that cannot be studied with the same approach, thus decreasing its appealing.

Nevertheless, *cox1* is at the moment one of the best possibility available to detect intrageneric and patterns of molecular variability. In this work I obtained *cox1* sequences for all the ciliates I studied. I used it for purposes of identification (see chapter 3), new data recovering (see chapter 4) and biodiversity survey (see chapter 5)

1.2.4. Mitochondrial LSU rRNA gene sequence

Mitochondria produce their own ribosomes, which have some bacterial properties. Hence, they have their own set of rRNA genes.

There is no data whatsoever, at least for ciliates, about the properties of mitochondrial large subunit ribosomal RNA (mtLSU rRNA) as a molecular marker for phylogenetic inference or other analyses. It will turn out a promising marker if it could be proven that it has indeed the good properties of both rRNA genes and mitochondrial genes. With the results of this work and of future, more thorough researches, it could be possible to propose this marker as a contender with similar functions to *cox1*, but easier to obtain.

1.3. Overview of the work

My thesis work comprises three different projects, united by the features discussed in sections 1.1 and 1.2: all three of them have ciliates as model organisms, and use a multi-markers molecular characterization approach as a methodology. The differences reside in the species actually studied, in the purposes and in the use of the obtained data.

The first system studied, detailed in chapter 3, was a population of *Paramecium jenningsi* whose macronuclei were infected by *Holospora*-like bacteria. The purposes of this study were straightforward and well rooted in literature. *Holospora*-like bacteria are indeed, as will be said later, very well known and relatively common organisms, whose characterization from a molecular point of view is lacking and somehow requested. Therefore, when a ciliate-bacterium symbiosis with the right characteristics was discovered in our lab during sample screening, the path was clear: obtaining as much commonly used molecular data as possible. I've done this both for the bacterium and the host. Results about the host opened some interesting speculations for the future. Results on the bacterium are both in accord to previsions and original. They allow us to classify this organism in a new genus, contributing at the same time to clarify the taxonomy of this group.

The second system studied concerned a much less known group of ciliates, the genus *Condylostoma*. I worked with different strains, some of them very carefully characterized from a morphological point of view, in order to evaluate the systematic of this genus. The result was the first phylogenetic analysis of class Heterotrichea focused on the relationships inside this very tangled taxon. In the meanwhile, as will be explained in chapter 4, we found some hints of the presence of a bacterial endosymbiont inside the cells of one of the *Condylostoma* strains, and I tried to collect molecular sequences also from it. While the previous project was supposed to add more data to an already well-studied group of organisms, the interest of this symbiont resided in its novelty. There aren't other known symbionts of genus *Condylostoma*, and not many from the marine environment.

The third sub-project was focused on the species *Paramecium duboscqui*, which is known to sometimes contain symbionts. As will be explained in chapter 5, just few of the more than 20 species in genus *Paramecium* are indeed widely used by biologists as models, and are hence very well known. Most of them, especially those from brackish water, are much more elusive and unfamiliar. In recent years, many of them were surveyed for the first time from a molecular point of view (e.g. Catania et al., 2009; Przybos et al., 2009). I performed a similar work on *P. duboscqui*, working on DNA of strains collected from very different sources around the world and confronting the results with the very few others present in literature. I focused on assessing the intraspecific molecular

diversity and its relationship with biogeographic pattern. I also tried to evaluate the hypothesis evoked by other authors about the possibility that *P. duboscqui* is indeed, like *P. aurelia* (and probably *P. bursaria*) a complex of sibling species. Moreover, I used the DNA material in my possession as a starting base for assessing a completely new molecular marker (the sequence of mitochondrial large subunit rRNA) and evaluate its possibilities.

2. Methods

2.1. Practical ciliatology

2.1.1. Sampling and sample screening

Ciliates sampling techniques are very variable according to the purpose for which they are made. For example, paramecia are never collected in marine environments, and there are entire groups of pelagic, actively swimming ciliates (like choreotrichs) as well as others specialized in crawling in interstices of the substrate (like hypotrichs). In the field, the act of sampling is fairly different if the researcher is looking for a particular taxon or if he wants to assess the overall biodiversity of the place.

Our laboratory is usually focused on interstitial ciliates from shallow body waters, both freshwater and saline. To obtain these organisms (which often include hypotrichs, stichotrichs, heterotrichs, oligohymenophoreans, karyorelicteans, and more rarely litostomats and prostomats) the sampling method simply consists in picking up an aliquot of substrate as well as the above medium in a rough proportion of 1:2 in a sterile container of glass or plastic. Most useful for this procedure is the 50 ml Falcon tube.

When the sample is taken, it is always important to note and write on a label data like the general conditions of sampling (morphology of the water body, time of the day, climate, even climatic event of the previous days…) and the exact position of sampling. It is not always possible, but generally recommended, to note also a few basic physico-chemical properties of the water, most importantly temperature and salinity.

The sample can usually be stored for some days (1 at most if hermetically closed) after sampling. Waiting for too long can lead to severe alterations of the physico-chemical properties of the medium and/or of the biological communities. Common cases include anoxia (or oxygenation),

heating and predatory activity of metazoan on the microbial community, especially in freshwater samples rich in small crustaceans.

The sample is then screened in the lab under a stereoscope. A small aliquot of the substrate/liquid mixture is taken after some gently mixing, and poured in a Petri dish. It is generally easy for an experienced observer to approximately determine the number of different ciliate taxa present, and their class of affiliation. Genus/species level identification in most cases requires the use of a histological microscope.

The original sample can then be kept for some time, even for months, in its original container or in a Petri dish or Boveri. It is necessary to maintain the amount of trophic resources in order to sustain the biological community; this can be accomplished adding a rice grain (which constitutes a nutrient substrate for bacteria and fungi) or some drops of algal culture. Anyway, generally the ciliate community changes in many ways over time. Some taxa can indeed become more abundant after the enrichment, or even appear anew after some excistation process. Some other can quickly decrease in number and finally disappear.

2.1.2. Culturing

When the researcher is interested in studying a specific taxon found in the sample, it is always better to culture it. In general, the term culturing refers to an increase of the organism's cells number under partially controlled conditions. In this work, and as a rule in our lab, culturing specifically means making the cell number increase in a medium fully depleted of other eukaryotic organisms, with the possible exception of the "food" organism(s). Most importantly, no other ciliate taxa should be present. Fungi and flagellates are other common contaminants that should be removed. When the result has been achieved, it is important to handle the culture under the most sterile conditions possible.

It has to be stressed that not all ciliates can be cultured in this way, for reasons not completely understood. It is also important to remember that some degree of bacterial contamination is unavoidable, even when the ciliate are fed with other eukaryotes. Thus, every time that some completely axenic material is needed, it is important to use additional precautions (see next session) in order to eliminate the contamination even from the most "clean" culture.

Another important point, often underestimated in the literature, is to note clearly if the culture was started from either a population or a single cell. The latter case is usually preferable, and the resulting culture is called a monoclonal culture or a strain. As long as a sexual process (conjugation or autogamy) doesn't happen, all cells in the culture can be treated as a genetic clone, all with the very same genetic material.

The first step towards a monoclonal culture is to isolate a single cell of the target organism from its original medium, "washing" it (see next session) and put it in a new, sterile medium. For a variety of reasons, some cell can die after this treatment, so it is generally preferable to make several attempts. The isolated cell is then fed with an appropriate prey organism – known from literature, experience or attempts. Whenever it divides asexually a certain number of times, it is important to add more volume of the medium. Some ciliate – like most *Paramecium* species – can grow quickly, doubling their number every 1-2 days. We speak of a mass culture when the ciliate culture reaches the volume of some liters.

In our lab there are three available kinds of food organisms, each one cultured in its own way: the unicellular green alga *Dunaliella salina* (Chlorophyceae, Volvocales), the diatom *Phaeodactylum tricornutum* (Bacillariophyceae, Naviculales) and bacteria belonging to the species *Raoultella planticola* (*Gammaproteobacteria*, *Enterobacteriales*).

In addition to food, some physical parameters are also kept constant. The cultures are stored in a thermostatic chamber with a temperature range between 19°C and 20°C. Lamps producing a radiation spectrum similar to natural light and performing a day/night cycle of 12 hours provide the light source. The optimal salinity of the medium of each culture is checked during the process of feeding and adjusted in order to remain stable.

2.1.3. Handling

Ciliates swim, crawl or stay on the top of their stalks inside the culture medium. They can usually be brought in suspension by gently shaking the container and then be transferred to another container pouring some volume of medium. In this way the medium can also be filtered with sterilized gauzes in order to eliminate some debris, like sheet of dead algae or bacterial floccules.

When axenic cells are needed for DNA extraction or FISH experiments, however, it is necessary to collect one by one the cells, virtually without bringing along the surrounding medium. This is generally accomplished using a glass micropipet obtained from a Pasteur pipet stretched on the flame of a Bunsen burner. The diameter of the micropipet should be roughly similar to the size of the ciliate, thus making glass micropipets a quite flexible instrument.

In order to reduce the the presence of the "food" organisms, ciliates targeted for the experiments are let starving for some days.

In addition, it is sometime useful to "wash" ciliate cells in order to remove all traces of the previous medium and other organisms it contained. This is accomplished in the simplest way with a three-depression glass. Each of the three depressions is filled with sterilized medium. The ciliate is brought to the first depression and allowed to swim for some seconds while residual original medium gets diluted. The operation is then repeated two other times (or more, if necessary) before considering most contaminations removed.

"Cleaned" cells can then be used. In this works, their fate could be being fixed on a slide for FISH experiments or stored in ethanol 70% in an Eppendhorf tube for DNA extraction.

2.2. Molecular characterization

2.2.1. DNA extraction and storage

The first step towards obtaining DNA sequences is obtaining the DNA itself. The starting material always consisted of isolated cells from the ciliate population/clone of interest stored in ethanol 70% at -22°C. I worked under the assumption that there were neither other ciliate contaminants in the sample, nor other eukaryotic organisms large enough to be visible at the stereomicroscope. Some degree of bacterial contamination in the medium is inevitable, and it is also wise to consider the possibility of some residual DNA from the food organism.

All this considered, the strategy was simply to isolate the total genomic DNA (tgDNA) present in the sample. This includes the macro- and micronuclear and mitochondrial genomes of the ciliate, and the chromosomal and plasmidic DNA of the bacteria, both endosymbionts and contaminants.

To this purpose, it was always used the Nucleo Spin® Plant II kit (Macherey-Nagel), performing a slightly modified version of the protocol for fungal DNA extraction. The choice to use a kit was somewhat obliged because of the very small amount of starting material. Albeit more expensive, commercial kit reactions are known to give higher yield than manual protocols.

The tgDNA obtained was always eluted in distilled and autoclaved water and stored frozen at -22°C. The final concentration of DNA in the solution was never quantitatively estimated, in order not to lose material. Elution volumes were decided each time, proportionally to the number of stored cells and the species they belonged to. Usually the first PCR reaction, aimed at the amplification of SSU rRNA of the ciliate, was used as an indicator of the good quality of the DNA.

2.2.2. PCR, electrophoresis and primers design

PCR reaction is probably the most commonly used tool in molecular labs, and it needs not to be reviewed here. Many practical and theoretical laboratory manuals and textbooks devote chapters to this topic (e.g. Wilson & Walker, 2005; Sambrook & Russell, 2000).

For my purposes, I performed classical PCR reactions and some simple variants, like semi-nested or nested PCR when one amplification reaction didn't produce an adequate amount of material. I also sometimes performed PCR reactions directly on the organisms, without isolating the DNA. This was the case for single-cell PCR on ciliate (simply pipetting the PCR solution on the cell stuck to the bottom of an Eppendhorf tube) and control PCR on bacterial clones during cloning (see next section). When some degree of interaction between the primers was suspected, I performed some amplification cycles in half the final volume with only one primer. This led to a linear amplification of the region of interest that becomes enriched for the actual two-primers PCR.

I employed the "Primus 96 Plus" (MWG Biotech) thermal cycler for control PCRs and "C1000TM" (BioRad) thermal cycler for all other purposes. Reagents used includes ex Taq (TaKaRa®) polymerase and the buffer solution and dNTP mixture of the same company.

PCR products were always evaluated through electrophoresis on 1% agarose gel. I used Agarose Electrophoresis Grade (Invitrogen®) diluted in TBE buffer (10,8 g Tris; 5,5 g boric acid and 4,0 ml EDTA pH 8.0 0,5 M in 1 L of distilled water). As loading buffer, I used a solution of glycerol (30% in water) with 0,25% bromophenol blue and 0,25% xylene cyanole. After a 30-45 minutes run, the gel was soaked in an ethidium bromide solution (0.5 μg/ml) for at least 20 minutes, and then observed with a transilluminator.

In order to exploit (e.g. sequencing or using as a template for a semi-nested PCR) the PCR products, they had to be purified from excess primers, PCR mixture salts, polymerase residues and low-weight aspecific products. To this end, the kit NucleoSpin® Extract II (Macherey-Nagel) or Quantum Prep® PCR Kleen Spin Columns (Bio-Rad) were employed.

Many primers and probes were developed for this work, especially for mitochondrial markers. A detailed list is shown in the Appendix, and here I just briefly review the methods and criteria used in designing them.

All primers and probes were developed with a degree of specificity in mind. For example, it is often necessary to design a probe that matches specifically a single sequence, allowing discrimination between the organisms that possesses that sequences and other, closely related, organisms. In other instances a more general primer was needed. In both cases the first step always consisted in finding the most appropriate region of the target sequence – a conserved or variable one. This was manually accomplished using the ARB software package (Ludwig et al., 2004) and its graphic editor, visualizing the database of choice.

For rRNA sequences, forward and reverse primers can simply be designed using the opposite DNA strands as template. Many features have to be considered in order to evaluate the potentiality of the primers pair. First of all, the more the two primers are distant, the better. This allows the amplification of longer (under a threshold given by limits of the PCR) fragments of DNA. The melting temperatures of two primers used contemporaneously should be approximately the same, and in the range of

50-60°C (usually 18-20 nucleotides long). It is also important to check that they don't bind to regions other than the one of interest, and that they don't interact with each other. Because the 3' end of the primer is the most important for the polymerase activity, this end has to match more precisely as possible the template – mismatches in the last 5-6 nucleotides have to be absolutely avoided.

For probes targeting RNA, of all the aforementioned features only the melting temperature is important. Probes used in this projects bind directly to the rRNA, so the secondary structure of the molecule plays a part in the binding process and has to be considered when evaluating how promising a region is. Mismatches are more penalizing for binding when in the middle of the probe's sequence.

For *cox1* sequences amplification, as for all protein gene sequences, the problems are bigger, especially if a "general" primer has to be designed. The reason is that even in the more conserved regions the third position of each codon is prone to silent mutations. The sequences are just too variable, at least above the species level. Trying to develop as general a primer as possible, a different approach was used. I started with the aligned aminoacidic sequences and searched for a conserved 5-6 aminoacid fragment. Not all aminoacids had the same appeal during this search: the less degenerated (encoded by few codons, like methionine or tryptophan) were preferred. After that, I generated the set of all nucleotide sequences that encode for those aminoacids. Then I designed a "degenerated primer" that matches all of these sequences (in practical terms, a mixture of all possible primers – this poses the constraint of using less degenerated aminoacids in order to avoid too high numbers of possible sequences). This constitutes the degenerated portion of the primer.

The degenerated portion is about 15 bp long, usually not enough to reach the desired melting temperature. In order to accomplish this goal, a consensus portion was added at the 5' end. This portion is designed on actual nucleotide sequences in the database, choosing the most appropriate (usually the more frequent) base whenever there is a non-conserved position.

As will be briefly explained in section 2.2.4., sequencing is usually accomplished with internal primers. Because the design of these primers has the same difficulties of the design of amplification primers, it was not practical to develop many of them for *cox1* gene sequences. I employed

then a different approach, adding to the 5' end of the consensus portion an arbitrary sequence (that of a universal primer like T7 or M13R) that could be targeted by sequencing primers in order to obtain the entire sequence of interest.

cox1 primers developed for this project are thus composed of three portions: the universal primer portion, the consensus portion and the degenerated portion. Even those primers obtained from literature (consisting only of a degenerated portion) were modified in this way.

2.2.3. Cloning and RFLP screening

Cloning is an older method than PCR, and has many applications. In my work, I used it only in conjunction with PCR, when PCR products were not suitable for direct sequencing because of the presence of contaminating amplicons. Cloning allows separating different amplified DNA fragments, in order to sequence and identify each of them.

The materials and protocol used were those of the kit TOPO TA Cloning® (Invitrogen). The plasmid is in its linear form, and has a free thymidine residue at the 3' end, that should pair with a free adenosine residue at the 5' end of the amplified fragment. (This is normally generated by the activity of the ex Taq polymerase, but can be lost after some time. Thus, if the cloning reaction was performed on a different day than the PCR reactions, an A-tailing step was added. To perform this, a mixture of buffer solution, dATP and exTaq was incubated at 70°C for 30 min.) The topoisomerase I of *Vaccinia* virus is covalently bound to the plasmid, and trigger the insertion reaction.

The PCR product that is suspected to contain DNA fragments of different origins in unknown relative ratio is included in a linear plasmidic vector that circularized. When the insertion reaction is successful, a plasmidic gene responsible for X-gal catabolism is interrupted, and is no longer functional.

Through a process of induced transformation in competent cells of *E. coli* (Mach1-T1®), each plasmidic vector is introduced in the cytoplasm of a bacterial cell. It is assumed that, thanks to an extremely low efficiency of the transformation process, a single cell assimilates at most one vector.

E. coli cells are inoculated in a liquid medium, where they can grow for 30'', and then plated on a Petri dish with LB-agar medium with ampicillin and X-gal added. Only transformed bacterial cells can grow on ampicillin, because an ampicillin-resisting gene is coded by the plasmid. X-gal allows discrimination of cells transformed with a plasmid vector that lacks the insertion. These cells can break down X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) through the beta-galactosidase encoded by the plasmidic functional gene to galactose and 5-bromo-4-chloro-3-hydroxyindole. The latter dimerizes into 5,5'-dibromo-4,4'-dichloro-indigo, an insoluble blue pigment.

As an assumption, each colony on the Petri dish originates from a single cell (and hence is a clone). White (non-blue) colonies should contain one of the inserts of interest inside their plasmid. A certain number of them are thus transferred to one or more other Petri dishes, called Master Plates, and numerated in order to guarantee their subsequent identification.

The screening of Master Plate colonies follows. A small aliquot of each colony is subjected to a control PCR. The primers used are plasmid-specific and target DNA regions immediately external to the insertion point. Presence of a band of the expected weight in the electrophoresis run confirms the success of the insertion. Bands of different molecular weight usually belong to unspecific PCR product included in the plasmid instead of the target gene. When they appear, the corresponding clone should be discarded from subsequent analyses.

All control PCR products are then digested with a restriction enzyme, and the restriction patterns are compared through an electrophoretic run on 2% agarose gel. One or more clones representative of each pattern are chosen and inoculated from the Master Plate in liquid medium in order to obtain more bacterial cells with the insert of interest. Eventually, a plasmid DNA extraction is performed on the clone grew on liquid medium (in my case, using the kit ZR Plasmid Miniprep^{TM}-Classic of Zymo Research). The obtained DNA concentration is evaluated through electrophoresis. The product is then sent for sequencing using the same plasmidic primers used for control PCR.

I've always sequenced three clones for each pattern in order to obtain a consensus sequence. This wariness is absolutely needed to circumvent random PCR errors fixed by cloning procedure. The degree of difference

from each clone sequences is in the order of 0-3 bases every 1.000 bp for bacterial SSUrRNA.

2.2.4. Sequencing and preliminary sequences analysis

DNA sequencing was performed by MACROGEN Company. I sent the DNA processed as requested, and the outputs were electropherograms encoded in different formats.

A good sequencing reaction usually gives an electropherogram with an unambiguous 600-800 bp long sequence. This isn't enough to cover all the length of, for example, SSU rRNA gene sequence. Thus, more sequencing primers were requested for each sequence. Carefully choosing these internal primers allows to cover all the sequence and to have at least some regions of overlap that can be crosschecked. Sequences of the amplification primers that are present at both ends of the complete fragment sequence were always removed.

For visualizing and editing the sequences, I used the software Flip4Mac on a Macintosh computer and Chromas Lite on a PC. During this step, electropherograms were crosschecked, their quality was assessed (by eye and with the help of parameters calculated by MACROGEN) and the final sequence was assembled.

The very first analysis performed on each sequence employed the free BLAST tool of NCBI. The outputs of BLAST were employed to verify the most probable affiliation of the sequence and to make same preliminary considerations. This was a simple method for discovering sequences, like those of common environmental bacteria, which were very unlikely to belong to the organisms of interest and could thus be ignored.

2.3. FISH experiments

FISH is a very useful and common tool in molecular biology studies with different aims. In the field of ciliates bacterial endosymbionts, it is useful to start with a FISH experiment with universal bacterial probes in search for bacteria in one of the cellular compartments. They are often not easily recognizable with histological microscope techniques, and difficult to catch looking at random sections at the TEM. FISH is also one of the last experiments that have to be made during a molecular characterization study. When a rRNA gene sequence from a putative bacterial symbiont is obtained, it is mandatory to prove that the sequence actually belongs to the organisms of interest. This is accomplished designing a sequence-specific probe on the obtained sequence and validating it by hybridization against fixed host cells.

In the form I used them, FISH experiments can be summarized in this way. Host cells were fixed on a slide and covered with a mixture of two oligonucleotide probes with different fluorophores and a hybridization solution. The latter is needed for adjusting the binding conditions, and can be modified to increase or decrease the stringency (and thus the probes specificity). The slide got incubated at 46°C for some hours. Afterwards, it is subjected to some washing steps in order to eliminate excess and unspecifically bound probes. A solution that preserves the fluorescence was pipetted on the hybridized cells. In this solution is also present the DAPI (4',6-diamidino-2-phenylindole), a fluorescent dye intercalating DNA. The sample observation is accomplished under an epifluorescence microscope.

In my FISH experiments I always used this two-probes combination that allows testing more than one hypothesis at the same time. The two fluorophores always emit in the wavelength of red and green respectively, and the DAPI emits in the range of blue, so that the signals are different and can be overlapped on the same picture in order to confront them. The DAPI is not a probe, but is still useful as a generic marker for the presence of genetic material. Comparing DAPI fluorescence it with the probes' signal can help finding some organisms that the probes didn't bind to. Moreover, it also highlights the macro- and micronuclei morphology, providing a data that can be compared with other, more specific morphological techniques.

As chemical fixatives I used, according to necessities and species, formaldehyde 4% in PBS buffer (directly added in 1:1 ratio to the drop of medium containing the cells), osmium tetraoxide 4% in aqueous solution, directly or as vapours (one drop on a different slide put near the one with the cells for about 1 minute).

2.4. Phylogenetic analyses

2.4.1. On the way to the tree

When sequences are obtained, one of the ways to use them is building a molecular phylogenetic tree based on them. By definition, this is a representation of the dichotomous process of lineages' branching from a common ancestor to the living organisms that we are studying. It is composed of branches and nodes. Nodes represent ancestors, and from each of them two branches depart, symbolising different and isolated descendent lineages. Terminal branches coincide with existing taxa, or OTU (Operative Taxonomic Unit), a term useful to avoid tricky labels like "species" or "populations".

In reality, linking the diagram that we obtain with the definition given above requires a lot of assumptions. The first is that the evolution of the involved organisms was indeed tree-like. There are other ways of seeing molecular evolution, but in what follows I will always talk of dichotomous, rooted trees. This means that I assume that the process of speciation is nearly instantaneous (relative to the period of time the tree analyses), always leads to two daughter taxa, both different from the ancestral one, and that I know which split is the oldest.

A second problem is to assume that gene trees reflect organism trees. This is probably true with the systems we are dealing with, but there are very well known examples involving Lateral Gene Transfer of genes in bacteria and even eukaryotes and gene family whose paralogs split before the species that carried them.

A third assumption is that our method of inferring the phylogenetic relationships from the sequences is correct. This constraint is a bit relaxed

when we use methods based on different principles and they all give the same, or similar, tree topology. This doesn't mean that all methods are equally reliable, though. And some basic assumptions are often shared by many methods. For examples, all of them explicitly rely on the fact that each base substitution is an independent event, a statement that is clearly violated in many cases.

Finally, there is the question of using molecular sequences with the correct evolutionary rate for the analysis we want to make (see section 1.2). Sequences saturated by substitutions are useless, because they don't contain anymore the unbiased evolutionary information required.

On every single problem listed here, and many others, hundreds of articles and many books have been written, and it is not possible even touching most of the disputes here. In what follows, I'll give just an extremely succinct summary of the process of inferring phylogeny from molecular sequences as I have performed it. For the details, I refer to the two textbooks I used to learn the basic of this discipline (Felsenstein, 2004 and Lemey et al., 2009). One last important remark is that, although I will always refer to nucleotide sequences, many of the methods I will cite can be used, with some modifications, for aminoacid sequences, morphological characters and every other kind of character subjected to evolutionary processes.

All of my phylogenetic analyses started choosing the correct sequences to analyse and then aligning them. The result is the data matrix, which is then processed differently according to the method of choice. Before and after the main analysis it is better – and sometimes essential – to make some statistical tests on the data matrix and/or on the tree topology. The final step consists in summarizing the results in a clear and comprehensive way. All of these steps will be briefly explained in the next sections.

2.4.2. Choosing sequences

It is obvious that the sequences obtained during a work will be exploited for the subsequent phylogenetic analysis. Sometimes this is enough, because all we want to do is comparing them with each other. In most cases, though, we need also other sequences, previously obtained by other researchers, to make a more extensive analysis. This is the case, for

example, when we want to know which are the most strictly related taxa of a newly discovered organism.

There are many ways to choose these sequences from the online databases, and they all try to answer two main questions: which ones and how many? We need sequences clearly homologous to our own, of about the same length, of good quality, possibly from unambiguously identified organisms. Concerning the number, from a theoretical point of view the more, the better, but increasing the number of sequences and their lengths makes quickly the analysis computationally too costly. It is indeed true that the magnitude of the data is limited by the computational capabilities of the computer and the time we can afford to dedicate to the analysis.

I always picked the sequences from the SILVA database (Pruesse et al., 2007) imported on the ARB software package, that can manage a vast number of sequences. The general tree built by the software is useful in quickly find other sequences of the clade of interest. The most important characteristics I soughed were length, quality (assessed mostly by the year of sequencing and by characters clearly divergent from those of other similarly classified sequences) and presence in articles of interest. Sequence similarity must remain well above 50% for nucleotide sequences. I tried to avoid clone sequences not presented as a *consensus* sequence from at least three clones. I included sequences labelled as "uncultured" (mostly from environmental studies) only when they were useful for the analysis, very similar to mine or interesting for some other reasons.

I constrained the total number of sequences in the analysis between 50 and 70. As said before, this limit is imposed just for reasons of time and computational effort. With these numbers, most of the analyses could be performed in 1-10 hours.

In every extensive analysis, I always included "outgroup" sequences. If one doesn't assume a molecular clock-like behaviour of evolution (and I didn't; Zuckerkandl & Pauling, 1965), this is the only safe way to root the tree. It is indeed possible to put the root between the outgroup clade and the rest of the sequences. Sequences were always considered good as outgroup based on the literature of the analysed clade. To avoid notorious problems like long-branch attraction and similar artefacts, whenever possible I choose sequences from the most related taxa outside the clade of interest. It should be remembered that phylogenetic relationships within the outgroup cannot be relied upon.

A list of sequences, their references and accession numbers can be found at the end of the Appendix.

2.4.3. Alignment

Alignment is probably the most tricky and underestimated problem in molecular phylogeny. One of the most basic and fragile assumptions for all the subsequent steps is that all single bases are confronted with homologous bases in different sequences. This is obtained placing them in columns that, on the whole, form the data matrix.

In sequences of the same lengths, without indels and with conserved regions at the two ends, this is easy to obtain. But for sequences coding for rRNA, for example, this is almost never the case. There are lots of indels, and often some taxa or groups of taxa contain a very long "insert" in their sequence. So, a lot of additional assumptions have to be made in order to obtain an alignment.

The job can be done with the help of a software or manually – nearly always, using both approaches sequentially. Whenever homologous sequences are not of the same lengths (which means they have indels), gaps have to be included. Every gap is an assumption. Moreover, adding gaps indiscriminately is a good way to align if we look just at the final similarity value, but is very unreasonable from a biological point of view. Software deals with this problem giving all gaps some sort of penalty. Penalties are subtracted from the final scoring value, a quantity calculated in different ways that is proportional to the similarity of the aligned sequences. The algorithms usually try different alignments and keep that with the higher value. Scoring formulas for gap penalties are more based on common sense than on some biological or statistical grounds. Moreover, the goal of maximising the final score trying all possible alignment and searching for the best one is computationally burdensome, so all software use heuristic algorithms to speed up the process. Heuristic algorithms don't guarantee to find the best alignment according to the optimality criterion they try to reach (in this case, the maximisation of the score).

I used a mixture of the alignment algorithms of the ARB software, ClustalX and manual editing of the sequences. ClustalX (Thompson et al., 1997) is one of the two versions of Clustal, the software most used for multiple sequence alignments of both nucleotides and aminoacids. It uses a

heuristic algorithm of the class called progressive alignments. Progressive alignment starts building a preliminary tree on the unaligned sequences using some fast method (usually Neighbor Joining, see below). The sequences are then ordered according to their similarity. The algorithm then uses exact methods to align the two most similar sequences, and bind them to share all subsequent modification. This step is then repeated for all other sequences, and thus requires just N-1 iteration for a group of N sequences. Even if it is not the most accurate of methods, it is still the most commonly used. Another useful feature of Clustal is the possibility to convert the input file containing the sequences in many other formats.

Clustal, like most software, uses generalized weight matrices and gap penalties in calculating the final score of the alignment. ARB package is instead optimized for rRNA sequences, which produce secondary structure like stems and loops. In portions of the sequence predicted to be stems, a change in one nucleotide can selectively drive a compensatory change in another. The graphical interface of ARB shows the predicted positions of stems and loops, helping manual editing.

Once the alignment step is over there are still some decisions to make before obtaining the final data matrix. Using the ARB package, all of these decisions are summarized in a filter. For all my analyses, I reduced the number of characters (columns in the alignment) to that of the shortest sequence. It is trickier to decide if delete some part of the aligned sequences. Reasons to do that are many: for example, that portion could be too difficult to align, and many ambiguities remained. Moreover, some sequences can have unique inserts, more or less long strings of nucleotides with no homology with other sequences; inserts may lead to an overstating of genetic distances. On the opposite side, every decision to exclude some part of the alignment can be seen as an arbitrary loss of information. The literature is not helpful in answering these questions, and there aren't shared solutions.

ARB filters can exclude columns according to the relative abundance of gaps and the values of base frequencies. I usually remained very conservative and kept most or all characters. Exceptions and reasons to make them are explained case by case. A filter at x% is a filter that excludes all columns in which the most common base is present in less than x% of the sequences.

2.4.4. Evolutionary models

All phylogenetic methods, with the exception of those of Maximum Probability, require an explicit, stochastic evolutionary model to make accurate calculations. Evolutionary models of molecular evolution, however simplified, give a much more reliable theoretical background for subsequent calculations.

Without a substitution model the only reasonable ways to quantitatively compare two sequences are unsophisticated. The most commonly used is the value of similarity, defined as the percentage of identical bases shared by the two aligned sequences.

A genetic distance between two sequences can be defined as any quantity proportional to the evolutionary divergence occurred between the sequences since the event of speciation from the most recent common ancestor. The inverse of similarity, called observed distance or p-distance, can be used as a very approximate genetic distance. The approximation comes from the incapability of p-distance to take into account any kind of multiple changes on the same site. Two changes in one position are counted just once or not at all, if the second reverts the first. Thus, p-distance always underestimates the actual genetic distance.

Building a model of molecular evolution means trying to evaluate different parameters and analytically define a set of probability functions that give, among other things, a corrected formula for the value of genetic distances. I will only briefly summarise this field, because it is built much more on maths than on biological questions. The best achievement that can be reached is the development of a statistical model practical to use, biologically reasonable and capable of explaining the patterns observed.

In order to generalize the discussion, it is useful to think of an evolutionary model as a 4x4 Q matrix, with the four nucleotides on rows and columns. Values in every cell correspond to the instantaneous substitution rate from one nucleotide to one another. Here we have already made an assumption, albeit a somewhat reasonable one: that the substitution process always follows the same rules, and that it is not affected by past history. In mathematical terms, that it is a Markov process.

Another assumption that is usually made is to consider just time-reversible models, in which every change from status A to status B has the

same rate and probability of a switch from status B to status A (note that if we knew for sure that this is not the case, we would have an additional way to root a tree). This gives symmetry to the Q matrix; because terms in one of the diagonals assume sense only from a mathematical point of view (the transition from A to A is of no particular biological interest), we obtain just 6 free parameters. This is called a General Time-Reversible (GTR) model (Lanave et al., 1984). If we put more constrains on the parameters, we obtain a vast number of simpler models, the most simple of which is the Jukes-Cantor (JK) model that assume all of the 6 parameters equal to 1 (Jukes & Cantor, 1969).

It is possible to show that there are (conceptually) simple formulas linking the Q matrix, the probability of observing a particular status character and the genetic distances between two sequences. For our purpose, the important thing is that these formulas exist, software can use them and can also estimate the free parameters of whichever model from the data matrix. It is important to highlight that formulas for estimating the parameters values are based on the concept of likelihood (see below). Parameters values are indeed the maximum likelihood estimators (MLE), constrained by the data, of the true, unknown values of the model.

Some additional notions are needed. One rough way to evaluate parameters values from the data matrix is to consider all columns (nucleotide homologous sites) as equally prone to variation and substitution processes, thus deriving the values as a mean from evaluation on each column. But it is biologically much more reasonable to think that different positions can have different rates of substitution, and hence different parameters values in the model. The mathematical tool to simulate this is the Γ function. Γ function has no particular reason to be used, outside the fact that it is mathematically tractable and has a useful shape. However, it is not used in its continuous form, but is discretized into n categories. The shape of the Γ function depends on a single parameter α, which can be estimated from the data. In most cases, α is less than 0,5 and the Γ function has a decreasing shape. Then, each column is assigned to one of the n categories, from the less to the most variable. In practice, there will be n sets of parameters for each analysis.

Γ function is introduced to correct for unequal substitution rates along the sequences. It has been showed that its impact is heavy, and it is generally recommended to use it in the model. However, the time and

computational effort grow linearly with the number of discrete categories. When I've used a Γ function inside an evolutionary model, I've always used 4 discrete categories as a compromise between accuracy and rapidity.

Another component of substitution models is the percentage of invariant sites, also estimated as a MLE from the data set.

2.4.5. Methods employed

Given the data and an explicit evolutionary model, to build a tree it is still necessary to use explicit rules for clustering the sequences. There are many possible methods, each one with philosophical and practical advantages and drawbacks. No one can be considered the best under every possible aspect. What I've done was performing the analyses with a fair sample of the most commonly used methods. Agreement between the results of methods based on very different assumptions is usually considered a good way to corroborate a phylogenetic hypothesis.

There are different ways of categorizing phylogenetic methods. One is to distinguish between clustering algorithms and methods based on optimality criteria. Another is to distinguish between distance-matrix methods and characters methods. As said above, all of them with the exception of MP methods, employ (or should employ) an explicit model of molecular evolution.

Statisticians prefer methods that search for the best tree according to a clearly defined optimality criterion. These include some distance-matrix methods, Maximum Parsimony, Maximum Likelihood and Bayesian Inference. All of them work in a similar way, but with different criteria. First of all, they need an algorithm to evaluate, for a given tree, its value according to the adopted criterion. Theoretically, one possibility is to evaluate all possible existing trees and choose that with the higher value. For more than about ten aligned sequences, this option is absolutely impracticable from a computational point of view, though, and heuristic methods are needed. There are literally hundreds of these methods, but in principle most of them can be used for all criterion-based phylogenetic methods. What they do is sampling the tree space, the universe of all possible trees, keeping the "best" tree they can find according to the criterion stated. Accuracy of the results and speed of computation are the

main properties of a heuristic method, and it is usually necessary to find a compromise between the two.

Clustering algorithms have the advantage of being immensely faster. They consist of a set of fixed rules and require a number of steps that increases linearly with the number of sequences. The major disadvantage is that they are based on very fragile theoretical assumptions, and are nearly non-testable from a statistical point of view. Although some of them are sometime considered an approximation of a related criterion-based method, the accuracy is usually low. UPGMA, WPGMA and Neighbor Joining are examples of clustering algorithms, and they are all based on the distance-matrix. Only the latter is still largely used.

The difference between distance-matrix methods (DDM) and characters method is the data on which they work. While the latter uses the aligned data matrix as a whole, distance-matrix methods first transform it into a pairwise distance matrix. For N sequences, this is an NxN matrix with values of genetic distance (usually corrected using an explicit evolutionary model) between each pair of sequences. All the information content is then reduced to N^2 numbers; the advantage is that much greater amount of data can be handled. Supporters of DDM showed that the informative content of the distance matrix is proportional to the total amount of information in the original data matrix. This would imply that genetic distances are all we need to infer phylogenetic relationships. Most of authors are sceptical, but DDM are still commonly used for analysing enormous data sets or as a first, not very accurate but faster attempt. DDM can use clustering algorithms (like Neighbor Joining) or optimality criteria (like Minimum Evolution, which search for the tree with the minimum total branch length). Characters methods almost always use an optimality criterion and a search in the tree space.

Neighbor Joining (NJ) (Saitou & Nei, 1987) is a distance-matrix method and a clustering algorithm. It is sometimes described as an approximation of the Minimum Evolution criterion, but this is a mathematically wrong description. Like other clustering methods, it has the disadvantage of poorly tractable statistical properties and, like other DDM, it works on a simplified data set, is heavily dependent on the correction used in calculating genetic distances and is prone to the problem of long-branch attraction for distantly related sequences with a non-clocklike behaviour. These drawbacks notwithstanding, it is still very

commonly used. It can build trees with hundreds of sequences in a matter of seconds, and it has been proved that in most cases the results are very similar to those of DDM using optimality criteria. It is also used as a step in many heuristic methods as a starting point for the search in the tree space.

I used NJ as a representative of DDM, and also as a first attempt in order to quickly find problems or difficulties in the analysis. The version used is the one included in the ARB package, and is known as ARB Neighbor Joining. The correction used was always the felsenstein one.

Maximum Parsimony (MP) (Edwards & Cavalli-Sforza, 1963) methods are among the oldest criterion-based methods. They all search for the tree that requires the minimum amount of evolutionary changes – they are thus character-based counterparts of the Minimum Evolution methods. MP theoretical background has its root in the first years of the cladistic evolutionary school and is often linked to Occam's razor. MP methods usually use exact algorithms for evaluating each tree and heuristic algorithms for searching the tree space. They actually count single base differences between sequences. Even if the "cost" of each change can be evaluated differently (with a cost matrix), MP methods aren't stochastic and don't rely on an explicit model of evolution. This is one of the many criticisms moved against the use of MP methods. Another is that they are known to be often non-consistent (they don't converge to the "true" tree adding more data). MP methods are also very prone to long-branch attraction problems, because of their incapability to take into account multiple changes at a single locus. On the other hand, they behave well in presence of different rate of evolution on different region of the same sequence. So, they can still be used safely for inferring very recent evolutionary events, like intraspecific relationships.

I've used MP methods mostly for historical reasons, knowing that for the most difficult analyses they were far inferior to stochastic methods. I've used the software Phylip DNAPARS, provided by the ARB package, always setting all parameters in order to maximise accuracy at the expense of speed.

Maximum Likelihood (ML) (Edwards & Cavalli-Sforza, 1964) methods are among the most commonly used. They are considered accurate and based on solid theoretical background. Their major drawback is the computational effort they require, especially when coupled with a

very complex evolutionary model. The criterion they use to evaluate a tree is its likelihood. The likelihood of a hypothesis (in this case, a tree) is, by definition, the probability of the data given the hypothesis. This means that for each evaluated tree, given an evolutionary model (whose parameter have to be calculated for each data set) one algorithm have to calculate the probability of the data as an outcome of the hypothesis. Mathematical details are more complex than those for other methods, and this justify the time needed for computation. ML methods have good statistical properties and, if the adopted model is correct and its parameters well estimated, are less prone to long-branch attraction problems and other artefacts. They also converge to the "true" tree when the amount of data increases.

ML methods were the core of my phylogenetic analyses. I used the PHYML software provided with the ARB package, and the evolutionary model suggested by ModelTest (or one more complex, see below) for each analysis. All parameters were empirically calculated on each analysed data set.

Quartet puzzling is indeed a category of ML methods. It is performed by the TreePuzzle software, and requires the same input of PHYML, with one exception: for complex evolutionary models, TreePuzzle is not able to calculate parameters values on its own. So they have to be inserted manually, according to the results of ModelTest (see below).

TreePuzzle (Schmidt et al., 2002) uses quartets of species. Group of four species have the interesting characteristics of being the smallest for which more than one unrooted tree exists. Any kind of analysis, even the most complex, is computationally easy on such a small data set. Moreover, the number of possible quartets of species in a dataset of N sequences grows fast with N, but is usually tractable. What TreePuzzle practically does is computing the likelihood of all possible quartets tree. Then, it performs a fixed number of independent puzzling steps (in my analysis, 100.000): every time, it adds all sequences in a randomized order to a tree, on the branch least contradicted by the quartets topology. The final tree is obtained as a consensus from all the trees obtained during the puzzling steps.

Bayesian Inference (BI) shares many analogies with ML methods. They're both character- and criterion-based methods based on stochastic models of evolution and statistical operators. But while ML aims to find

the tree with the maximum likelihood, BI actually searches for the tree with the maximum probability. For mathematical reasons, it is impossible to derive the probability of hypothesis θ (in our case, the tree with all of its parameters: topology, branch length, evolutionary model…) given a data matrix X without knowing (or guessing) the so-called prior probability P (θ), an absolute probability of the hypothesis unconstrained by the data. Bayes theorem states that:

$$\mathbf{P}(\theta|X) = \frac{\mathbf{P}(\theta)\mathbf{P}(X|\theta)}{\mathbf{P}(X)}$$

P(X|θ) is actually the likelihood of θ. P(X) is the absolute probability of X, a normalizing factor that can be calculated. But the prior probability is essential in order to obtain P(θ|X), the so-called posterior probability – the true probability of the hypothesis given the data.

The posterior probability function cannot be derived analytically. Bayesian phylogenists, in recent years, have found a way to bypass the problems. The method is called Markov chain Monte Carlo (MCMC), and is a heuristic algorithm that allows sampling the tree space in order to find the tree with the maximum posterior probability. It starts with a random set of parameters (a hypothesis θ_1) and calculate the posterior probability of the corresponding tree; then, it alters slightly one or more parameters (generating hypothesis θ_2) and calculate the new posterior probability value. If $f(\theta_2)>f(\theta_1)$ the new set is kept, and sometimes it is kept also if $f(\theta_2)<f(\theta_1)$, according to precise rules that try to avoid the stacking in some local maximum. After many iterations (called "generations") the posterior probability distribution can be assessed, thanks to the intensive sampling in a region around a presumed maximum. The final tree is accordingly derived.

The problem of prior probability functions remains. Indeed, those have to be determined by the researcher before starting the analysis. The most common approach is to set all kinds of prior probability distributions to be uninformative. This means that they consist of functions that give no particular emphasis on any hypothesis with respect to any other. But BI software always allows changing this approach in a variety of ways: the researcher can thus put more emphasis on some kinds of parameters simply choosing the appropriate probability distribution. For example, it is

possible to make some particular node in the topology more likely to appear, or render lower value of the α parameter of distribution Γ more probable, and so on… Bayesian phylogenists think that this versatility is one of the strength of BI. Critics focus on the same point, because using different prior probabilities distributions may change the final output of the analysis, and there are no universal rules to help with the decision. So, BI suffers from some kind of subjectivity. It is interesting to note that the argument between Bayesian and non-Bayesian statisticians is an old one, and has nothing to do with phylogenetic methods themselves or any specific biological question.

I've used BI mostly because it is one of the currently most used and requested methods for inferring phylogenies. I used MrBayes (Huelsenbeck & Ronquist, 2001), by far the most known software. I've set the evolutionary models in accordance to those indicated by ModelTest and the prior probability distributions as uninformative as possible. I ran the MCMC analysis for 1.000.000 generations and eventually excluded the first 25% outputs as an initial burn-in, far away from the desired region of the tree space. MrBayes can run more than one analyses ("chain") independently and in parallel. They start from different initial points, so not only this options gives a more intensive sampling, but also a statistical tool to evaluate convergence and problems of local maxima. I've always run 3 chains for every analysis. Moreover, for every so-called cold chain I've used the Metropolis Coupling. This consists in running also some (in my case, 3) heated chains, which sample a flattened posterior probability distribution in which it is more likely to jump from one peak to another. At regular intervals, parameter status sets running in the heated chains are evaluated by the cold chain and, if accepted, can allow the cold chain to escape from a local peak.

2.4.6. Hypotheses testing: the data matrix

After the alignment step is complete, the data matrix obtained and before starting the analysis itself, many decisions have to be made. One is of course the choice of the method but, as said above, it is generally preferable to try many of them and confront the results. Another is choosing the appropriate evolutionary model. Even more important, it would be better to evaluate if the data matrix is properly fit for the

phylogenetic analysis, or if the sequences contain too much noise and ambiguities.

For all these questions statistical tests have been developed in recent years. I've used two of them: Likelihood mapping and ModelTest.

Likelihood mapping (Fig. 2.1) is a function of TreePuzzle, and is one of the possible ways to evaluate the amount of evolutionary information brought by the data matrix. This can be lacking for different reasons: bad alignment, low quality of the sequences, poor sampling or choosing of the available data.

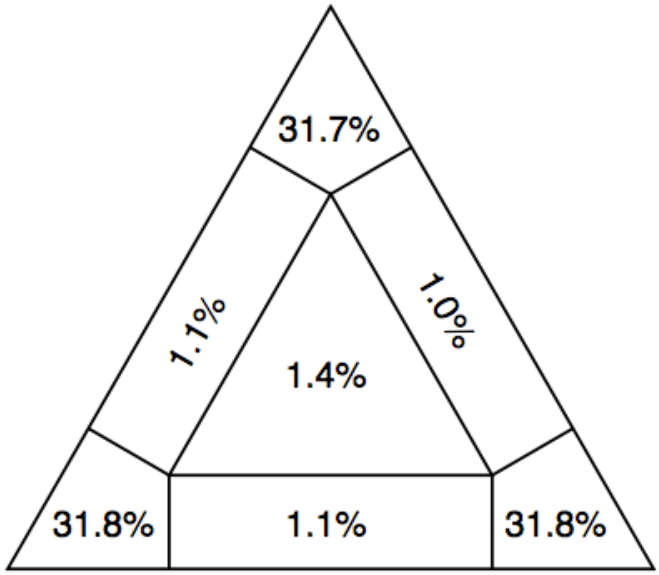

Fig 2.1. Output of Likelihood mapping using TreePuzzle. Sum of the values in the corner areas give the percentage of resolved quartet tree topologies.

Likelihood mapping is concettually simple and performs one of the steps already seen in building a quartet-puzzling tree. It computes the likelihoods value for all (or a subset of all) possible quartet trees. Because each quartet has 3 possible unrooted trees, it is easy to plot the results of the analyses on a triangle graphic. Each dot corresponds to a quartet. Dots near one of the vertices correspond to very strongly supported topologies. Dots near the sides or the centre correspond to tangled or unsupported topologies. The percentage of dots in the three regions adjacent to the vertices is an estimation of the good information contained in the data

matrix. Values below 80% suggest a very strong noise, and make futile any kind of analysis unless the data matrix is somehow modified (excluding bad or too similar sequencing, adding information…).

ModelTest (Posada & Crandall, 1998) is a free software that helps making decision about the choice of the correct evolutionary model. One naive option is to always use the most complex and parameters-rich model. But more complexity and more parameters to be evaluated means much more time requested. More than that, parameters value estimation is a process that by itself can add errors to the analysis. Thus, it is preferable to use a more complex model only when it gives a significantly better explanation of the data than a simpler one.

ModelTest uses the data matrix to calculate the parameters value and likelihoods of 88 different models. The likelihood, as said above, is the probability of the data given the hypothesis (in this case, of a given alignment given the model). Once the calculations are over, it is possible to choose a statistical test (AIC, BIC, hLRT) that gives the most complex model with a likelihood value statistically higher than simpler ones. The result can be different with different test and, moreover, not all 88 models can be easily included in all software. My strategy was to pick the next more complex model available in all the software I used (with the exception of ARB NJ, that can exploit as the most complex model the felsenstein one).

2.4.7. Statistical support to tree topology

Once the analysis is over and a tree is produced, it is difficult to say which node, if any, of the tree topology is reliable and which isn't. Some clades can be found also with different methods and nonetheless be prone to confutations. This can be seen when adding or subtracting certain sequences or excluding some portion of the sequences leads to a different topology. To find these less reliable nodes, some kind of statistical support has to be provided.

Quartet puzzling and Bayesian Inference have their intrinsic methods for providing statistical support to each node. TreePuzzle produce what is called a majority-rule consensus tree from a comparison between all trees found in the puzzling steps. A percentage is associated with each node,

and indicates the fraction of different trees that possessed that node. Percentages below 50% lead to a polytomy.

All recent BI methods use the MCMC, which samples extensively from the space tree region of interest. Thus, they can deduce the posterior probability distribution for all parameters involved in the analysis, nodes and branch lengths included. In literature, it is common to find a number simply labelled as "Posterior Probability" near each node.

For NJ, MP and ML methods there are no such shortcuts. One of the most commonly used method, and the one I chose, is called non-parametric bootstrap. Bootstrapping is a statistical method, and has no relationship whatsoever with the biological framework of the study. It consists in running N times the same analysis on a modified version of the data matrix called a pseudoreplicate. A pseudoreplicate is generated this way: given that the original analysis was performed on a matrix of n characters (columns), for each iteration a new matrix of n characters is obtained sampling randomly n times, with reintroductions, the character set of the original one. Some columns may be picked more than once and others can be end up being excluded from the analysis. The bootstrap value associated with each node in the initial tree is equal to the percentage of pseudoreplicate trees that had the same node. Because the tree shown is the one obtained from the original data matrix, and not a majority-rule consensus tree, values below 50% can be produced.

Non-parametric bootstrapping is another very controversial step in the field of molecular phylogenesis inference. Everybody uses it, but no one can actually tell what is its exact meaning and how the value should be interpreted. There is no official threshold below which a node has to be considered unreliable, even if usually values higher than 70% are considered good. From a practical point of view, high bootstrap values prove that, even if the data get slightly modified, that node is still recovered, and is thus a good hypothesis (there are however artefacts that can bring high statistical support to wrong topologies). This is a good reassurance against small doubts on the alignment.

In any case, the number of pseudoreplicates has to be adequate in order to give a serious support. Once more, there is a conflict between accuracy and rapidity, because the analysis time increases linearly with the number of iteration. 200-1.000 is usually considered a good interval, and I

always used 500 pseudoreplicates in MP and ML trees and 1.000 in NJ trees.

2.4.8. Drawing the tree

A phylogenetic tree is the conclusion of a long work, a scientific hypothesis and in some case, the final result of a project. Thus, it has to be presented in a correct but also appealing way. This means that it should be readable and comprehensive of all data. To accomplish this goal, many problems are of aesthetic nature. The dimensions of the picture, the groups of taxa that should be merged, the labels on terminal branches and the fonts, dimensions and positions of all words and numbers should be carefully chosen.

Moreover, like for all hypotheses and results, the eye of the researcher filters phylogenetic trees, too. Colours, boxed regions, bold letters and whatever are sometimes useful to highlight some important feature, but always reflect the personal choices of the author.

Aside from the aesthetic work, I always faced some decisions about how to present the phylogenetic results in a clear way. I thought, for example, that showing all trees produced for every single analysis would be confusing and redundant. So I chose to present only the ML tree for every analysis. Given that in most cases the well supported nodes were the same in all kinds of analysis (with the important exceptions being described in the text), to summarize them I simply wrote a set of values for each node. The values represent the statistical support (bootstrap, posterior probability or quartet puzzling) for NJ, MP, ML, quartet puzzling and BI trees respectively.

The second choice I've made was to cut the values of bootstrap and quartet puzzling lower than 70%, and posterior probability values lower than 0.80. These are of course arbitrarily chosen thresholds. I've done this choice for reasons of clearness. Low-supported nodes with all statistical values lower than these thresholds are unreliable anyway, and leaving them without labels renders the tree easier to read.

3. *Paramecium jenningsi* and *Gortzia*

3.1. Background

Paramecium jenningsi is a species of the *aurelia* subgroup of genus *Paramecium*, morphologically similar to *P. aurelia* itself but more closely related, according to SSU rRNA phylogeny, to the recently discovered *P. schewiakoffi* (Fokin et al., 2004). It is not among the most studied species of *Paramecium*, and the literature on it is quite scattered. Several authors argues that it is a complex of at least two sibling species (e.g. Maciejewska, 2007), one from Japan and the other from mainland Asia. No report of endosymbiosis inside *P. jenningsi* cell has been yet published but, given its closeness to many other species that commonly harbour bacteria (e.g. *P. aurelia* and *P. caudatum*) this absence is more probably due to lack of studies than to an actual peculiarity.

The ciliates studied on this first project belong to two coexisting populations of *P. jenningsi* and *P. aurelia* (in most of what follows, I will use the species complex name because the biological species identification is still underway). Along with other elements of interest that will be detailed below, it is important to notice that the original sample came from Thailand. Tropical and equatorial countries are much less studied than temperate ones by ciliatologists exclusively for practical reasons. Data coming from such ecosystems are always welcome to test generalizations of previous hypotheses and development of new ones.

Bacteria belonging to genus *Holospora* (*Holosporaceae*, *Rickettsiales*, *Alphaproteobacteria*) are probably the most studied among the ciliate endosymbionts (Fokin & Görtz, 2009; Fujishima, 2009). They were the first to be discovered (Hafkine, 1890), and among the first to be re-discovered in the second wave of studies on symbiosis in the ‘70s (Gromov & Ossipov,

1981; Preer & Preer, 1982). One of the reasons is probably their conspicuousness and their peculiar morphology. Another one – and related – reason is that they are the only ciliate endosymbionts for which there are extensive proofs of horizontal transmission between hosts.

The term *Holospora*-like bacteria (HLB) is commonly employed. HLB are usually defined as alphaproteobacteria inhabiting the macro- or micronuclei of a *Paramecium* cell, capable of horizontal transmission and differentiation of a morphologically distinguishable infectious form (IF) (e.g. Görtz & Dieckmann, 1980; Fig. 3.1). Their exclusive presence in *Paramecium* is now challenged, because very similar bacteria (from a morphological, life-cycle and systematic point of view) have been found in *Frontonia* and other ciliate genera (Fokin, unpublished data).

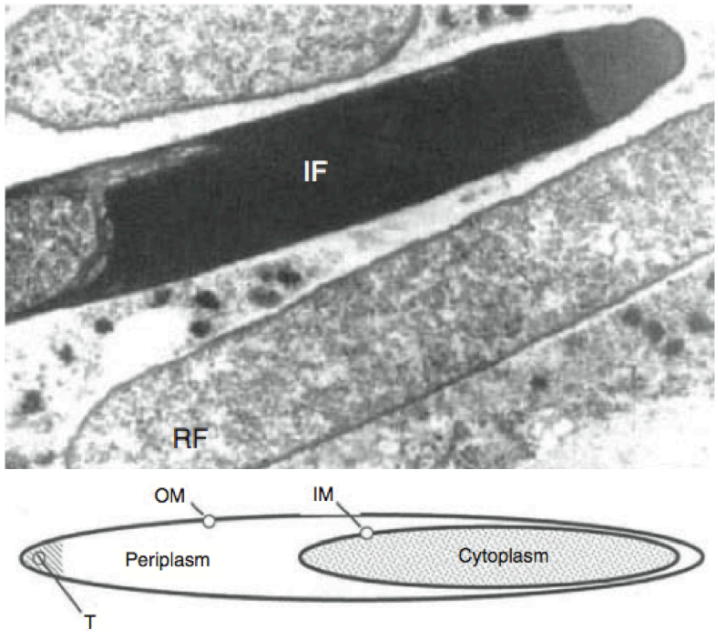

Fig 3.1. Ultrastructure (above) and schematic drawing (below) of *Holospora obtusa* Infectious Form (IF). From Görtz, 2006. Infectious Form (IF), Reproducing Form (RF), Infectious Tip (T), Outer Membrane (OM), Inner Membrane (IM).

HLB life cycle is shown in figure 3.2. Vegetative forms grow by binary division in the macro- or micronucleus of the hosts. When the ciliate cell divides, more or less half of the bacterial cells end in each of the two daughter nuclei. Under some circumstances, IF are produced. These much elongated, electron-dense cells develop a huge periplasmic space, with a

differentiated end full of specific proteins, called the infectious tip. In some *Holospora* species (like *H. obtusa* and *H. undulata*) the IF are collected in a portion of the dividing nucleus, called the connecting piece. This vesicle is segregated in one of the daughter cells, and eventually fuses with plasmalemma and release IF on the outside medium.

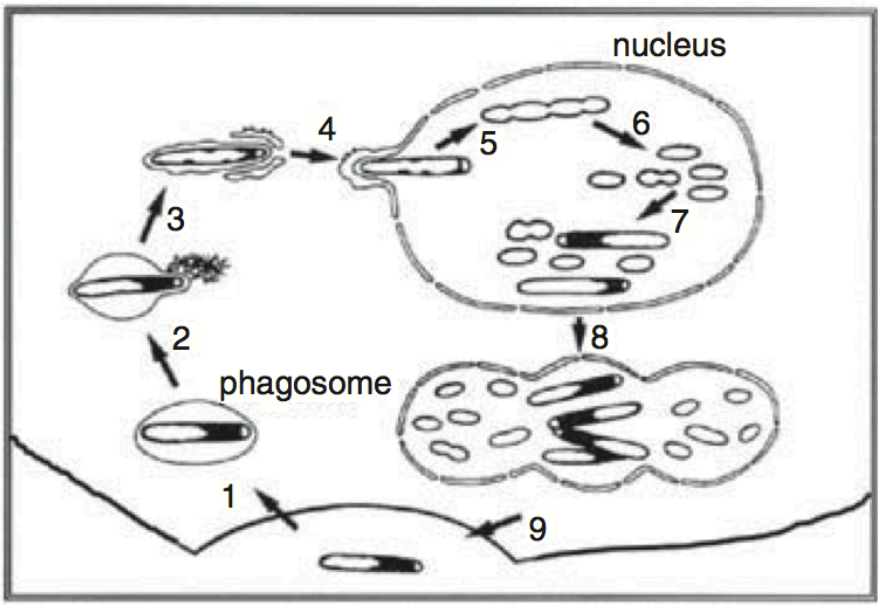

Fig 3.2. Infection cycle of *Holospora obtusa*. A connecting piece in the dividing macronucleus can be identified. Description provided in the text. From Görtz, 2006.

IF are capable of surviving in the environment outside their host for a certain amount of time, and are responsible for the infection of new hosts. Infection process in *Paramecium* starts when the ciliate ingests IF of HLB through phagocytosis at the cytostome. IF, usually with other food bacteria, are collected in a phagosome where digestion processes start. But in the acidification phase, lowering of pH triggers the escape of IF – infectious tip first – from the digestive compartment. It is not clear if the bacterium is then surrounded by a eukaryotic membrane derived from phagosome membrane or if it is free in the cytoplasm and gets eventually surrounded by an ER-derived membrane. In any case, the IF travels to the target nucleus, probably aided by a structure involving host's actin filaments, and contacts the nuclear envelope with the infectious tip (Fujishima, 2009).

After the entrance in the nucleus and elimination of any host's membrane, the IF divides through multiple scissions and originates a number of vegetative forms, thus closing the cycle. However, if the bacteria end up in a nontarget nucleus or host, they are actively destroyed or expelled by the ciliate (e.g. Skovorodkin et al., 2001)

HLB are intensively studied especially from the cytological/ biochemical (mostly in Prof. Masahiro Fujishima's laboratory) and ecological (in Prof. Oliver Kaltz's lab) point of view. These studies concentrate on a few well known species, like *Holospora obtusa*, *H. elegans* e *H. undulata* (e.g. Fujishima et al., 2005; Fels & Kaltz, 2006). In parallel, other surveys have revealed the great diversity of HLB, especially those supervised by Prof. Sergei Fokin and Prof. Hans-Dieter Görtz.

The systematic and taxonomy of HLB are tangled and messy fields. The main reason is that, although many papers are present in literature highlighting the diversity of these bacteria, most of them are not supported by careful molecular analyses. This has led to an unreliable, according to modern standard, taxonomic system (Görtz & Schmidt, 2005).

There are nine described species of HLB (Hafkine, 1890; Gromov & Ossipov, 1981; Fokin, 1991; Preer & Preer, 1982; Fokin, 1989; Fokin & Sabaneyeva, 1993; Tab. 3.1) all currently assigned to the genus *Holospora*. Five of these were not described properly according to bacterial nomenclature, and are considered invalid. *Holospora* species are discriminated according to a pattern of 4 characters: morphology (form and size of the IF), host specificity, target nucleus and presence/absence of the connecting piece. This set of characters is sufficient to discriminate among known species, but isn't enough for building a reliable phylogenetic tree. In addition, the recent discovery of "unusual" HLB, like those found in non-*Paramecium* hosts, can cast even more doubts on any attempt to provide an unambiguous systematic framework.

Only the SSU rRNA sequences of *H. obtusa* (Amann et al., 1990) and a very small fragment of the SSU rRNA sequence of *H. elegans* (Hori et al., 2008) are publicly available. To unravel the problem of HLB phylogeny, two paths have to be followed: first, it would be important to obtain strains of known *Holospora* species – preferably more than one – and characterize them from a molecular point of view (e.g. obtaining their SSU rRNA gene sequence); second, it is recommended that all studies

conducted on HLB diversity from now on include a multi-disciplinary approach, with molecular characterization alongside classical morphological and biological analyses. All additional data could then be compared in a unifying phylogenetic analysis on molecular sequences, supported by other kinds of characters. It is important to notice that some organisms found in the past and poorly described could never be collected and described again – at least, not with the certainty of handling the same biological object.

The project that follows is organized in accordance with the rules just listed. It is a characterization-oriented study on a ciliate-HLB system performed through use of the multi-marker approach. The original aims were to describe the organisms involved and, if possible, use the data obtained together with other already published in order to draw more general conclusions on the systematics of HLB.

Species	IF	Hosts	Target nucleus	Connecting piece
H. undulata	Spiral, ends tapered	*P. caudatum*	Micronucleus	Present
H. elegans	Straight, ends tapered	*P. caudatum*	Micronucleus	Present
H. caryophila	Spiral, ends tapered	*P. biaurelia*, *P. caudatum*, *P. novaurelia*	Macronucleus	Absent
H. obtusa	Straight, ends rounded	*P. caudatum*	Macronucleus	Present
"H. acuminata"	Straight, ends tapered	*P. bursaria*	Micronucleus	Present
"H. recta"	Straight, one end tapered	*P. caudatum*	Micronucleus	Present
"H. curviuscula"	Curved, ends tapered	*P. bursaria*	Macronucleus	-
"H. bacillata"	Straight, ends rounded	*P. calkinsi*, *P. woodruffi*	Macronucleus	Absent
"H. curvata"	Curved	*P. calkinsi*	Macronucleus	Absent

Tab 3.1. Diagnostic table of HLB. Modified from Görtz & Schmidt, 2005)

3.2. The system

The organisms studied in this project come from a sample taken on September 2010 in Thailand, from the freshwater Chaweng Lake on Samui Island.

Prof. Sergei Fokin screened the sample during the month of October 2010. He found at least two coexisting populations belonging to different morphospecies of genus *Paramecium*. According to morphological features, one could be assigned to the *P. aurelia* complex of sibling species and the other to the morphospecies *P. jenningsi*. Both of them were inspected under the histological microscope and contained HLB in their macronuclei (Fig 3.3).

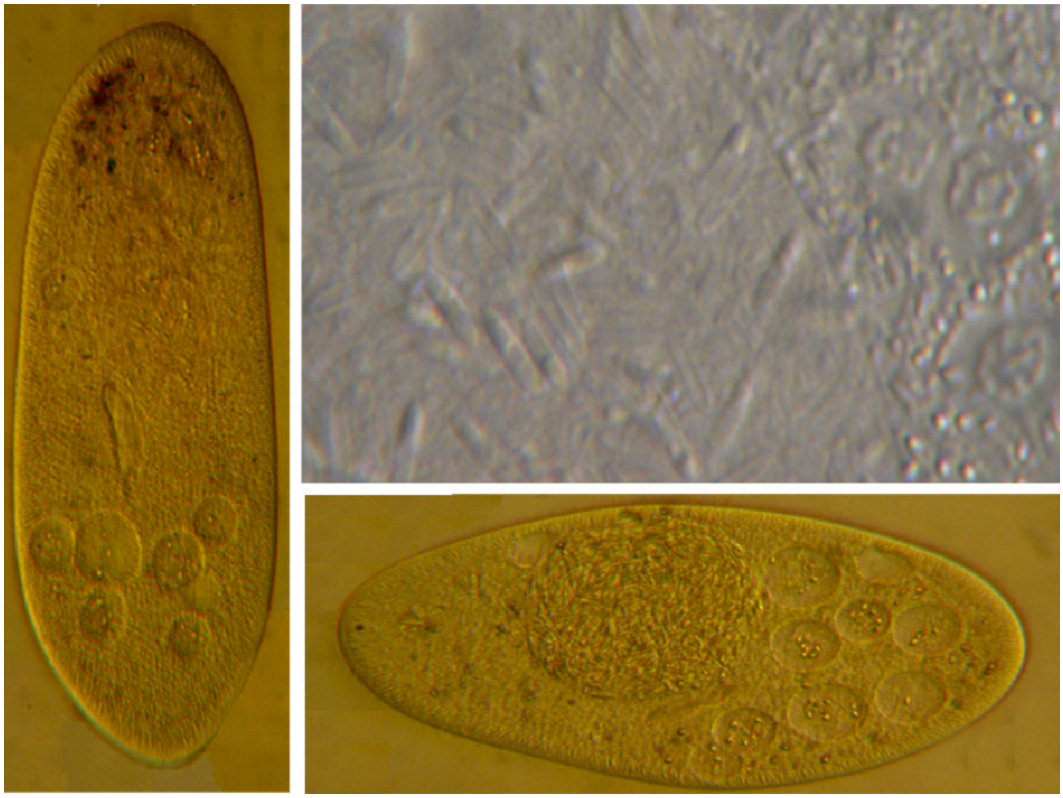

Fig 3.3. Pictures of the *Paramecium jenningsi* population. From left in clockwise order: a paramecium cell without evident infection; detail of the two micronuclei and the single macronucleus, inside which bacteria are visible; whole cell heavily infected. Photographs by Sergei Fokin.

Using dimensions as a discriminant, Prof. Fokin was able to separate the two populations and culturing both, starting with a few cells. All results were thus obtained on populations, and not on clonal cultures.

Before starting to feed the cultures with bacteria and transferring them in a new medium, about 50 infected cells picked from each morphospecies were fixed in ethanol 70% and stored at -22°C. Few days after the populations were separated, the *P. aurelia* cells lost all their macronuclear endosymbionts. A second isolation of cells for DNA extraction was performed some months later, to obtain more material.

While I was performing the molecular characterization, Prof. Fokin surveyed some aspects of the morphology, life cycle and host specificity of the HLB. Vegetative forms are 1-3 μm long and IF are straight, with rounded ends and 4-7 μl long, well in the usual range of classical HLB. The connecting piece is not induced in the dividing host.

Bacterized medium obtained from crushed *Paramecium* cells were used for cross-infection experiments. Strains of *P. schewiakoffi*, *P. caudatum* and *P. sonneborni* and aposymbiotic cells from the Thai *P. jenningsi* and *P. aurelia* populations were used as target. The HLB could enter in the macronucleus of all of them, but only in *P. jenningsi* the infection could be maintained.

Moreover, some cells of *P. aurelia* were sent to Prof. Ewa Przybos from Krakow in order to perform mating experiments and identify the correct biological species. According to the preliminary results, Thai paramecia belong to *P. tredecaurelia*.

3.3. Results

3.3.1. Molecular characterization of the hosts

From cells of both populations stored in ethanol, total genomic DNA was extracted using the NucleoSpin Plant II kit, and resuspended in 50 μl of distilled and autoclaved water. The solutions were stored at -22°C and thawed at 0°C for all subsequent uses.

Partial SSU rRNA sequences were obtained from *P. jenningsi* and *P. aurelia* tgDNA through PCR with primers 18S F9 and 18S R1513 Hypo performing 35 cycle at 94°C (30''), 50°C (30'') & 72°C (120''). PCR products were directly sequenced with three internal primers (18S R536, 18S F783 and 18S R1052) and gave unambiguous sequences.

The SSU rRNA sequence recovered from *P. jenningsi* is 1712 bp long and in analysis with the BLAST software resulted more similar to AF100311, attributed to *Paramecium jenningsi* partial 18S rRNA gene sequence (similarity 99,9%, 1650/1652 identities, no gaps; Strüder-Kypke et al., 2000).

The SSU rRNA sequence recovered from *P. aurelia* is 1711 bp long and is more similar to AF100315 attributed to *Paramecium primaurelia* partial 18S rRNA gene sequence (similarity 99,9%, 1649/1651 identities, no gaps; Strüder-Kypke et al., 2000).

For *P. jenningsi*, also ITS1+5,8S+ITS2 sequence was obtained, using PCR and direct sequencing [amplification primers: 18S F919 and RGD2; thermal pattern: 35x 94°C (30''), 50°C (30'') & 72°C (180''); sequencing primers: FG1400 and RGD2]. Electropherograms quality was very good.

The ITS1+5,8S+ITS2 sequence recovered from *P. jenningsi* is 1087 bp long and is identical to JF304167 attributed to the homologous sequence of *Paramecium sonneborni* strain ATCC 30995 (similarity 100%, 1064/1064 identities, no gaps; Tarcz et al., unpublished). The maximal similarity to another *Paramecium jenningsi* homologous sequence is to JF304171.1 attributed to strain SA (similarity 99,2%, 1055/1064 identities, no gaps; Tarcz et al., unpublished).

cox1 gene sequence was obtained twice for *P. jenningsi*. The first time, a PCR was performed [primers: cox1fT7_Cil_IMVFF and cox1rM13_Cil_HDTF; thermal pattern: 5x 94°C (30''), 45°C (30'') & 72°C (150''), 25x 94°C (30''), 50°C (30'') & 72°C (150''), 10x 94°C (30''), 45°C (30'') & 72°C (150'')]; the product was then diluted 1:100 with distilled water and used as a template for a second PCR amplification [primers: T7 and M13R; 25x 94°C (30''), 50°C (30'') & 72°C (180'')], whose product was directly sequenced [sequencing primers: M13R and cox1_Rs_Olig_AGWT]. The second time, one PCR and direct sequencing were sufficient [amplification primers: cox1fT7_DVA(FY) and cox1rM13_Cil_HDTF; thermal pattern: 5x 94°C (30''), 45°C (30'') &

72°C (180''), 25x 94°C (30''), 50°C (30'') & 72°C (180''), 15x 94°C (30''), 45°C (30'') & 72°C (180''); sequencing primers: FG1400 and M13R and cox1_Rs_Olig_AGWT]. Both results were good and identical in the overlapping portions, so the sequences were assembled together.

The partial *cox1* sequence recovered from *P. jenningsi* is 1169 bp long. Even though the maximal similarity is shared with EU086108 attributed to *Paramecium dodecaurelia* strain 246 *cox1* gene (similarity 86,4%, 367/425 identities, no gaps; Przybos et al., 2008), the query coverage is only about 36%. A more reliable result is JF304188 attributed to *Paramecium jenningsi* strain SA *cox1* gene (similarity 85,1%, 610/717 identities, no gaps; Tarcz et al., unpublished), whose query coverage is about 61%.

For *P. aurelia*, a partial *cox1* gene sequence was obtained performing a first PCR [primers: cox1_F543_SKmP and cox1_R1321_SKmP; thermal pattern: 20x 94°C (30''), 45°C (45'') & 72°C (60'') for each single primer, then 2x 94°C (30''), 45°C (45'') & 72°C (60''), 35x 94°C (30''), 55°C (45'') & 72°C (60'') with both primers]; the product was then diluted 1:100 with distilled water and used as a template for a second PCR amplification [primers: T7 and M13R; 30x 94°C (30''), 50°C (30'') & 72°C (90'')], whose products were directly sequenced [sequencing primers: M13R and T7]. The quality of electropherograms was very good.

The partial *cox1* sequence recovered from *P. aurelia* is 761 bp long. The most similar sequence is FJ003706 attributed to *Paramecium novaurelia* strain V9-6 *cox1* gene (similarity 85,2%, 421/494 identities, no gaps; Catania et al., 2009). A similarity of 82,7% (401/485 identities, no gaps) is obtained from the comparison with FJ003711 attributed to *Paramecium tredecaurelia* strain 209 (Catania et al., 2009).

3.3.2. Molecular characterization of the HLB

First attempts to obtain the SSU rRNA gene sequence of the HLB from *P. jenningsi* through PCR and direct sequencing failed. Thus, a cloning was performed on PCR products [primers: 16S Alphα F19a and 1492R; thermal pattern: 35x 94°C (30''), 50°C (30'') & 72°C (120'')].

60 clones positive to the control PCR were screened with RFLP. 4 patterns were represented by more than 1 clone: A (15 clones, 25%), B (6

clones, 10%), C (4 clones, 6,7%) and D (2 clones, 3,3%). 3 clones for each pattern were chosen to represent patterns A, B and C, and 1 to represent pattern D. Pattern A corresponded to the sequence of interest; the 1399 bp consensus sequence obtained had a similarity of 90,5% with that of *Holospora obtusa* 16S rRNA gene (accession number: Amann et al., 1991).

After that, another attempt with PCR and direct sequencing succeeded [amplification primers: 16S Alpha F19a and 16S R1488 Holo; thermal pattern: 35x 94°C (30''), 57°C (30'') & 72°C (120''); sequencing primers: 16S R515ND, 16S F343ND and 16S F785ND]. The sequence quality was good, and the obtained sequence was identical to the consensus one, albeit shorter, and was not utilised in subsequent analyses.

The approach to obtain the SSU rRNA gene sequence of the HLB originally infecting *P. aurelia* was the same. PCR product [primers: 16S Alpha F19a and 16S R1488 Holo; thermal pattern: 35x 94°C (30''), 57°C (30'') & 72°C (120'')] was cloned. 19 clones were screened with RFLP method. Patterns A (6 clones, 31,6%), B (2 clones, 10,5%) and C (2 clones, 10,5%) were identified. 3 clones were chosen to represent pattern A and 1 from every patterns to represent B and C. Pattern A corresponded to the sequence of interest. Indeed, the 1398 bp consensus sequence obtained was identical to the homologous part of SSU rRNA gene sequence of HLB from *P. jenningsi*.

3.3.3. FISH experiments

Because *P. aurelia* cells lost the endosymbionts early in the work, FISH experiments were performed only on *P. jenningsi* cells. These were put to starvation in sterilized medium for a week, washed and fixed on slides using formaldehyde (4% in PBS buffer) or osmium tetraoxide (4% in H_2O) vapours at least one day before the experiments.

From the SSU rRNA sequence obtained, I could observe that no existing HLB probes perfectly matched the bacteria from *P. jenningsi*. Thus, I designed a new probe, called Gortzia659, labelled with Cy3 at the 5' end.

In one experiment, I used Gortzia659 and Eub338, a general probe that matches most of *Bacteria*, labelled with fluorescein. Both probes gave positive results, binding to HLB reproductive form-like and infectious

form-like organisms inside the paramecia's macronuclei (identified by DAPI staining). Picture overlapping suggests that there are no other bacteria inside the ciliate cells (but it is not possible to exclude completely their presence in the macronucleus because of the great number of HLB cells providing much signal noise). All observed *P. jenningsi* cells (>20) were infected by at least a single bacterium. In the majority of cases the bacterial cells were too numerous to be exactly counted, but well over one hundred, with an approximate IF/VF observed ratio of 1:5.

Very similar results were obtained with the probe pairs Gortzia659 and ALF1b, a general probe for *Alphaproteobacteria* labelled with Alexafluo488 dye at the 5' end (Fig 3.4).

The probe Gortzia659 was also used in a FISH experiment on *Paramecium biaurelia* strain FGC3, kindly supplied by Valerio Vitali, whose macronucleus was infected by *Holospora caryophila*. The probe gave no signal, as expected by the 4 mismatches between the probe and the *H. caryophila* sequence (data not shown).

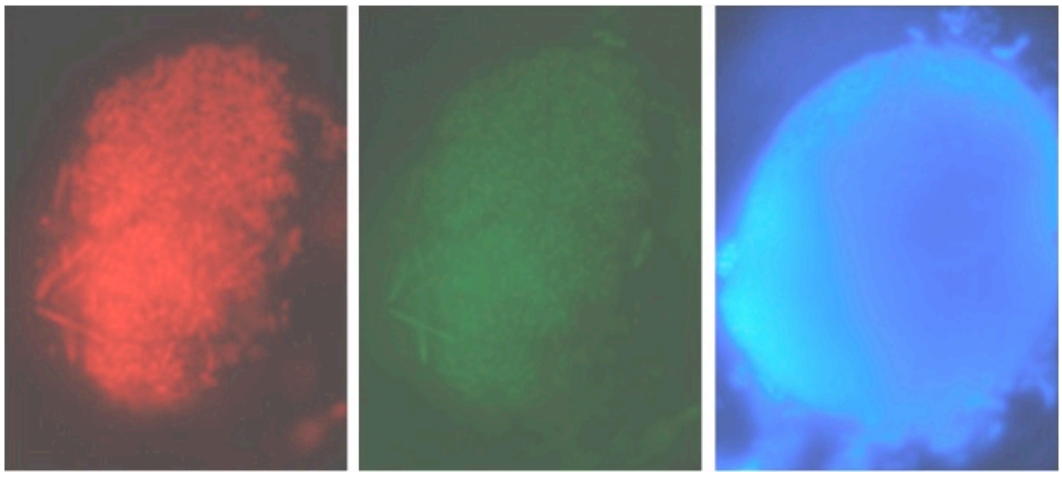

Fig 3.4. Epifluorescence microscope pictures of the macronucleus of a *P. jenningsi* cell. From the left: signal of the Gortzia659 probe (red), signal of the ALF1b probe (green), signal of the DAPI staining (blue).

3.3.4. Phylogenetic analysis of the bacterial sequences

A phylogenetic analysis was performed on both SSU rRNA sequences in order to assess their evolutionary position inside order *Rickettsiales* (Fig 3.5). In order to accomplish this goal, a total of 56 sequences were included in the analysis. Representing HLB I used my two sequences, the one available from *H. obtusa* plus three other previously obtained by other students in our lab and still unpublished: one from *H. caryophila* (obtained by Martina Schrallhammer and Stefano Galati), one from *H. undulata* and a third from an HLB found in the macronucleus of *Frontonia* (both obtained by Filippo Ferrantini). The sequence of "*Candidatus* Paraholospora nucleivisitans" was added; the affiliation of this organism to family *Holosporaceae* is not formalized (its biology is indeed quite different). The *Caedibacter*-like group is represented by 4 sequences (three published and one recently obtained from a putative *Caedibacter*-like bacterium found in *Euplotes harpa* by Claudia Vannini). "*Candidatus* Odyssella thessalonicensis" and "*Candidatus* Captivus acidiprotistae" sequences are added as other symbiotic members of order *Rickettsiales*, like a recently found endosymbiont from the *Euplotes* population EMP. Family *Rickettsiaceae* is represented by 11 sequences, 8 already published and belonging to genera *Rickettsia*, *Orientia* and "*Candidatus* Cryptoprodotis polytropus" and 3 others from RLO characterized in our lab and still unpublished. Family *Anaplasmataceae* is represented by 5 sequences from genera *Anaplasma*, *Ehrlichia*, *Wolbachia* and *Neorickettsia*. Sequences from "*Candidatus* Cyrtobacter comes", "*Candidatus* Midichloria mitochondri", "*Candidatus* Anadelfobacter veles" are known to be associated in a family-like clade (Vannini et al., 2010). 11 sequences were chosen from other orders of class *Alphaproteobacteria* as outgroup. 12 uncultured sequences were included because of their similarity with the target sequences or because they come from promising organisms; 6 of them came from symbionts of the amoeba *Acanthamoeba*.

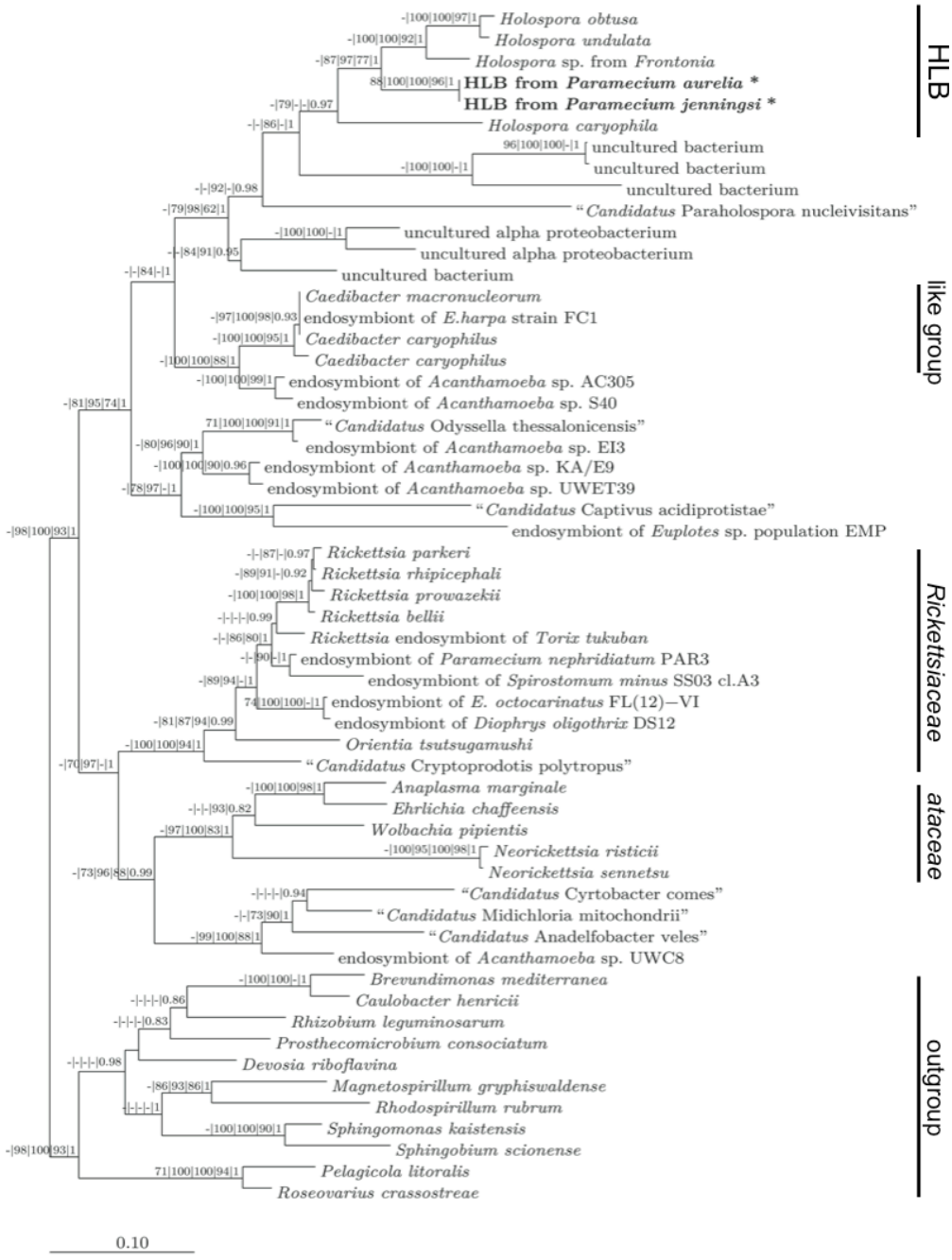

Fig 3.5. SSU rRNA gene sequences Maximum Likelihood tree. 56 sequences and 1257 characters were employed. The node values represent NJ, MP and ML bootstrap values, TreePuzzle quartet puzzling support and Posterior Probability respectively, cut as explained in section 2.4.8. Asterisks mark sequences obtained in this work. Supplementary information are provided in the text.

All sequences were automatically and then manually aligned using the ARB software package. Because of long inserts in some sequences and ambiguity in the alignment of some other regions, I trimmed the aligned data matrix with a filter at 10%. The analyses were thus conducted on 1257 characters. AIC and BIC tests, performed with the ModelTest software, both gave GTR+I+G (the most general time-reversible model) as the preferred evolutionary model output. Likelihood mapping analysis performed with TreePuzzle showed that 95,2% of the 100.000 sampled quartet trees were in the resolved regions, thus suggesting that the data-set was adequate for phylogenetic analyses.

ARB NJ, PHYLIP DNAPARS, PHYML (with the tools provided by the ARB package), TreePuzzle and MrBayes were used for producing NJ, MP, ML, quartet puzzling and Bayesian tree respectively. GTR+I+G was the model used in all model-based methods, with the exception of ARB NJ, in which the felsenstein (Felsenstein, 1981) correction was used instead. 1000 bootstrap pseudoreplicates were performed for ARB NJ, 500 for MP and ML trees. 100.000 puzzling step were performed in order to obtain the quartet puzzling support values, and posterior probability values are derived from 1.000.000 generations (burnin 25%) for each of the 3 runs of MrBayes (each consisting of 3 heated and one cold chain).

Most of the nodes in the trees are supported by high support values. Monophyly of order *Rickettsiales* is confirmed. Highly supported are also the clades of *Anaplasmataceae* and *Rickettsiaceae*, those of *Caedibacter*-like sequences and two family-like clades including symbionts: one with "*Candidatus* Odyssella thessalonicensis" and "*Candidatus* Captivus acidiprotistae" and another with "*Candidatus* Cirtobacter comes", "*Candidatus* Midichloria mitochondri" and "*Candidatus* Anadelfobacter veles". The *Acanthamoeba* endosymbionts are scattered throughout the entire tree, but none of them is strictly associated to the HLB sequences nor cluster inside families *Rickettsiaceae* or *Anaplasmataceae* as currently defined.

H. obtusa and *H. undulata* are sister group in this analysis. The HLB from *Frontonia* cluster with their clade, while the HLB from this study and *Holospora caryophila* are more basal. 6 sequences from uncultured bacteria are associated to the larger clade including HLB and "*Candidatus* Paraholospora nucleivisitans", 3 of them nested inside the clade and the other 3 in a basal position.

HLB, *Caedibacter*-like bacteria and the "Odyssella-Captivus" clade form one of the two major (and well-supported) clades in the order. The second one includes *Rickettsiaceae*, *Anaplasmataceae* and the "Midichloria"-clade.

3.4. Discussion

The molecular identification of the hosts is only partially in accordance with expectations. *P. jenningsi* SSU rRNA and *cox1* gene sequences are indeed similar to other *P. jenningsi* sequences already published, but not identical. There are 2 base differences with the SSU rRNA gene sequence of strain SA, while the *cox1* similarity is very low, set at 85,1%. Moreover, the ITS1+5.8S+ITS2 sequence is more similar (indeed, identical) to that of *P. sonneborni* strain ATCC 30995 than to that of strain SA of *P. jenningsi*.

It was already showed, at least for the species of the *aurelia* complex, that different markers could give contrasting results (Catania et al., 2009). Sometimes, even using many of them together the boundaries and relationships between biological species are blurred or ambiguous. *P. sonneborni* is considered the 15th species of that complex, and *P. jenningsi* is a close relative too, thus raising the suspicion that those problems are more spread than thought.

A simpler explanation, of course, is that ITS1+5.8S+ITS2 isn't a good marker for these organisms; its 100% value of similarity with that of a *P. sonneborrni* strain is, however, an oddity. A sample of the very same strain was asked to Prof. Ewa Przybos, who very kindly sent us some cells. We could then ascertain that there was no misidentification of the morphospecies. Strain ATCC 30995 possesses all the morphological features typical of *P. aurelia* complex, and can be distinguished from our Thai *P. jenningsi* population. Moreover, dr. Tarcz kindly showed me the not yet submitted SSU rRNA gene sequence of that *sonneborni* strain, and the similarity value with the homologous sequence of the *P. jenningsi* population is about 91.4%, pretty low.

cox1 sequences great variance can be ought to an intraspecific molecular variability inside the morphospecies *P. jenningsi*, that was on the other hand already proved for histon H4 sequences (Maciejewska 2006). It could also be a further prove of the presence of a sibling species complex.

About *P. aurelia*, that was preliminary identified as a member of the biological species *P. tredecaurelia* (Przybos, preliminary data), the similarity data are more ambiguus. *cox1* gene sequence is more similar to another attributed to *P. primaurelia* than to the one of *P. tredecaurelia* strain 209. The lack of more data for the species *P. tredecaurelia* (of which only strain 209 is known) should be taken into account. Until the mating identification won't be completed, further considerations would be premature.

Summarizing all the results, the identification of *P. jenningsi* population is only contradicted by the ITS1+5.8S+ITS2 marker. Our data are in accordance to other's suggesting a big intraspecific molecular variability. The identification of the second paramecium population to the *aurelia* complex is strongly confirmed by molecular data, that cannot confirm the affiliation to *P. tredecaurelia*.

Because this is one of the few molecular studies on paramecia from tropical countries, the emphasis should perhaps be placed on the similarities, and not the differences. It is interesting to note that both paramecia from the Thai sample belongs to known morphospecies, and their sequences associate clearly with already published ones.

About the symbionts, the results of molecular analysis were more coherent and unambiguous. Because of the identity of the sequences and the occurrence in the same habitat, from now on I will assume that *P. jenningsi*'s and *P. aurelia*'s HLB belong indeed to the same species (I will make some stronger speculations at the closure of this section).

The sequences of SSU rRNA showed high similarity with the only published sequence of an HLB (*H. obtusa*). FISH experiments proved that at least the sequence obtained from *P. jenningsi* tgDNA belongs indeed to macronuclear endosymbionts. At the same time, 90.5% is a similarity value low enough to allow the establishment of new species- and genus-level taxa. The development of a specific oligonucleotide probe that doesn't bind

to other species is another requirement accomplished for the formal characterization.

Phylogenetic analysis confirmed the marked differentiation of the new HLB from "classical" holosporas like *H. obtusa* and *H. undulata*, and even more from *H. caryophila*, which should assume a new genus name (and it is likely that this will soon occurs – Schrallhammer et al., 2011b and personal communication).

Although HLB sequences publicly available are still few (more should come in the next future – Raultian, 2011), preliminary phylogenetic results seem to be coherent with the pattern of some biological features. *H. obtusa* and *H. undulata* both induce the connecting piece and have narrow host specificity. All other HLB present in the tree cannot induce the connecting piece formation, as long as we know. Although for the new HLB studied in this thesis and for that inhabiting *Frontonia*'s macronucleus the data are too limited, it is known that *H. caryophila* has a much wider host specificity. So, not only bacterial nomenclature rules, but also biological features support our proposal of splitting the genus *Holospora* in at least three different genera.

One is genus *Holospora*, with *H. obtusa* and *H. undulata*. The latter is the type species, so it has the rights on the original name. Moreover, these represents the two most studied and well-known "classical" holosporas, alongside with *H. elegans*, and are united by the capacity of inducing the connecting piece and the narrow host specificity. A second genus has to be proposed for *H. caryophila*. And a third one is needed for the HLB of this study. We propose the name *Gortzia infectiva* gen. nov., sp. nov. (Boscaro et al., 2011 and Fokin et al., 2011) in honour of Professor Hans-Dieter Görtz, who spent many years studying this group of ciliate endosymbionts. For the purposes of this work, we left the HLB from *Frontonia* as *incertae sedis*, because of its borderline value of sequence similarity with "classical" holosporas (data not shown) and the current lack of more information on its life cycle and biology.

The HLB from Thailand has to be formally described according to the current rules of bacterial taxonomy. Its SSU rRNA sequence, the species-specific oligonucleotide probe, its phylogenetic position together with morphology, location of sampling, host and infective capabilities in different *Paramecium* species will be the proposed diagnostic characters (Fokin et al., in preparation).

From a broader perspective, this is one step further on the way to definitively assess the elusive taxonomy of this group of endosymbionts. It is important not to overlook the fact that, even if diversified and various, the HLB molecularly characterized until now form indeed a monophyletic group. It is thus highly probable that many features related to the infective life cycle, like the infectious form, arose just once in the evolutionary history. This is indeed a good example of a well-defined bacterial group.

The systematic hypothesis here expressed is based on some biological features and on SSU rRNA phylogeny. It gives great importance to the connecting piece, and allows to make testable previsions for the future. For example, other species capable of inducing the connecting piece (like *Holospora elegans*, *H. acuminata* and *H. recta*) should cluster in the "classical" clade with *H. obtusa* and *H. elegans*. The exact phylogenetic positions of the species lacking this feature is harder to infer, but they should all form a single clade together with other HLB, in one of the three proposed genera or in others.

Some other characters regarded as reliable may on the contrary turn out to be false. One is the exclusiveness of HLB in *Paramecium*, or even in ciliates. Some sequences from uncultured bacteria were included in my tree to show their close relationship to HLB. The three most related to the "*Holospora*" clade come from bacteria collected in fleas of genus *Oropsylla* and in the rhizosphere of trembling aspen, two improbable environments for *Paramecium*. Because a description of these organisms is lacking, we don't know which other features, if any, they share with HLB. It is however interesting to notice that their sequences fall in the larger clade including "*Candidatus* Paraholospora nucleivisitans", another *Paramecium* endosymbiont.

One important step to put some order in the research is a revision of family *Holosporaceae* at the light of the new data. It is important to define its boundaries, and hence its apomorphies with respect to other *Rickettsiales* bacteria.

All that is said above is, strictly speaking, valid for the HLB originally found in *P. jenningsi*. The lost of the symbionts from *P. aurelia* before it was possible to fix some cells for FISH experiments made a full characterization impossible. The sequence alone isn't enough for a formal

publication. Nevertheless, there is a possible speculation in accordance with all the results.

The symbionts were quickly lost after the isolation of *P. aurelia* cells; infection experiments on aposymbiotic *P. aurelia* gave results similar to those for other species, with the bacteria invading the macronucleus but disappearing after some days. Thus, it seems probable that *Gortzia* is indeed an exclusive stable symbiont of *P. jenningsi*, meaning that this species is the only one in which it can complete its biological cycle. Its presence in the original *P. aurelia* population can be then explained with the spatial proximity with the infected *P. jenningsi* population. The latter would act as a source for continual reinfection of aposymbiotic *P. aurelia* cells that repeatedly lost (perhaps actively eliminating) and regained the bacteria. If this hypothesis is true, it can be expected that no *Gortzia* will be find in the future in *P. aurelia* population non-coexisting with other infected paramecia.

4. *Condylostoma* and its endosymbiont

4.1. Background

Unlike the other two projects, this one was focused both on the host and on the bacterial endosymbiont. Genus *Paramecium*, because of its importance and notoriety in the biologists community, has been studied under many points of view, albeit not all of its species received the same amount of attention. Most other ciliate genera are known and studied only by the smaller community of ciliatologists, and most of their features, especially at the molecular level, are unknown. *Condylostoma* is among these genera, and this study provides some insights and data on its systematics and on the possible presence of symbionts in at least one strain.

The original aim was the molecular characterization of some strains that were, at the same time, morphologically characterised by dr. Letizia Modeo and her Bachelor student, Gabriele Tomei. Strains COL2 and YK2 were then studied by a thorough and multidisciplinary approach in order to provide modern and complete redescriptions of some morphospecies of genus *Condylostoma* (Fig 4.1). This was part of a greater project with the research group of Prof. I. D. da Silva-Neto from the Universidade Federal do Rio de Janeiro (Brazil). The ultimate aim of this collaboration was to revisit the systematic classification of the genus in the light of new studies and the use of more recent methods.

Genus *Condylostoma* consists of large filter feeders heterotrich ciliates, quite common and easy to find in the interstitial environment of marine and brackish waters. A huge paroral membrane opposite to the typical well-developed AZM, moniliform macronucleus and often high contractility characterize them. But the characters used for identifying different

morphospecies are ambiguous and nearly always shared by many species or difficult to measure (e.g. the use of cell length in contractile organisms that sometimes form giants). Moreover, many original descriptions are old, of poor quality, and are difficult to compare to each other (e.g. Spiegel, 1926). Some modern descriptions also have some flaws, and none attempts to actually grow monoclonal strains under standardized conditions (e.g. Song et al., 2003).

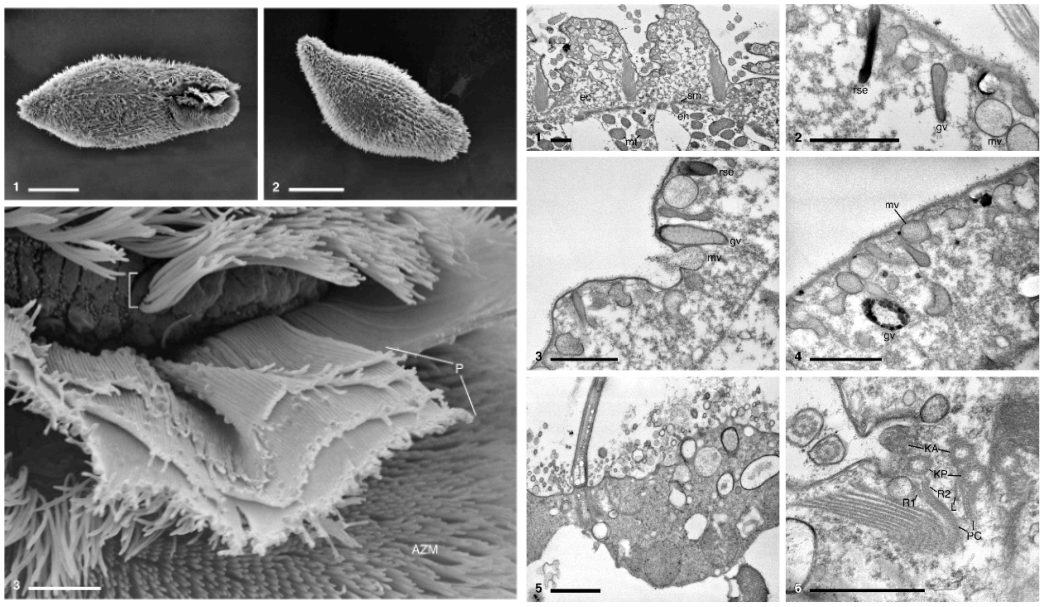

Fig 4.1. A small extract from the Bachelor Thesis of Gabriele Tomei. SEM and TEM tables showing the external morphology and the ultrastructure of *Condylostoma* strain COL2. Reproduced with the permission of the author.

To increase the accuracy and amplitude of the work, all already published molecular data were considered, and the SSU rRNA sequences of several different strains/populations of *Condylostoma* were obtained even in the absence of an accurate morphological analysis. The underlying topics were the existence of a correlation between morphological and molecular data, the validity (under a molecular point of view) of the diagnostic characters used for species identification, the assessing of a phylogenetic tree for the different strains of *Condylostoma* and for placing *Condylostoma* itself inside the class Heterotrichea.

While Gabriele Tomei carried on the multi-disciplinary characterization of strain COL2 with an aid of mine for the molecular

part, we suspected the presence of a bacterial endosymbiont inside the cytoplasm of the ciliate. This attracted our attention for more than one reason. First of all, there are currently no characterised endosymbionts from this genus. The second reason is that the preliminary results suggested the affiliation of this presumed endosymbiont to the order *Rhodospirillales* (class *Alphaproteobacteria*), for which no ciliate symbiont is currently known. Diversity is also at the base of the third reason: bacterial symbiont of ciliate from marine environment are, as a whole, largely unknown (Fokin, 2011).

To analytically review the systematics of all known ciliate bacterial endosymbionts (for which a systematic analysis is available) many pages would be necessary. I will just make a brief summary in order to highlight some common themes. A much larger, albeit slightly less updated, account can be found in Görtz, 2006.

First of all, virtually all bacterial endosymbionts belong to the phylum *Proteobacteria*. The epixenosomes of genus *Euplotidium* (which still lack a formal binomial name) belong to the very derived phylum *Verrucomicrobia*, but they are actually ciliate ectosymbionts. Methanogens of anaerobic ciliates belong to Archaea, and thus beyond the scopes of this thesis.

Proteobacteria is a huge and diversified phylum with no common autapomorphies outside molecular markers similarity. One simple explanation of this conspicuous bias towards ciliate/proteobacteria associations is that they are among the most common bacteria living in the mild environments that "common" ciliates inhabit. Another, less natural hypothesis is that common primers and probes works better with bacteria of this group. Nevertheless, symbionts are not equally distributed inside the phylum. The vast majority belongs to *Alphaproteobacteria*. The exceptions are few and instructive. About *Gammaproteobacteria*, there are some *Francisella*-like organisms (FLO; Schrallhammer et al., 2011a), *Caedibacter taeniospiralis* (Beier et al., 2002) and a still unnamed symbiont found in our lab in *Euplotes aediculatus* (Boscaro et al., 2010); all of them cluster together in phylogenetic analyses.

In *Betaproteobacteria*, there are *Polynucleobacter necessarius* and the recently characterized genus *Protistobacter* (Vannini et al., under revision), both symbionts of the same clade of *Euplotes* species and actually quite related to each other in order *Burckholderiales*, family *Burckholderiaceae*.

In *Alphaproteobacteria*, the majority of bacterial endosymbionts of ciliates can be found in order *Rickettsiales*: the HLB and most *Caedibacter* species belong to families *Holosporaceae* and *Caedibacteraceae* of this order, and also do *Rickettsia*-like organisms of family *Rickettsiaceae* and a handful of other ciliate endosymbionts (see chapter 3). There are then much rarer reports of symbionts from different orders, like one from genus *Devosia* (*Rhizobiales*) found in *Euplotes harpa* (Vannini et al., 2004).

This brief account should make clear why the discovery of a symbiont from a new order like *Rhodospirillales* is interesting to the eyes of the researchers in this field. Currently, the impression is that ciliate endosymbionts are not bacteria with endless different and independent origins but belong instead to some specialized clades from few bacterial orders. If this framework actually reflects the truth, or is instead a by-product of biased analyses, only more researches can tell.

Most of the aforementioned bacteria were found in freshwater or brackish environments. It is nevertheless difficult to sustain the hypothesis that these environments are more favourable for the development of symbiotic processes of this kind. The simpler explanation is that *Paramecium*, by far the most studied ciliate under this (and many other) point of view is a non-marine organism. *Euplotes*, another well-known ciliate, is instead common in a wide range of salinity, but species containing *Polynucleobacter* are almost all from freshwaters. It is probable that the diversity of bacteria from marine ciliate simply went nearly unnoticed until now.

4.2. The system

This project involved different specimens from *Condylostoma* genus quite differentiated by morphology, culture conditions and origins. Strains COL2 and YK2 were the original objects of interest. They were both carefully characterized from the morphological point of view while the molecular analyses were performed; results of the morphological study are summarized in the Bachelor Thesis of Gabriele Tomei.

To cite only the conclusions most relevant for my work, COL2 was morphologically identified as a strain of *C. magnum*, while YK2 (though with more uncertainties) as a strain of the *C. curva* morphospecies. Both strains were kept in culture (mass culture, in the case of COL2) and fed with the diatom *Phaedoctylum tricornutum*. COL2 grew well at 19°-20°C and often originated giant forms. YK2 grew more slowly at 14°C, and didn't form giants during this period of time. For both strains more than 100 cells were isolated from the main cultures, starved, washed and stored in ethanol 70% at -22°C for tgDNA extraction. For FISH experiments, COL2 cells were fixed instead with formaldehyde on slides.

Other sources of DNA sequences were the BG1 and BG2 populations from Guanabara Bay (Brazil), sent by Roberto Diaz and Naomi Fernandes as a stock of 20-30 cells each in ethanol 70%. BG1 was preliminary characterized as *Condylostoma arenarium*, while the morphological characterization of BG2 is still undergoing. Strains FO3 and GCO were stored as single cells in Eppendhorf tubes. FO3 was sampled in Foul Island in 2004, while Prof. Chris Lobban sent GCO from Guam. For GCO cells no morphological analyses was ever performed, but some data were available on the FO3 strain.

4.3. Results

4.3.1. Molecular characterization of *Condylostoma*

For molecular analyses purposes, an aliquot of COL2 culture was kept starving for a week. After this time, about 200 cells were isolated one by one and washed with 5 washing steps in marine (33%) sterile water and 3 washing steps in distilled water before being stored in ethanol 70%. When the cells precipitated (without the need of a centrifugation), the supernatant alcohol was mostly dropped off and other was added. All of these procedures were performed in order to eliminate or greatly reduce diatoms and bacterial contamination. A similar protocol was employed for YK2 cells storage.

From cells stored in ethanol, total genomic DNA was extracted using the NucleoSpin ® Plant II kit (Macherey-Nagel), and resuspended in 150 µl, 100 µl, 40 µl and 40 µl of distilled and autoclaved water for COL2, YK2, BG1 and BG2 respectively. The solutions were stored at -22°C and thawed at 0°C for all subsequent uses.

The amount of material for FO3 was small, so I didn't extract isolated DNA but simply centrifuged (500g, 5 min) the Eppendhorf tube, making the cell to adhere at the bottom and then removing the supernatant and pipetting the appropriate PCR solution (see below) on the cells. Exactly the same procedure was performed with the single cell from strain GCO.

Partial SSU rRNA sequences were obtained with primers 18S_F9 and 18S_R1513Hypo for COL2, YK2, BG1 and BG2 and with the pair 18S_F9 and RGD2 for FO3 and GCO. Thermal patterns were all some slight variations of the basic 35 cycle at 94°C (30''), 50°C (30'') 72°C (120''). Only the amplicon from GCO needed a second step of amplification using the primer pair 18S_F9 and 18S_R1513Hypo on a 1:50 diluted solution of the original PCR product [thermal pattern: 30x 94°C (30''), 55°C (30'') & 72°C (90'')]. Sequencing primers used were18S_R536, 18S_R1052 and 18S_F783.

A peculiarity of the electropherograms for the *Condylostoma* SSU rRNA sequences was that many of them were very good and unambiguous except for some single nucleotide positions showing a distinctive double peak. At least one of these double peaks was found in all SSU rRNA sequences obtained in this project except those of FO3 and GCO. Some homologous *Condylostoma* sequences in the database also contained at least one standard ambiguity symbol. It seemed then more correct to use the ambiguity symbol also in my sequences. The polymorphic site is not the same in all sequences.

The SSU rRNA sequence recovered from COL2 is 1644 bp long and in analysis with the BLAST software resulted more similar to DQ822482 attributed to *Condylostoma minutum* partial 18S rRNA gene sequence (similarity 99,6%, 1640/1646 identities, 1 ambiguous difference, 4 gaps; Guo et al., 2008). That from YK2 is 1670 bp long and more similar to AM295496.1 attributed to *Condylostoma* sp. strain Poe2.2 partial 18S rRNA gene sequence (similarity 99,8%, 1649/1652 identities, 1 ambiguous difference, no gaps; Modeo et al., 2006). That from FO3 is 1619 bp long and more similar to EU379939 attributed to *Condylostoma curva* partial

18S rRNA gene sequence (similarity 99,3%, 1606/1618 identities, no gaps; Guo et al., 2008). Those from BG1 and BG2 are identical; they're 1655 bp long and more similar to AM295496 attributed to *Condylostoma* sp. strain Poe2.2 partial 18S ribosomal RNA gene sequence (similarity 98,2%, 1627/1656 identities, 1 ambiguous difference, 7 gaps; Modeo et al., 2006).

ITS1+5.8S+ITS2 sequences were also obtained for COL2, YK2 and BG2 through PCR and direct sequencing [amplification primers: F919 and RGD2; thermal pattern: 35x 94°C (30''), 50°C (30'') & 72°C (120''); sequencing primers: FG1400 and RGD2]. Electropherograms quality was good with the exception of the first 30 bp of COL2 sequence, that probably contains an indel polymorphism and had to be dropped from subsequent analyses.

The resulting COL2 sequence is 710 bp. The most reliable BLAST result is sequence AY775567, attributed to the homologous sequence of *Stentor amethystinus* (similarity 78,6%, 515/655 identities, 13 gaps; Di Giuseppe & Dini, unpublished); but there is also Z49906, attributed to *Condylostoma magnum* partial 28S rRNA gene (possibly a mistake; similarity 94,4%, 336/356 identities, 1 gap; Tourancheau et al., 1995). The YK2 is 773 bp long, and is also similar to the aforementioned AY775567 from *Stentor amethystinus* (similarity 81,7%, 543/665 identities, 13 gaps). The BG2 sequence is 787 bp long, and gave a similar result (with similarity 81,4%, 543/667 identities, 15 gaps). There are no other *Condylostoma* homologous sequences in the public database.

cox1 gene sequence was obtained only for COL2 through PCR and direct sequencing [amplification primers: cox1F543_SKmP and cox1rM13_Cil_HDTF; thermal pattern: 5x 94°C (30''), 45°C (30'') & 72°C (120''), 35x 94°C (30''), 55°C (60'') & 72°C (120''); sequencing primers: T7 and M13R]. The resulting electropherograms were of very good quality.

COL2 *cox1* partial gene sequence was 1069 bp long. The most reliable BLAST result is FJ905166, a partial *cox1* sequence attributed to *Stentor* sp. (similarity 76,9%, 286/372 identities, 14 gaps; Strüder-Kypke & Lynn, 2010). It is to be noted that this is the first *cox1* sequence obtained for genus *Condylostoma*, and one of the very few for ciliates of class Heterotrichea.

4.3.2. Phylogenetic analysis of Class Heterotrichea

Phylogenetic analysis (Fig. 4.2) was only performed on SSU rRNA sequences because of data availability. The analysis included 59 sequences: the 6 obtained in this project, other 7 assigned to genus *Condylostoma*, 31 additional sequences representing other heterotrich taxa and 15 sequences belonging to karyorelicteans used as outgroup.

Inside genus *Condylostoma*, 4 sequences were already publicly available and used by other authors: those of *C. minutum*, *C. spatiosum* and *C. curva* and of the strain Poe2.2. The sequence once labelled as *Condylostentor auriculatus* more probably belongs to *Condylostoma wangi*, and as such is here presented. Strains P2CO5, Y2 and S1 were molecularly, but not morphologically, characterized in our lab (unpublished data).

Other Heterotrichea sequences were chosen in order to represent all other morphospecies whose SSU rRNA gene sequences were characterized and published. Those consists of 8 sequences of *Stentor* (Stentoridae), 6 for *Blepharisma* (Blepharismidae), 2 for *Fabrea salina* (Condylostomatidae), 3 for *Spirostomum* (Spirostomidae), 2 for *Peritromus* (Peritromidae) and one each for *Gruberia* (Spirostomidae), *Climacostomum* (Climacostomidae), *Chattonidium* (Chattonidiidae), *Eufolliculina*, *Folliculina* (Folliculinidae) and *Maristentor* (Maristentoridae). The only "uncultured" heterotrich sequence in this tree was collected in our lab, and is added in order to stabilize a long branch.

All sequences were automatically and then manually aligned using both the ARB software package and ClustalX (for some difficult regions in the Karyorelictea sequences). The sequences were trimmed to the shorter one, keeping all the resulting 1674 characters. AIC and BIC tests, performed with the ModelTest software, gave TIM2+I+G and TrN+I+G as preferred evolutionary model outputs. Likelihood mapping analysis

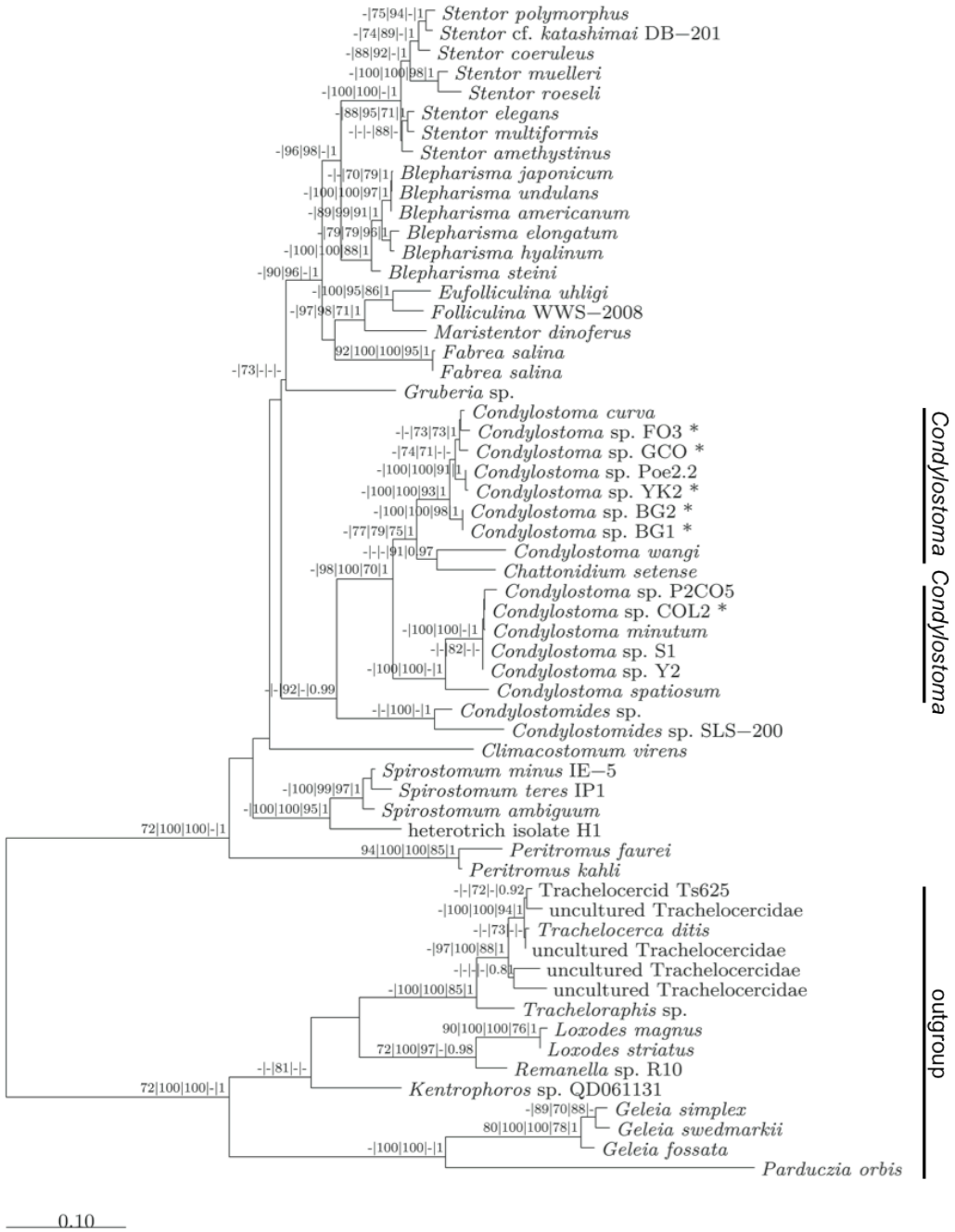

Fig 4.2. SSU rRNA gene sequences Maximum Likelihood tree. 59 sequences and 1674 characters were employed. The node values represent NJ, MP and ML bootstrap values, TreePuzzle quartet puzzling support and Posterior Probability respectively, cut as explained in section 2.4.8. Asterisks mark sequences obtained in this work. Supplementary information are provided in the text.

performed with TreePuzzle showed that 92,9% of the 100.000 sampled quartet trees were in the resolved regions, thus suggesting that the data-set was adequate for phylogenetic analyses, but also highlighting the presence of some noise (some of which coming from identical sequences like those of BG1 and BG2).

ARB NJ, PHYLIP DNAPARS, PHYML (with the tools provided by the ARB package), TreePuzzle Tree Reconstruction and MrBayes were used for producing NJ, MP, ML, quartet puzzling and Bayesian tree respectively. GTR+I+G was the model used in all model-based methods, with the exception of ARB NJ (in which the felsenstein correction was used instead). 1.000 bootstrap pseudoreplicates were performed for ARB NJ, 500 for MP and ML trees. 100.000 puzzling step were performed in order to obtain the quartet puzzling support values, and posterior probability values are derived from 1.000.000 generations (burnin 25%) for each of the 3 runs of MrBayes (each consisting of 3 heated and a cold chain).

Monophyly of class Heterotrichea and of *Stentor*, *Blepharisma*, Folliculinidae, *Spirostomum* and *Peritromus* are recovered with high support values. Family Climacostomidae, Spirostomidae and Condylostomidae appear instead paraphyletic. Higher-level relationships are all weakly supported, with the exception of the cluster *Stentor*-*Blepharisma*-Folliculinidae-*Maristentor*-*Fabrea*. The clade including *Condylostoma* and *Chattonidium setense* sequences is also highly supported.

4.3.3. Molecular characterization of the COL2 symbiont

The partial SSU rRNA gene sequence of the presumed bacterial endosymbiont of COL2 was obtained through PCR and direct sequencing [amplification primers: 16S_alpha_F19a and 16S_R1492; thermal pattern: 35x 94°C (30''), 55°C (30'') & 72°C (90''); sequencing primers: 16S_R515ND, 16S_F343ND and 16S_F785ND]. The resulting electropherograms was of good quality and without ambiguities.

The sequence obtained is 1422 bp long. The highest similarity sequence obtained with a BLAST research is FJ403068.1, attributed to the

partial 16S rRNA of the uncultured alphaproteobacterium clone MD2.19 (similarity 93,4%, 1339/1433 identities, 18 gaps; Johnson et al., unpublished). The most similar results from a sequence identified at the genus level is FJ952806.1 from *Thalassospira* sp. (similarity 89,2%, 1186/1329 identities, 40 gaps; Rypien et al., 2010).

4.3.4. FISH experiments

Isolated, starved and washed COL2 cells were fixed with formaldehyde on slides for FISH experiments. The probe COL2_1249 was designed in order to match the sequence of the SSU rRNA gene of the presumed endosymbiont, and was labelled with the fluorophore Cy3.

COL2_1249 and Eub338I labelled with fluorescein were exploited jointly in the experiment (Fig. 4.3). The results were not completely encouraging. Only half of the observed *Condylostoma* cells gave positive signals on both probes. Besides, the number of bacterial cells was low, never higher than 10 (and more often near 1-2). On the other side, these cells seem to be outside the digestive vacuoles, apparently inhabiting the host cytoplasm.

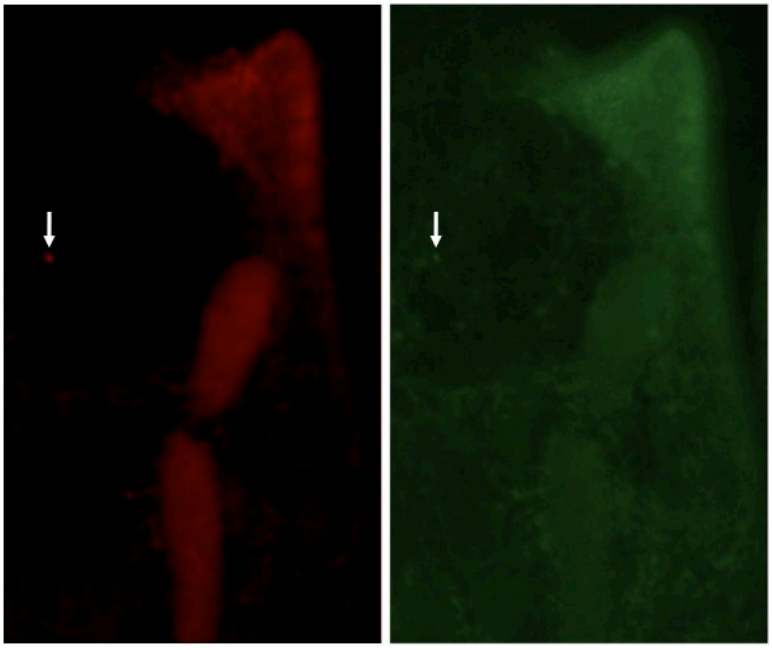

Fig 4.3. Epifluorescence microscope pictures of a COL2 cell. Signal of the COL2_1249 probe (red) and of the Eub338I probe (green). Arrows point at the single bacterium present. The moniliform macronucleus and the oral region show a high level of autofluorescence.

4.3.5. Phylogenetic analysis of the bacterial sequence

Phylogenetic analysis was performed on the SSU rRNA sequence. The analysis included 93 sequences, 77 from *Rhodospirillales* taxa (included the one obtained in this work), 11 from other *Alphaproteobacteria* orders, as outgroup and 5 because of their similariy with the COL2 putative symbiont sequence.

The 77 sequences of *Rhodospirillales* were chosen in order to represent all genera of the order for which at least one SSU rRNA sequence was available. They belong to one of the two families of *Rhodospirillales*: *Rhodospirillaceae* and *Acetobacteraceae*. 3 sequences of uncultured bacteria and from non-Rhodospirillales genera (*Terasakiella pusilla* and *Kopriimonas byunsanensis*) were included because of their high similarity value with that of the putative symbiont.

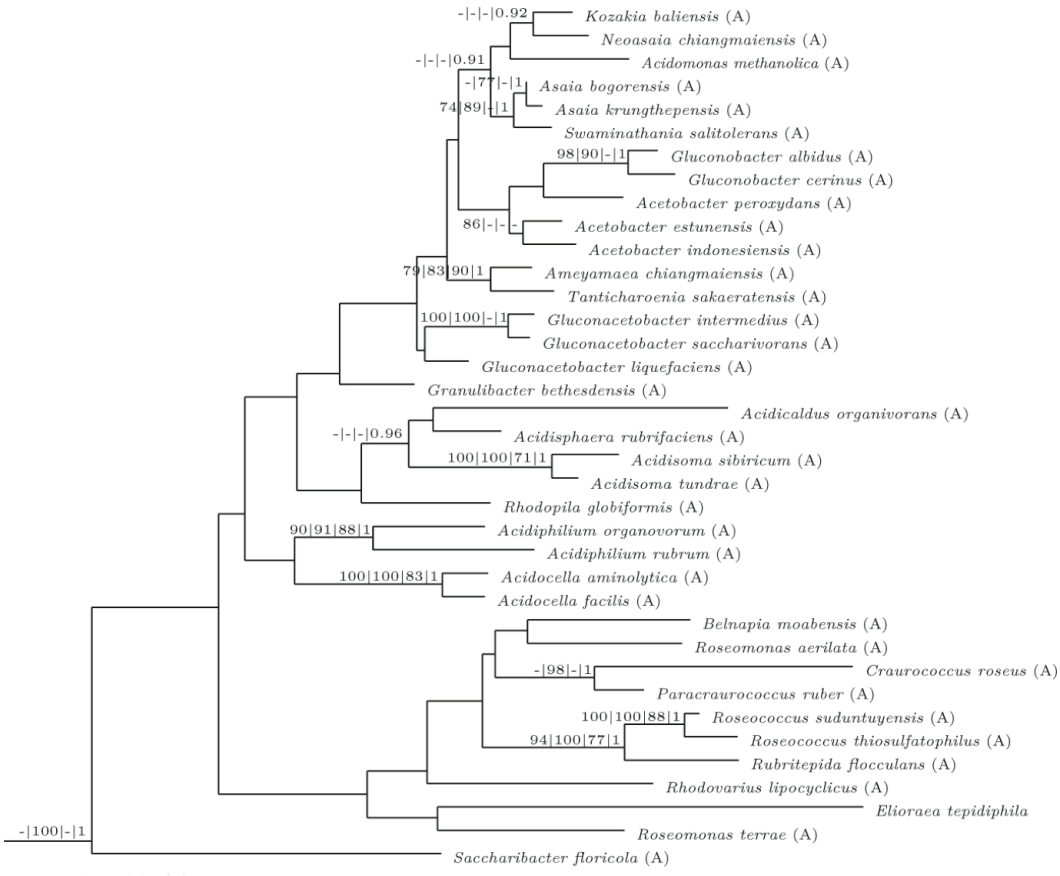

Fig 4.4. (I)

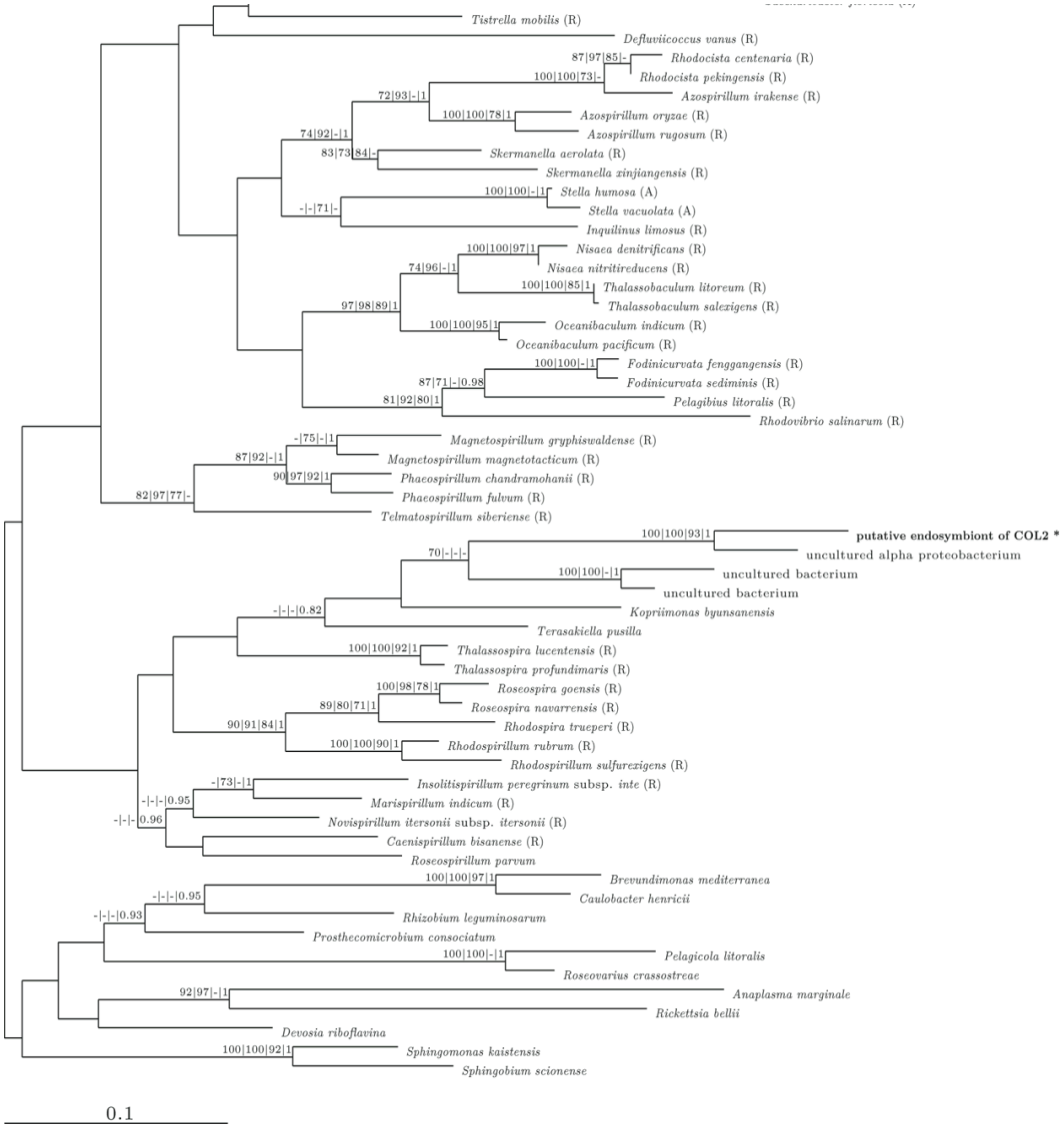

Fig 4.4. (II) SSU rRNA gene sequences Maximum Likelihood tree. 93 sequences and 1218 characters were employed. The node values represent MP and ML bootstrap values, TreePuzzle quartet puzzling support and Posterior Probability respectively, cut as explained in section 2.4.8. [A] stands for Acetobacteraceae, [R] for Rhodobacteraceae. NJ bootstrap values are not shown. The asterisk marks the sequence obtained in this work More information are provided in the text.

All sequences were automatically and then manually aligned using the ARB software package. The sequences were trimmed to the shorter one, and a 5% filter was employed to buffer some possible ambiguity in the alignment, keeping 1218 characters. AIC and BIC tests, performed with the ModelTest software, both gave GTR+I+G as preferred evolutionary model. Likelihood mapping analysis performed with TreePuzzle showed that 97,3% of the 100.000 sampled quartet trees were in the resolved regions, thus suggesting that the data-set was perfectly adequate for phylogenetic analyses, with a very low percentage of unresolved topologies.

ARB NJ, PHYLIP DNAPARS, PHYML (with the tools provided by the ARB package), TreePuzzle Tree Reconstruction and MrBayes were used for producing NJ, MP, ML, quartet puzzling and Bayesian tree respectively. GTR+I+G was the model used in all model-based methods, with the exception of ARB NJ (in which the felsenstein correction was used instead). 1.000 bootstrap pseudoreplicates were performed for ARB NJ, 500 for MP and ML trees. 100.000 puzzling step were performed in order to obtain the quartet puzzling support values, and posterior probability values are derived from 1.000.000 generations (burnin 25%) for each of the 3 runs of MrBayes (each consisting of 3 heated and a cold chain).

Most of the nodes in the tree were recovered by all phylogenetic methods, but show nevertheless very low support values. The monophyly of the order itself is not supported. Both family, as currently defined, are paraphyletic, although all *Acidobacteraceae* genera except *Stella* cluster together in a highly supported clade. Inside this clade, relationships receive low support because of the instability of *Saccharibacter floricola*, that shifts its phylogenetic position according to the method used.

Most of the genera represented by more than one sequence appear monophyletic, although the monophyly is not always supported by high support values.

The sequence of the putative COL2 symbiont clusters with acceptable support in a clade with the three uncultured bacteria, *Terasakiella pusilla* and *Kopriimonas byusanensis*.

4.4. Discussion

As said above, this molecular study was conducted in parallel with a deep morphological characterization of two of the strains. Comparison with all the available material in the literature suggested the classification of strain COL2 as *Condylostoma magnum* and of strain YK2 as *Condylostoma curva*. But another equally important result of the analysis was the recognition that different morphospecies of genus *Condylostoma* are distinguished by blurred and not very reliable characters. This, alongside poor old descriptions and the common use of populations instead of monoclonal strains, were identified as the major problems for good identification.

Because of all this issues, it is hard to settle the question of the correct species identification for COL2 – notably, there are no SSU rRNA sequences in the public database assigned to *C. magnum*. The similarity of the ITS1+5.8S+ITS2 sequence of COL2 to another one attributed to *C. magnum* is of little importance, because there are no other homologous *Condylostoma* sequences to make comparisons with. As I will detail below, molecular data seem to cast some more serious doubts on the identification of YK2.

Generally speaking, there are many disagreements between the molecular tree and morphologic expectations. Nevertheless, the intrageneric molecular phylogenetic tree obtained has some points of interest that deserve to be discussed. These can constitute a basis for subsequent works and hypotheses in a more modern and multidisciplinary context, in order to reach a new classification system for this genus. Because this thesis is concerned with molecular characterizations, I will skip many morphological issues, assume that there is no accordance between the two approaches, discuss the molecular results and then look again at the morphologic questions.

First of all, and as already showed by other authors (e.g. Guo et al., 2008), the genus *Condylostoma* is not monophyletic with respect to *Chattonidium setense*. This result seems to be very robust, and has to be reconciled with morphological cladistic analyses. There are more than one formally correct method of resolving this issue, but the most simple one is renaming *Chattonidium setense* as *Condylostoma setense*, abandoning the

monospecific genus *Chattonidium* and the monogeneric family Chattonidiidae – providing that some morphological apomorphy for the resulting clade can be found.

The second evident result about the intrageneric systematic (including *Chattonidium* in genus *Condylostoma* for simplicity in the rest of the discussion) is the presence of three well-supported clades (Fig. 4.5). One is labelled as the "*curva* group" by the name of its only sequence attributed to a species-level taxon. This clade contains 8 sequences: the one from *Condylostoma curva* and 7 from strains or populations studied in our lab (5 during this project). Only BG1 was provisionally assigned to a morphospecies (*Condylostoma arenarium*) and, together with the identical sequences of BG2, places in basal position. The morphological variability is fairly high, while the molecular one is the lowest of the three clades. YK2 and *C. curva* both belong to this group, but don't cluster together. Strain FO3 is indeed more closely related to *C. curva* than YK2 is, but its morphology is rather different.

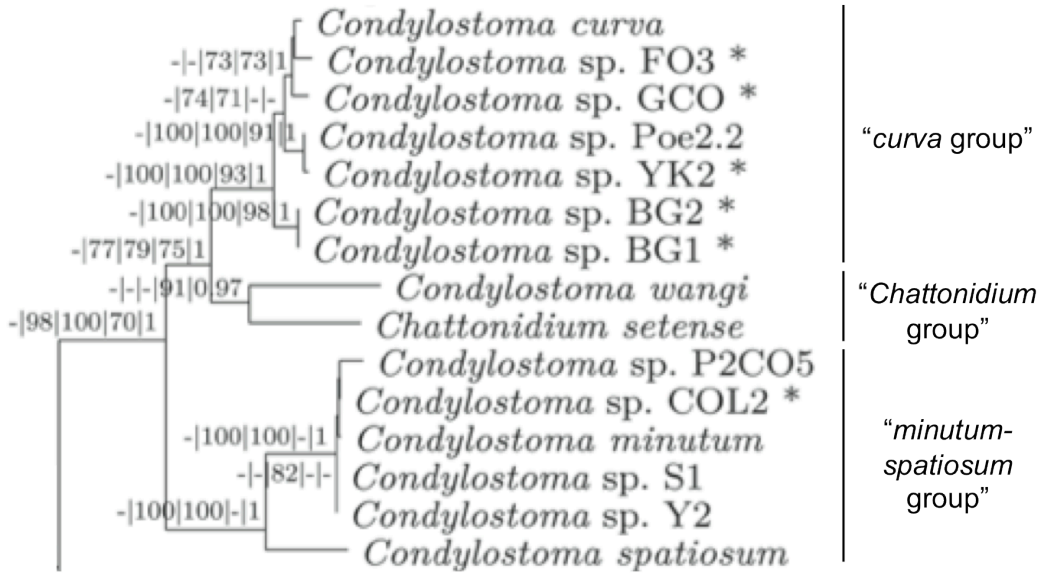

Fig 4.5. Detail of the phylogenetic tree showed in Fig. 4.3. showing the three *Condylostoma* clades described in the text.

The second clade is labelled "*Chattonidium* group" and includes only the sequence from *Chattonidium setense* and another deposited as *Condylostentor auriculatus* but later corrected into *Condylostoma wangi*.

Morphological and molecular diversity are both high – on morphological grounds, *Chattonidium* was even classified in a separate family.

The third clade is labelled "*minutum-spatiosum* group" and includes the SSU rRNA sequence of the strain COL2. With the exception of the sequence of *C. spatiosum*, the other sequences are highly similar, thus making their relationships impossible to evaluate with the use of this marker alone. Moreover, the morphological analysis performed on COL2 strain and the one published for the *C. minutum* give very different outcome, thus making their huge molecular affinity difficult to explain. This is the most problematic clade of all.

There is no infallible similarity threshold that can separate "species". It is also interesting to note that there aren't any mating-compatibility studies in the literature of this genus, so we don't know nothing about the correlation between morphospecies and biological species. This study shows that there is little between morphospecies and molecular diversity. For all that we know, and depending on the definition of species we use, COL2 and the *C. minutum* studied by Guo et al. could belong to different species that share a strong phylogenetic bond, or to the same species that shows a huge morphological variability.

It is possible, of course, trying to perform a more thorough meta-analysis on all the data at our disposal in order to find an *ad hoc* way to reconcile all of them. The odds in favour of a success are however low. It is simply possible that the current knowledge on the systematic of this genus is wrong or inadequate. Molecular studies cannot, on their own, establish a new classification, but they are giving us the only stable data on which we can rely, because of the difficulties met by morphological studies detailed above. It is much safer to wait for other multidisciplinary data before even thinking about reviewing the systematic status of this genus and the species belonging to it. If the results of this study will be corroborated, many identified morphospecies (and hence the diagnostic characters on which they were defined) will be proven inconsistent, and a classification based on the three-clade tree will be established. The conclusions of this project have been already presented to an international congress (Modeo et al., 2011).

The phylogenetic analysis performed can tell us something also on the general systematics and phylogeny of the whole class Heterotrichea. The results are indeed much similar to those already published, thus

corroborating them. Among the well-supported clade there are a very stable "crown group" consisting of genera *Stentor* (Stentoridae) *Blepharisma* (Blepharismidae), *Eufolliculina* and *Folliculina* (Folliculiniidae), *Maristentor* (Maristentoridae) and *Fabrea*; and the association between *Condylostoma* (as defined above) and *Condylostomides*. All genera containing more than one species (with the already described exception of *Condylostoma*) result monophyletic, as does the whole class.

All other relationships vary between studies, methods, sequences chosen and are however supported by very low values of bootstrap. In particular, there is no certainty on the most ancient split inside the class – in my tree the one between the genus *Peritromus* and the rest – and on the position of *Gruberia* and *Climacostomum*. Family Climacostomidae seems to be clearly polyphyletic, with its members *Fabrea* and *Condylostoma* never clustering together despite their morphological affinities. Family Spirostomidae, represented in this tree by genera *Spirostomum* and *Gruberia*, is probably not monophyletic too.

About the bacterium, obtained results have to be considered as preliminary and thus treated cautiously. Some data support my hypothesis that *Condylostoma* strain COL2 hosts a bacterial endosymbiont of order *Rhodospirillales*. The first is the recovery of a TEM picture showing a single bacterial cell in the cytoplasm near the oral region, outside the digestive vacuole (Modeo, personal communication). The second is that I obtained an alphaproteobacterial SSU rRNA sequence through PCR with non specific primers and direct sequencing. The perfect quality of the electropherograms suggests that the bacterium whose the sequence belongs is at least predominant in the studied system. As already explained, the isolation and washing of single ciliate cells before tgDNA extraction cannot be used as assurance against any possible bacterial contamination from the original medium. On the other side, if we are dealing with a contaminant, its DNA preponderance over that of other bacteria in the prokaryotic community is very strange.

The sequence itself supplies another reason inducing optimism. It is largely different from any other in the database, although clustering inside *Rhodospirillales* sequences. Common environmental bacteria sequences are often recovered by environmental studies or cloning, and submitted as

such. The alternative hypothesis is that the sequence I retrieved belongs to a prolific contaminant of the COL2 culture that belongs to a never befored observed free-living genus.

An observation that somehow contrasts the symbiont hypothesis comes from the FISH experiment. Not many COL2 cells gave positive results with the specific probe, the signals were weak and few in number. As showed in the first project with Gortzia (chapter 3), bacterial infections usually appear quite differently, with hundreds or thousands of bacterial cells equally distributed in one specific cellular compartment.

Provisionally assuming that one bacterial symbiont is indeed present and that I obtained its SSU rRNA gene sequence, the results of the phylogenetic analysis are interesting. Not many certainties can be drew from the phylogenetic tree, because all support values are very low. The monophyly itself of order *Rhodospirillales* is not strongly supported. Besides, the two families of this order appear clearly paraphyletic. Nevertheless, the COL2 symbiont always clusters with a variety of other basal sequences, mostly affiliated to *Rhodospirillaceae* organisms.

Many *Rhodospirillaceae* bacteria belong to the metabolic group of the non-sulfur purple bacteria, facultative photoheterotrophic organisms. Would this association be confirmed, and the phylogenetic relationships established, this symbiont could represent a hugely interesting organism to study under a metabolic point of view, especially in the light of the possible physiological interactions with the host. Given the highly preliminary status of this characterization, I have to postpone such wild speculation to the future.

One more reason to be careful comes when looking with more attention at the phylogenetic tree. The putative symbiont's sequence cluster together with all the others that are not formally included in order *Rhodospirillales*: the three uncultured, *Terasakiella pusilla* and *Kopriimonas byusanensis*. The clade received moderate support in the analysis. One possibility is that these organisms truly belong to the order *Rhodospirillales*; this is supported by their clustering inside the order with all the methods employed, albeit with low support values. An alternative that should be considered is the presence of an artifact that "force" this long-branching clade inside the order *Rhodospirillales*, lowering all support values in the analysis.

In order to test these two hypotheses, more sets of sequences should be tested, removing and adding some of them searching for a more supported tree. This procedure requires much time and efforts, and should be considered only when the endosymbiotic nature of the bacterium will be definitively proven.

5. Molecular survey of *Paramecium duboscqui*

5.1. Background

As already mentioned, the general fame of *Paramecium* is not equally shared between all of its species. Molecular biologists preferably use *Paramecium tetraurelia*, of the *P. aurelia* group. *P. caudatum* and *P. bursaria* are two other very well-studied taxa.

Even if the number varies according to the author assessing them (there are many once- or very rarely found morphospecies formally described), there are about 20 different morphospecies in the *Paramecium* genus, and probably much more biological species. Some of the *aurelia* species and most of the brackish-water paramecia are very poorly known. Furthermore, as sometimes happens for model organisms, although we know many details about the molecular biology and genetics of some species, the naturalistic knowledge on many aspects of their biology is much less developed. Ecology, distribution and phylogenetics of this genus are not completely described, and intraspecific variability analyses only started to appear in the last two decades.

Paramecium duboscqui (Fig 5.1) is one of the "neglected" species. Originally described in 1933 (Chatton & Brachon, 1933) it was later considered a vary rare or even non valid morphospecies, until it was rediscovered and validated again in recent years (e.g. Shi et al., 1997). It is more often collected in brackish-water environment, or in polluted bodies of water, and it seems to have a preference for lower temperatures than other paramecia. It is generally collected in late autumn, winter or spring, or throughout the years at northern latitudes (but a recently recovered strain included in this survey was sample in Tunisia). In laboratory, it can survive and divide up to 20°C, but is unable to do so over 25°C. It has

rounded ends, 1-8 (usually 2) spindle-form micronuclei, two contractile vacuoles without connecting canal and with a single pore each. The oral aperture is in the middle of the kidney-shaped cell (Fokin et al., 1999).

My work was aimed at surveying the molecular variability between different strains of *P. duboscqui* collected through the years in different countries and environments. The paper of Fokin, Stoeck and Schmidt (Fokin et al., 1999) was very influential, and it can indeed be considered the base of this project. That paper was in fact the only one attempting to review the ecology, distribution and molecular diversity of this species, exploiting a high number of strains. A subset of the very same strains was used in the present study. It was hence possible to perform a reasonable comparison of the results.

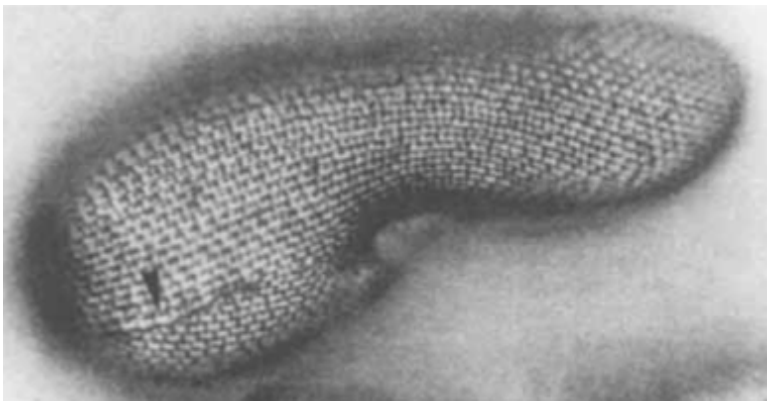

Fig 5.1. *P. duboscqui cell* stained by silver impregnation, right side. The arrowhead points to the long cytoproct. From Shi et al., 1997.

Although I will refer to that paper for morphological and ecological considerations, the weak point of that pioneering work was the molecular method used. Different strains' DNA were confronted with the RAPD technique. This method was then commonly used for many *Paramecium* species and proved to be quite sensible (e.g. Stoeck & Schmidt, 1998), but is much less flexible as gene sequencing. The latter is, as I will show, equally or even more capable of discrimination, much more reliable for phylogenetic inferences and more fit for comparisons among different studies. In the following sections, I will analyse my results in the light of those of the 1999 article, defending the arguments just stated and integrating all the analyses in a coherent output.

This third project aims the ciliated hosts and their characterization, but *P. duboscqui* is known to contain at least one bacterial endosymbiont (Schrallhammer et al., 2006). The strain containing that symbiont was part of the collection of our lab, kept by Prof. Fokin, but unfortunately became extinct some years ago. It is a shame, because its already published SSU rRNA gene sequence was by far the most divergent of all, and a comparison with the *cox1* gene sequence data would have been very valuable.

I also used the material from *P. duboscqui* strains, alongside with some other *Paramecium* species to obtain preliminary results for a further project. While I relied mostly on *cox1* and SSU rRNA sequences as markers for my purposes, I obtained some sequences of mitochondrial LSU rRNA, a never-before used (at least in ciliate) molecular marker. At the end of the Discussion section, I will present some preliminary evaluations on the use of this marker, looking at the whole genus *Paramecium*.

5.2. The system

The organisms studied in this project were collected in different times, places, environments and by different researchers (Tab 5.1). They were all assigned to the *P. duboscqui* morphospecies by means of morphological diagnostic characters, and are nearly impossible to distinguish from one another.

With one exception, I used as starting material cells fixed years ago in ethanol 70%. The exception is the strain Tub2, which I isolated during my bachelor traineeship in 2009 from a sample collected by Prof. Fokin in the Orbetello lagoon. I kept the monoclonal mass culture since then in 5% salt water, fed with *Dunaliella salina* (about 15 microliters of dense culture once a week). I identified them through SSU rRNA gene sequencing, and Prof. Fokin later confirmed their belonging to the *P. duboscqui* morphospecies.

Strain	Sampling	Reference
Ku4-8	Kunashir Island (Russia) September 1990	Schrallhammer et al., 2006; Fokin et al., 1999;
AWH9-4	Woods Hole (USA) March 1997	This study; Fokin et al., 1999
BB8	Berezovye Island, Gulf of Finland (Russia) May 1993	This study; Fokin et al., 1999
702	Harbin (China) -	This study; Fokin et al., 1999
IG2-1	Giglio Island (Italy) March 1996	This study; Fokin et al., 1999
In05	Naples (Italy) April 2005	This study
Pd-2	Pisa (Italy) March 2006	This study
Ppd-3	Procida Island (Italy) May 2006	This study
Tub2	Orbetello Lagoon (Italy) April 2009	This study
TWH2	Tunisia, December 2006	This study

Tab 5.1. List of the *P. duboscqui* strains surveyed in this work, for which molecular sequences are available.

Strains AWH9-4, 702, IG2-1 and BB8 were studied also in the 1999 article by Fokin et al., so the results are directly comparable. Strains Tub2, In05, Pd-2 and Ppd-3 from Italy, and strain TWH2 from Tunisia are exclusive of the present study. Strain Ku4-8 from Kunashir Island was present in the '99 paper, but is now lost. However, its SSU rRNA gene sequence is available in the online database (accession number: AM236094, 1709 bp long, Schrallhammer et al., 2006). This was also the strain harbouring the symbiont *Caedibacter macronucleorum*.

P. duboscqui is one of the *Paramecium* species that grow preferably in water with some salinity. It can be effectively fed with algae grew at the same salinity, and needs significantly less food material than other *Paramecium* species. It also grows well at lower temperature.

Molecular data from other *Paramecium* species were recovered and exploited for this project. These are: *P. schewiakoffi* strain SH1-38 from Shangai (China), that Prof. Fokin gave me in the form of ethanol-stored cells; *P. caudatum* strain CG6 from Vercelli (Italy), whose tgDNA I obtained from a mass culture during my Bachelor Thesis in 2009; and the *P. jenningsi* and *P. aurelia* Thai populations described in chapter 3.

5.3. Results

5.3.1. Sequences obtained from *P. duboscqui* strains

For all analysed *P. duboscqui* strains, the starting material consisted in 50-70 cells stored in ethanol 70% since the time of sampling. The only exception was strain Tub2, whose DNA was already extracted and isolated during my Bachelor traineeship.

NucleoSpin ® Plant II kit (Macherey-Nagel) was used to obtain total genomic DNA from ethanol-stored cells. The DNA was eluted in distilled and autoclaved water in a volume of 100 µl. The solutions were stored at -22°C and thawed at 0°C for all subsequent uses.

Partial SSU rRNA sequences were obtained through PCR and direct sequencing [amplification primers: 18S_F9 and 18S_R1513Hypo; thermal pattern: 30x 94°C (30''), 50°C (30'') & 72°C (120''); sequencing primers: 18S_R536, 18S_R1052 and 18S_F783]. All resulting electropherograms gave good quality results without ambiguities, and sequences 1722 bp long.

Complete ITS1+5.8S+ITS2 sequences were obtained through PCR and direct sequencing [amplification primers: 18SF783 or 18SF919 and RGD2; thermal pattern: 35x 94°C (30''), 50°C (30'') & 72°C (180''); sequencing primers: FG1400 and RGD2]. The electropherograms were of good quality in a range of 921 bp and contained ample overlapping regions with those obtained for SSU rRNA.

cox1 gene sequences were obtained through initial PCR amplification [primers: cox1F543_SKmP and cox1_F543_SKmP; thermal pattern: 20x 94°C (30''), 45°C (45'') & 72°C (60'') with single primer, then 2x 94°C (30''), 45°C (45'') & 72°C (60'') 35x 94°C (30''), 55°C (45'') & 72°C (60'') with both primers] and subsequent semi-nested on the products [primers: T7/PdubR1088 and PdubF922/M13R; thermal pattern: 30x 94°C (30'') 50°C (30'') & 72°C (60'')] and direct sequencing [respectively with primers T7 and M13R]. The electropherograms were of perfect quality and without ambiguities in a range of 660 bp.

For comparison purposes, molecular sequences were collected also from other *Paramecium* strains. *P. schewiakoffi* strain SH1-38 tgDNA was

extracted from cells stored in ethanol with the kit NucleoSpin Plant II and eluted in a final volume of 100 µl.

P. schewiakoffi cox1 gene sequence was obtained through initial PCR amplification [primers: COIfT7_Oligo_INHK and COIM13R_Cili_HDTF thermal pattern: 20x 94°C (30''), 37°C (30'') & 72°C (120'') with single primer, then 1x 94°C (30''), 45°C (30'') & 72°C (120'') 30x 94°C (30''), 50°C (30'') & 72°C (120'') with both primers] and subsequent reamplification on the products diluted 1:100 with distilled water [primers: T7 and M13R; thermal pattern: 25x 94°C (30'') 50°C (30'') & 72°C (120'')] and direct sequencing [sequencing primers: T7 and M13R]. The electropherograms were of good quality and without ambiguities.

P. caudatum cox1 gene sequence was obtained through PCR and direct sequencing [amplification primers: COI_F543_SKmP and COI_R1321_SKmP; thermal pattern: 5x 94°C (30''), 45°C (60'') & 72°C (60'') 35x 94°C (30''), 55°C (60'') & 72°C (60''); sequencing primers: T7 and M13R] The electropherograms were of perfect quality and without ambiguities.

5.3.2. Comparative analyses on the *P. duboscqui* sequences

SSU rRNA gene sequences were trimmed to 1710 bp in order to be easily compared with the published sequence of strain Ku4-8. This 10 sequences can be classified into three groups: group A, with Tub2, TWH2, Pd-2, Ppd-3, In05, IG2-1 and AWH9-4; group B, with BB8 and 702; and group C, with Ku4-8.

Sequences inside each of the three groups are identical. Between groups A and B there are 2 single-base differences (at sites 602 and 998) and no gaps; the similarity between the two is thus 99,88%. Group C is much more divergent: using the alignment produced by BLAST software, it has 8 differences and 9 gaps with respect to the sequences of the other groups, thus sharing with them a similarity of just 99,01%.

Groups A, B and C can be geographically delimited. Group C contains only the Ku4-8 strain, from Kunashir Island (Russia), just north to Japan. Group B extends from China to the Gulf of Finland. Group A is typical of

the Mediterranean Sea (Tunisia and Italy) and contains the strain AWH9-4 from the Atlantic coast of the USA.

ITS1+5.8+ITS2 sequences of the 9 strains of this project are slightly more variable. The same groups A and B identified by SSU rRNA marker can be recovered, and they differ by 3 bases (at sites 56, 99 and 411). Site 433 is hypervariable: it hosts a cytosine in group B sequences and In05, Ppd-3 and TWH2, and a tyrosine in the other group A sequences. Average similarity between groups A and B is thus 99,61% using this marker. Group C cannot be evaluated because an homologous sequence from strain Ku4-8 was not available.

cox1 sequences can be classified in up to four groups. Group A1 contains the Mediterranean strains of the A group identified by the other markers, while group A2 contains only the american strain AWH9-4. These two groups differ only by one base at site 34; the similarity value is thus 99,85%. Strains BB8 and 702 belong respectively to group B1 and B2. This two sequences differ by 2 bases (site 142 and 278) and their similarity value is 99,70%. Average similarity between group A sequences and group B sequences is 93,64%, with 42 different bases on average.

Tab 5.2 shows some values of dissimilarity (p-distances) between selected pairs of *cox1* nucleotidic sequences from paramecia. In order to make them comparable, they were aligned with ClustalX and trimmed to the shortest of them (493 bp).

Besides similarity assessment, phylogenetic analyses were performed for all three molecular markers (Fig 5.2). For SSU rRNA, 10 *P. duboscqui* sequences were included in the analysis (alongside the homologous sequence of *P. jenningsi* obtained in the first project, as detailed in chapter 3). For ITS1+5.8S+ITS2 and *cox1* sequences, there are only 9 *P. duboscqui* sequences, because Ku4-8 is missing. In all cases, the sequences number was very low, and ambiguities virtually absent. Thus, only a Maximum Likelihood analysis with the PHYML software provided by the ARB package was performed. The values associated with each node are the bootstrap values from 500 pseudoreplicates.

Pairs	Dissimilarity value	References
P. duboscqui strains In05/BB8	6,1%	This study
P. multimicronucleatum strains IN1/IP10	8,7%	Barth et al., 2006
P. septaurelia strain AZ5-2 / *P. octaurelia* strain K9	0,2%	FJ003697, FJ003702; Catania et al., 2009
P. primaurelia strain V7-6 / *P. pentaurelia* strain Nr1-1	6,1%	FJ003654, FJ003693; Catania et al., 2009
P. primaurelia strain V7-6 / *P. sexaurelia* strain AZ8-4	18,1%	FJ003654, FJ003695 Catania et al., 2009
P. jenningsi / *P. schewiakoffi* strain SH1-38	14,7%	This study

Tab 5.2. *cox1* gene sequences dissimilarity values of selected paramecium specimens pairs. *P. duboscqui* strains belong to the two major groups identified in the comparison analysis; the strains of *P. multimicronucleatum* belong to two clade suspected to be sibling species (Barth et al., 2006); other strains belong to different species of the *aurelia* complex (Catania et al., 2009).

The sequences were previously aligned with the ClustalX software. For SSU rRNA, the analysis was performed on 11 sequences and 1716 characters. For ITS1+5.8S+ITS, 10 sequences and 923 characters were used. For *cox1*, ten sequences were included, with 660 characters; GTR+I +G was the model of choice in the analyses.

The SSU rRNA and *cox1* phylogenetic trees are in accordance with expectations. They identified the three groups (2 with *cox1* sequence), that can thus also be called clades. The most ancient split is those between group C and the others, according to the SSU rRNA phylogeny. ITS1+5.8S+ITS tree doesn't support the two-clade split provided by SSU rRNA analysis. While group B appears as a clade, group A form a paraphyletic group without bootstrap support.

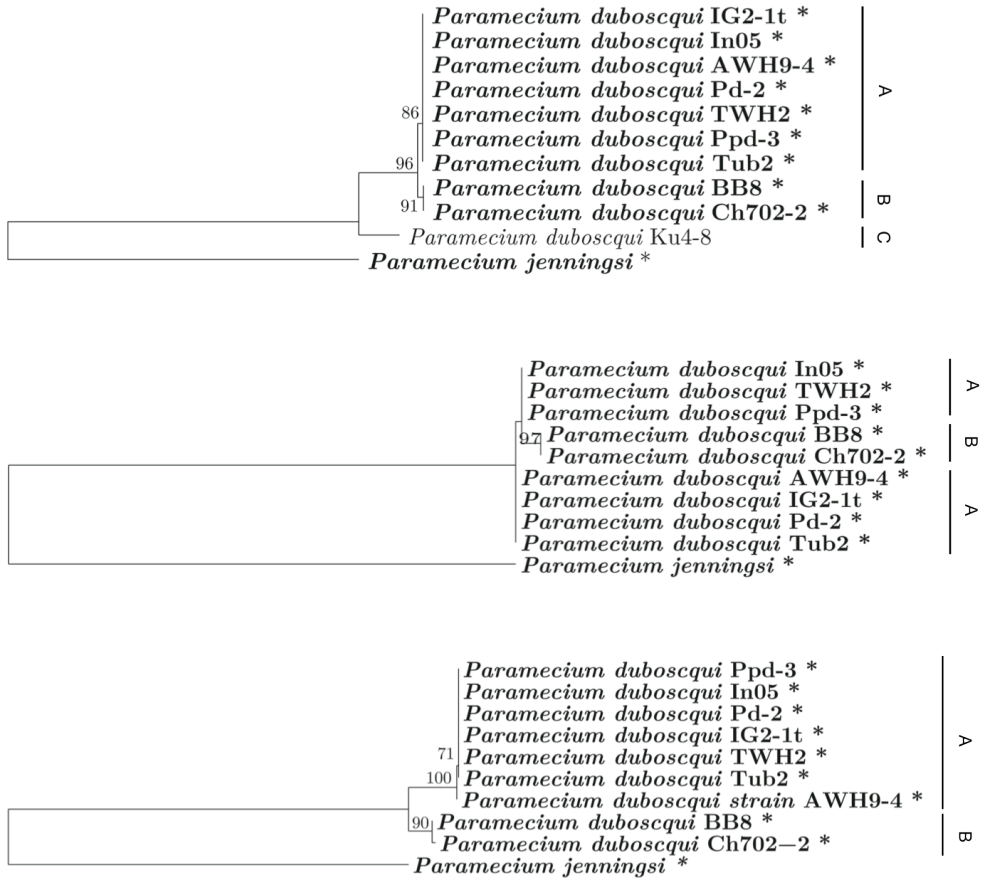

Fig 5.2. From the top: SSU rRNA, ITS1+5.8S+ITS2 and *cox1* gene sequences Maximum Likelihood tree. The node values represent ML bootstrap values cut as explained in section 2.4.8. Asterisks mark sequences obtained in this work. Supplementary information are provided in the text.

5.3.3. mt LSU rRNA gene sequences

Primers for amplification of a partial mt LSU rRNA gene sequence were developed exploiting the available sequences of some complete mitochondrial ciliate genome, like those of *Paramecium caudatum* (NC014262, Barth & Berendonk, 2011), *Paramecium aurelia* (NC001324, Burger et al., 2000), *Euplotes minuta* (GQ903130, de Graaf et al., 2009), *Tetrahymena termophyla* (NC003029, Brunk et al., 2003) and *Tetrahymena pyriformis* (NC000862, Edqvist et al., 2000). These sequences were aligned using the ARB software package and manually edited.

Partial mtLSU gene sequences from *P. duboscqui* were successfully obtained only for strains BB8 and In05, through PCR and direct sequencing [amplification primers: MLSU_F768Par and MLSU_ R2057Par; thermal pattern: 35x 94°C (30''), 57°C (45'') & 72°C (120''); sequencing primers: MLSU_F1824Par and MLSU_R1958Cili]. Only the optimal quality portion of the electropherograms were kept and used for subsequent analyses.

Partial mtLSU gene sequences of *P. jenningsi* and *P. aurelia* were obtained through PCR and direct sequencing [amplification primers: mtLSU_F548Par and mtLSU_R2436Cili; thermal pattern: 35x 94°C (30''), 50°C (40'') & 72°C (120''); sequencing primers: mtLSU_F670, mtLSU_F1776 and MLSU_R2057Par]. The electropherograms obtained using mtLSU_F670 were of bad quality, showing a pattern of double peaks after about 100 unambiguous bases. Only the good quality portions was utilised in subsequent analyses.

The partial sequence from *P. schewiakoffi* strain SH1-38 was obtained through two PCR and direct sequencing [amplification primers: MLSU_F768Par and MLSU_R2057Par; thermal pattern: 35x 94°C (30''), 57°C (45'') & 72°C (120''); sequencing primers: MLSU_F1824Par and MLSU_R1958Cili. amplification primers: mtLSU_F548oligo and mtLSU_R2436cili; thermal pattern: 5x 94°C (30''), 53°C (30'') & 72°C (150'') 10x 94°C (30''), 50°C (30'') & 72°C (150'') 30x 94°C (30''), 47°C (30'') & 72°C (150''); sequencing primers: mtLSU_F1776].

The same apply to *P. caudatum* strain CG6 [amplification primers: MLSU_F768Par and mtLSU_R2436cili; thermal pattern: 5x 94°C (30''), 53°C (30'') & 72°C (150'') 10x 94°C (30''), 50°C (30'') & 72°C (150'') 30x 94°C (30''), 47°C (30'') & 72°C (150''); sequencing primer: MLSU_R2057. amplification primers: mtLSU_F548oligo and mtLSU_R2436cili; thermal pattern: 35x 94°C (30''), 50°C (40'') & 72°C (120''); sequencing primers: mtLSU_F1776oli].

For the subsequent phylogenetic analysis (Fig. 5.3), I employed a total of 15 sequences: 6 from this study and 9 from complete ciliate mitochondrial genomes (of genera *Paramecium*, *Tetrahymena* and *Euplotes*).

All sequences were automatically and then manually aligned using the ARB software package. After trimming the sequences ends to the shorter

one, the analyses were conducted on 774 characters. AIC and BIC tests, performed with the ModelTest software, gave TVM+G and TPM3uf+G as preferred evolutionary models output. Likelihood mapping analysis performed with TreePuzzle showed that 94,7% of all the quartet trees were in the resolved regions, thus suggesting that the data-set was adequate for phylogenetic analyses.

ARB NJ, PHYLIP DNAPARS, PHYML (with the tools provided by the ARB package), TreePuzzle and MrBayes were used for producing NJ, MP, ML, quartet puzzling and Bayesian tree respectively. GTR+I+G was the model used in all model-based methods, with the exception of ARB NJ, in which the felsenstein (Felsenstein, 1981) correction was used instead. 1000 bootstrap pseudoreplicates were performed for ARB NJ, 500 for MP and ML trees. 100.000 puzzling step were performed in order to obtain the quartet puzzling support values, and posterior probability values are derived from 1.000.000 generations (burnin 25%) for each of the 3 runs of MrBayes (each consisting of 3 heated and one cold chain).

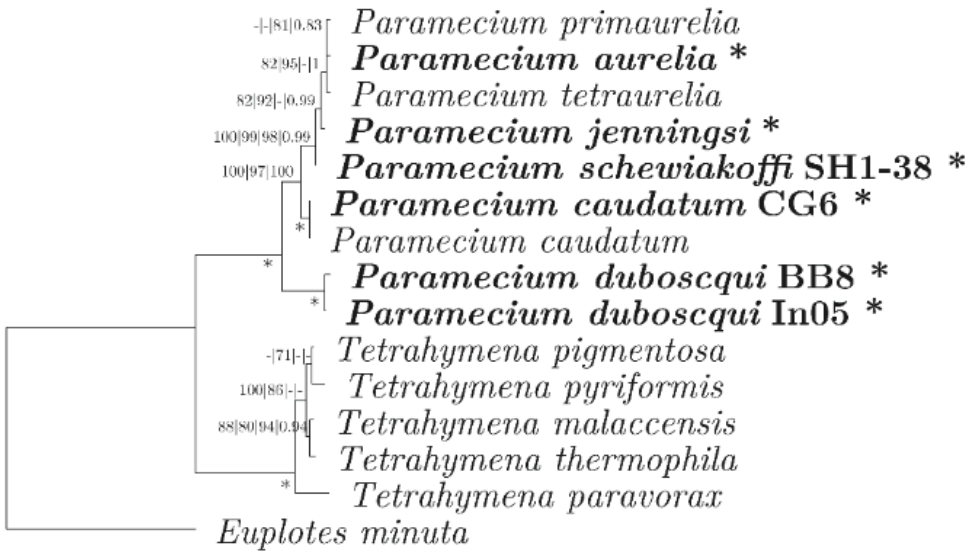

Fig 5.3. mt LSU rRNA Maximum Likelihood tree. 15 sequences and 774 characters were employed. The node values represent MP and ML bootstrap values, TreePuzzle quartet puzzling support and Posterior Probability respectively, cut as explained in section 2.4.8. NJ bootstrap values are not shown. Asterisks mark sequences obtained in this work, and node supported by all values. Supplementary information are provided in the text.

The obtained phylogenetic tree of genus *Paramecium* is congruent to others published in literature using SSU rRNA gene sequences (e.g. Struder-Kypke et al., 2000). Monophyly of the genus, of the *P. duboscqui* strains and the *Paramecium* subgenus (*sensu* Fokin et al., 2004) are highly supported. The only difference from published data on SSU rRNA is that *P. jenningsi* and *P. schewiakoffi* don't associate to each other. *P. jenningsi* clusters inside with the aurelia sequences, while *P. schewiakoffi* is in a more basal position.

5.4. Discussion

Before trying to evaluate the results, it is important to revise again the *P. duboscqui* strains directly or indirectly involved in this survey. Strain Ku4-8 was not available for further analyses, but its SSU rRNA gene sequence is published (Schrallhammer et al., 2006) and was compared with the others. It was sampled in Kunashir Island, like another monoclonal strain called Ku4-11; both were present in the '99 paper by Fokin et al. Strain AWH9-4 cam from the atlantic coast of the United States, and was present in the '99 paper together with another strain from the same sampling location, called AWH9-3. A similar situation apply for strain IG2-1 from Giglio Island; in the paper, it was associated to IG2-2. Strains BB8 and BB1-11 share exactly the same history and were collected in the Gulf of Finland, while strains 702 and 101 came from eastern China. Strains WL2-6 (White Sea) and GN3-3 (Baltic Sea) are the only one present in the '99 paper for which no subsequent data are available. Strains Tub2, Pd-2, Ppd-3 (from Italy) and TWH2 (from Tunisia) are exclusive of the present study.

Fig. 5.4 summarize the results obtained by RAPD approach in the '99 paper. It is useful to confront it with phylogenetic trees in Fig 5.2.

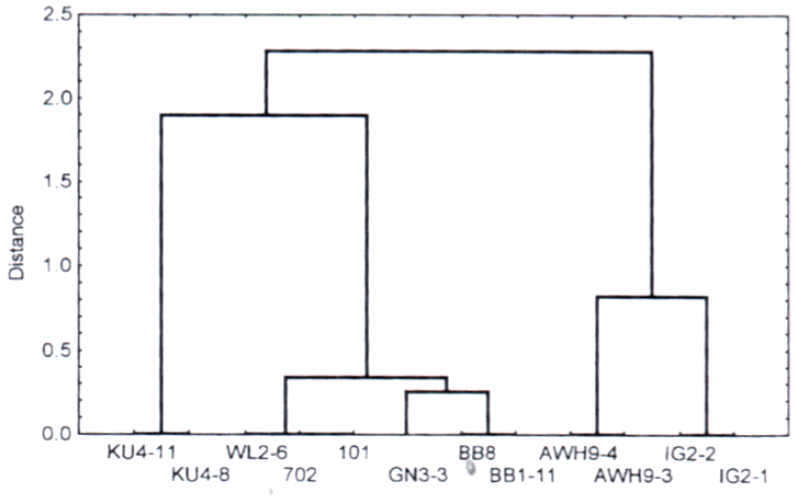

Fig 5.4. Cluster analysis performed through UPGMA method on Euklid's distances between the similarity index of RAPD patterns (from Fokin et al., 1999). Strains Ku4-8, 702, BB8, AWH9-4 and IG2-2 were also surveyed in the present work.

SSU rRNA gene sequences allow the discrimination of three different groups/clades, that I called A, B and C respectively. Authors of the '99 paper also identified three major clades through their RAPD analysis, assumed that they were probably sibling species and called them *P. duboscqui I*, *P. duboscqui II* and *P. duboscqui III*. The classification of the strains present in both works lead me to identify the A, B and C groups to the *P. duboscqui* putative biological species *I*, *II* and *III* respectively. From this point onwards, I will use my own nomenclature, without assuming that each clade form a separated species.

SSU rRNA and RAPD data gave perfectly comparable results. The phylogenetic inference is slightly different: in the original cluster analysis on RAPD distances groups B and C resulted the most closely related, while in my SSU rRNA ML phylogenetic inference group C is the most basal. Because my analysis is based on much more informative characters and includes an outgroup, it is probably more reliable.

These major clades inhabit distinct geographic areas (Fig 5.5). Group C is exclusive of Kunashir Island, politically belonging to Russia but actually situated north to Japan. Group B can be found in the northeast of China amd in northern Europe (in the Finland Gulf, and also in Baltic Sea and White Sea, if my inclusion of strains WL2-6 and GN3-3 is correct). There are no other asiatic strains, but one prevision can be made:

any *P. duboscqui* strain collected from central- or northern Asia brackish waters should belong to group B; its SSU rRNA gene sequence should be identical to those of the other strains of this group, or at least cluster with them in phylogenetic analysis. Thus, my data not only can supply predictive hypotheses, but are also well suited to test them. The comparison with RAPD results was possible only because the very same strains (or at least most of them) were exploited in two different works.

Group/clade A can be found on the atlantic coast of the United States and on both sides of the Mediterranean Sea (represented by Italy and Tunisia). The tunisian strain added in this study is precious for extending the biogeographical conclusions. Nevertheless, previsions on future affiliations of strains from Europe have to be cautious. Clade B should be present on the northern coasts, but clade A is probably distributed in all the Mediterranean and perhaps also on western coasts of Europe and northern Africa.

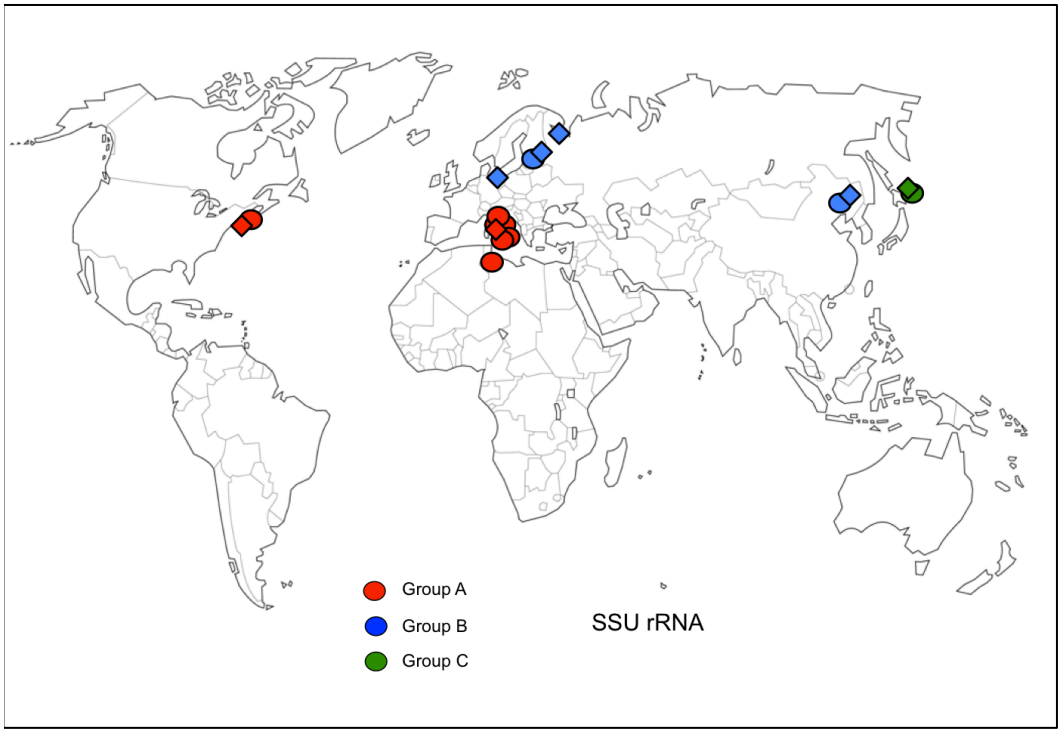

Fig 5.4. Biogeographic pattern of the *P. duboscqui* strains based on SSU rRNA gene sequences. Circles represent strains surveyed through sequence markers in this study. Squares represent strains surveyed only through RAPD in Fokin et al., 1999. The attribution of these strains to the three major clades is hypothetical.

Geographic separation can reflect an old migration event from Eastern Asia towards Europe and America. The origin of the event is deduced by the most ancient split in the phylogenetic tree, between clade C and clades A/B. It can also be a consequence of a niche separation triggered by environmental temperature, because strains from groups A and B are separated by a latitudinal gradient. The latter hypothesis is a mere speculation, because actual temperature sampling are not available for all strains and were not repeatedly measured in the different environments. On the basis of the measures available values are moreover similar between the two groups, averaged at 13,5°C.

ITS1+5.8S+ITS2 sequences gave no additional insights with respect to the SSU rRNA marker. It possesses a slightly (probably non-significant) lower similarity value and a base polymorphism in the group A strains (even inside the set of the Italian strains). These polymorphism is with all evidence the reason why this marker cannot properly recover the two-clade split between groups A and B in phylogenetic analysis. The paraphyly of group A is not supported by bootstrap values, and comparisons with the other two markers and the biogeographic pattern makes this result unreliable. Like in the first project (see Chapter 3), ITS1+5.8S+ITS2 turned out to be the less useful of the three markers. Moreover, it is not available for group C strains.

cox1 gene sequence differs from the nuclear markers in its capabilities of discrimination, as should be expected by the properties of this mitochondrial gene. In phylogenetic analysis, the two major groups A and B can be recovered (unluckily, data from clade C are missing). But the sequences of the strains are not all identical inside them. BB8 and 702 on one side, and AWH9-4 and the Mediterranean strains on the other, form a total of 4 haplotypes. Only considering data from my own research, these haplotypes should be called “northern european”, “chinese”, “american” and “mediterranean” respectively. I will label them B1, B2, A1 and A2 respectively.

But perhaps further assumptions can be made (Fig. 5.5). It seems reasonable to think that *cox1* gene sequence has at least the same discrimination capabilities of RAPD analysis. In the ’99 paper, only strains from the same sampling location were not distinguishable (namely, they had the same RAPD pattern). The only exception was WL2-6, from the Baltic Sea german coast, that shared the same RAPD pattern with chinese

strains 702 and 101. I will thus make the reasonable speculation that the "chinese" haplotype extends to the eastern coasts of Scandinavia, in the White Sea. It is also to be noted that the similarity between B2 ("chinese", with 702) and B1 (Gulf of Finland, BB8) haplotypes is very high.

Using similar assumptions, the german strain GN3-3, that RAPD analysis can tell apart from the Finland strains, possibly belongs to a fifth haplotype (B3) or to haplotype B1 itself. In any case, the major gradient step inside group B is situated between the western and eastern sides of Scandinavia (this conclusion, I must stress, heavily relies on the RAPD analysis).

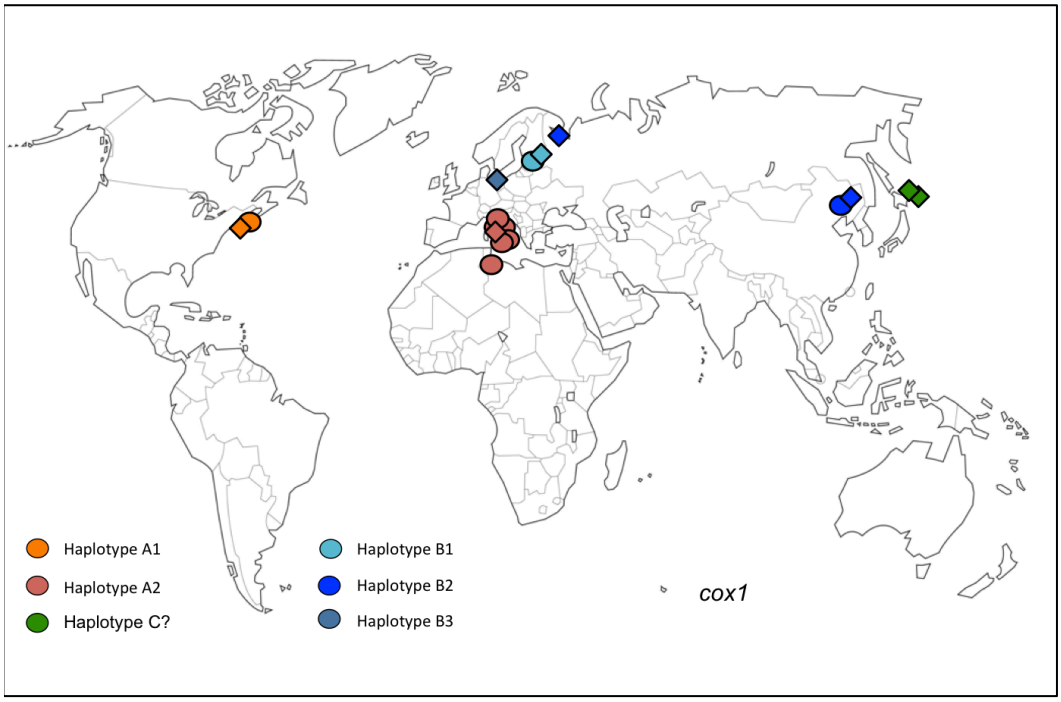

Fig 5.5. Biogeographic pattern of the *P. duboscqui* strains based on *cox1* gene sequences. Circles represent strains surveyed through sequence markers in this study. Squares represent strains surveyed only through RAPD in Fokin et al., 1999. The attribution of these strains to the haplotypes is hypothetical. Haplotype C is only theoretical.

Less can be said about group A. Like the RAPD analysis, but unlike the nuclear marker, *cox1* sequencing can discriminate between the american and mediterranean strains. It is interesting to note that the Italian and Tunisian strains all share the same haplotype, thus suggesting a somewhat recent separation. Moreover, the american strain has just one

base difference with the mediterranean ones. The average difference between A and B group consists instead of 42 bases, a much more impressive separation. It is likely that the similarity with the C group would have been even lower, and hopefully further studies will obtain new strains from the Pacific area in order to test this prevision.

From a systematic and biogeographic point of view, the picture is clear. The questions of elevating the three groups to the biological species status is however yet to be answered.

Molecular markers, even gene sequences, cannot give a definitive answer. As explained above, a threshold value of similarity below which new taxa can be established has never been fixed for eukaryotes. The barcoding properties of *cox1* could be used in the future for identification purposes, but to exploit them in order to establish new species would be very unwise. Only mating experiments will resolve this issue.

Nevertheless, a correlation between sequences divergence and sexual isolation probably exists. It is interesting, then, to make some comparisons. In Tab 5.2 the dissimilarity values for two strains of *P. duboscqui* from the clades A and B are showed, alongside the homologue sequences of *P. schewiakoffi* and *P. jenningsi* (two closely related but morphologically different specie), some *P. aurelia* biological species (selected in order to give an idea of the differences range in the complex) and two strains of *P. multimicronucleatum* suggested to belong to sibling species only because of their molecular difference (Barth et al., 2006).

It can be seen that the dissimilarity value between clade A and B inside *P. duboscqui* is well in the range that can be found among *aurelia* species. It is also similar to that between the two *P. multimicronucleatum* putative sibling species.

As already stressed, these data cannot prove that *P. duboscqui* is indeed a complex of sibling species (at least one from Asia and Northern Europe and the second from America and the Mediterranean), but surely suggest to test this hypothesis with further researches. Besides, data are missing for the most divergent group, that I called C; if the correlation between molecular and mating data would turn out to be true, this third clade could become even more likely a separated species. By the way, the biogeographic pattern detected in this work would be a support to future hypotheses of allopatric speciation.

A final and separate discussion has to be made for the mt LSU rRNA data from the *Paramecium* genus. This project is at a very preliminary phase, and thus only preliminary conclusions can be listed.

First of all, I was successful in obtaining primers that work at least for genus *Paramecium*, as proved by the partial sequences extracted from 6 different sources. The comparisons in Tab 5.3 show that evolutionary rate for this gene seems to be lower than that of *cox1*, but higher than those of the nuclear marker SSU rRNA, at least in the studied systems. This is a first interesting property; *cox1* is not usually suitable for phylogenetic inferences above the genus level, while nuclear markers are too conserved for intraspecific studies. mt LSU rRNA seems to possess intermediate qualities.

Pairs	**SSU rRNA**	***cox1***	**mt LSU rRNA**
P. duboscqui strains In05/ BB8	0,1%	6,1%	2,3%
P. jenningsi / *P. schewiakoffi* strain SH1-38	0,5%	14,7%	1,1%
P. jenningsi / *P. caudatum* strain CG6	4,7%	24,6%	8,8%
P. caudatum strain CG6 / *P. duboscqui* strain In05	8,4%	26,6%	14,6%

Tab 5.3. mt LSU rRNA gene sequences dissimilarity values of selected paramecium specimens pairs.

Because it was obtained only from two *P. duboscqui* strains (In05, from clade A, and BB8, from clade B) no further considerations can be made on the questions above. Values from *P. aurelia* sibling species are not available for comparisons, but it is interesting to notice that the two morphologically distinguishable species *P. jenningsi* and *P. schewiakoffi* are more similar to each other than the two clade of *P. duboscqui*, using this marker.

Finally, a phylogenetic analysis was performed on all available ciliates mt LSU rRNA sequences. This is only an initial step in order to evaluate

the possibility of this marker for phylogenetic inferences; however, it is good to notice that the tree topology for the *Paramecium* genus is nearly the same of the classical systematics based on SSU rRNA sequences.

mt LSU rRNA is a molecular marker whose possibilities have yet to be explored in depth. My preliminary results seem to induce a cautious optimism, though. It shares many good features with both nuclear rRNA genes and mitochondrial genes. And it also has a great advantage on the protein-coding genes like *cox1*: the possibility of being the template for nucleotide FISH probes. One application that could be exploited in the next future on the *P. duboscqui* system is the development of clade-specific probes for strains BB8 (clade B) and In05 (clade A). Nuclear rRNA sequences are in fact not different enough to be used as good target for specific probes, and *cox1* is a protein-coding gene present in few copies that would give a weak signal. The development and success of these clade-specific probes would be an additional contribution on the process of establishing new species.

6. Conclusions

Each one of the three projects is a slightly different application of the multi-marker molecular characterization approach to a ciliate/ endosymbiont system.

The first one is mainly focused on the bacterial endosymbiont found in a *Paramecium jenningsi* population from Thailand. Soon identified as a HLB, this organism has been studied with two parallel approaches, molecular (this study) and morphological/biological. The aim was to carefully characterize an organism belonging to a known group in order to add data and considerations useful for evaluating the entire group. The conclusions reached are:

1 – The HLB SSU rRNA gene sequence is closely related to those of the other HLB publicly available or previously obtained in our lab. All of them form a monophyletic group.

2 – The similarity with the sequences of the *Holospora* species is low enough to allow the establishment of a new genus for this HLB. Biological features like the absence of the connecting piece induction support this hypothesis. We propose the name "*Candidatus* Gortzia infectiva" gen. nov. sp. nov.

3 – *Holospora caryophila* SSU rRNA gene sequence is even more dissimilar from the others, thus requiring an additional genus according to the prokaryotic taxonomy rules. Biological features support this differentiation too. Thus, all HLB molecularly characterized until now can be classified in three different genera.

4 – Molecular and experimental data both support – but don't prove definitely – that *Gortzia infectiva* has a narrow host specificity, and is restricted to the macronucleus of *P. jenningsi*. This character, along with all the others obtained from the jointed characterizations, will be used as a diagnostic feature in the future formal description of this bacterium.

The second project is concerned with both ciliates and a putative endosymbiont. The emphasis is placed on the phylogenetic analyses, unfortunately carried on just one molecular marker (the SSU rRNA gene sequence) for reasons of limitations of the database. On the ciliate side, the aim was resolving the phylogenetic relationships inside genus *Condylostoma*, obtaining a systematic framework independent from the morphological one. Concerning the bacterium, the interest resided in its novelty, having it been found in unusual environment and host. The conclusions can be so summarized:

1 – The genus *Condylostoma* is not monophyletic unless the ciliate *Chattonidium setense* (now formally included in its own family) gets included in it. This was showed also in works from other authors.

2 – The molecular phylogenetic analysis suggests the presence of three clades inside the genus *Condylostoma* (including *Chattonidium*). There are currently no available morphological apomorphies for these clades. Moreover, the molecular results are not coherent with the morphological one. One possible explanation is that the morphological diagnostic characters traditionally employed for identifying the species in this genus are unreliable.

3 – Phylogenetic analysis conducted on the entire class Heterotrichea produced some robust results: the class itself is monophyletic, and so are all of the genera examined (except *Condylostoma*). Some families, especially Climacostomidae, are clearly paraphyletic. On the contrary, family Condylostomatidae and the cluster of family Stentoridae, Blepharismidae, Maristentoridae, Folliculiniidae plus *Fabrea salina* are highly supported by bootstrap values.

4 – There are reasons to think that strain COL2 harbours a bacterial endosymbiont. If the preliminary data are correct, this bacterium belongs to the order *Rhodospirillales* of class *Alphaproteobacteria*, and is related to the non-sulfur purple bacteria of family *Rhodospirillaceae*.

5 – The phylogenetic analysis of SSU rRNA sequences from order *Rhodospirilalles* gave very low support values for most nodes, included that sustaining the order itself. Families *Acetobacteraceae* and *Rhodospirillaceae*, as currently defined, are paraphyletic.

The third project didn't directly focus on endosymbionts, but on the molecular characterization of a ciliate species known to harbour them. It is the first molecular survey of the poorly known morphospecies *Paramecium duboscqui* assessed with modern markers. In addition to the results of this study, conclusions rest also upon comparisons with the literature, especially the paper of 1999 by Fokin et al. These conclusions are:

1 – All available molecular markers highlight the presence of three major clades inside the *P. duboscqui* morphospecies. Subclades of sequences differing by at least 1 base (haplotypes) can be identified using *cox1*.

2 – The major clades reflect the biogeographic pattern of the strains. A Pacific clade, an Asian-Northern European clade and an Atlantic-Mediterranean clade can be identified. *cox1* haplotypes (subclades) are more spatially limited.

3 – The most divergent strain is Ku4-8 from Kunashir Island, the only representative of the Pacific clade and the only strain known to contain a bacterial endosymbiont.

4 – Although molecular markers alone cannot prove the presence of a sibling species complex inside the *P. duboscqui* morphospecies, sequence similarity values between clades are lower than those between other identified species in the *Paramecium aurelia* group. This corroborates the hypothesis and suggests further researches.

In addition, two strains of *P. duboscqui* and specimens from other 4 *Paramecium* morphospecies were exploited for a preliminary evaluation of the properties of mt LSU rRNA gene sequence as a molecular marker. The preliminary results show that it is possible to develop useful primer pairs for this marker, at least for the genus *Paramecium*; that the evolutionary rate seems to be lower than that of *cox1*, but higher than that of nuclear genes; moreover, although the experiment has not been performed yet, oligonucleotide probes for FISH experiments can be designed, that are capable of discriminating between sub(morpho)specific taxa inside *P. duboscqui*.

Acknowledgements

After so many pages on ciliates and bacteria, I think I'm allowed to write one or two about human beings. The Master Thesis is supposed to be the last act of one phase of life, and the beginning of something else. Thus, I will make use of this opportunity to thank all of the people who played an important role, for one reason or another, in the last 5 years of my life.

Sadly, the thesis is primarily the product of hard work, so I will start with people of the lab or that were somehow involved in my scientific life. To list all of the merits and demerits of my supervisor, dr. Giulio Petroni, would take a whole book on its own, so I will constrain myself citing him. Thanks to his allowance of sincerity in human relations and my own outspokenness, he knows what I would write, given the space. Immediately after, and equally important for my scientific formation, I have to thank dr. Claudia Vannini, my former supervisor for the Bachelor Thesis and continuous support in the everyday lab life. I'm grateful to her also for letting me participate to some off-thesis projects that interested me. I've also worked with Prof. Sergei Fokin and dr. Letizia Modeo, benefiting from their knowledge in different areas. And I owe thanks also to Simone Gabrielli for his help and kindness. All of these work relationships were really important in enjoying the everyday life, giving the feeling of a united research group and the advantage of varying the objects of study and following several different projects in parallel.

And if one spends so many kind words for the teachers, he cannot be less grateful to the collegues. Beatrice Moretti and the already *dr.* (but yet *padawan*) Gabriele Tomei were the undergraduate students with which I actually worked… and joked and laughed and dined and screamed. But another great part of the entertainment came from two other master students, Alessandro Ristori (life in the first months would have been much less easy without him) and Matteo Meliani (life in the last months would have been just as easy, but a lot less funny, without him), and our

general commander and PhD student dr. Carolina Chiellini (hi Caro, I know you're reading!).

And if it is true that I already knew the lab and its environment, during this last year I really really really enjoyed another side of the scientific life: congresses. They were really amusing and enthralling. All nice memories: Pisa, Roscoff (what a place!) and Berlin. So, I want to collectively thanks all the interesting people I met there, and especially Prof. Linda Sperling, for being so kind with me and other students in all the three occasions… and after! It is also fair to thank the Scuola Normale Superiore of Pisa for actually paying my participation to those congresses and Filippo Lipparini, a *precious* friend who suggested me to *ask* for those moneys.

Enough for the scientific hours of the day. Thankfully, there were still many others. And for them I have so many people to thank, starting with Antonio Casini (you've saved me again!), Michele Felletti (we made it, Mik!) and the rest of the Pack: Lorenzo Romeo, Beatrice Pozzetti, Nicola Pierazzo, Franco Egidi, Fabio Miazzi, Giulio Belletti, Luca Battistella and Dario Trevisan (before he went crazy – sad things happen!), then Pietro D'Amelio for always inflating my ego and Umberto Binetti for his old habit of deflating it, and finally some lost companions like Luca Tosti, Riccardo Sarti and Dalila Colucci. Oh, also a special thank to Enzo Busseti; he's much less pleasant than all of them, but as a matter of fact he made me met them all! And of course, there is love and affection also outside Pisa: Teo, Tia, Totta and Luca are still there (and I now think they will be for some time still).

Last but not least come Dad, Mom and Valeria, more supporter than ever in this last heavy months. They still don't understand exactly what a protist is, but that's not a big problem… neither do most of my cooler molecular biologists friends.

In the end, in homage to biodiversity, some non-human thanks. The first is for Pisa. I know you're small, stinky and not that funny, but I irrationally love you anyway. There's not a single, tiny spot of this town that doesn't have its memories. Good and bad. The second and last for all those tiny, loving furry ciliates and their internal deseases I've studied. Sorry for all that I've done to you!

Appendix

A.I. Protocols

A.I.1. Protocol: PCR

1. Thaw on ice:
    - Distilled and sterile water
    - 10x exTaq buffer solution (TaKaRa®)
    - 10x dNTP mix (TaKaRa®)
    - Primers aliquote (100 pmol/µl)
2. For each reaction, mix in Eppendhorf 0.2 ml:
    - 36.7 µl of water
    - 5 µl of 10x exTaq buffer
    - 5 µl of 10x dNTP mix
    - 0.25 µl of each primer
    - 2.5 µl of each primer
    - 2.5 µl of DNA solution
    - 0.3 µl of exTaq (TaKaRa®)
3. Place each Eppendhorf in the Thermal Cycler with the desired thermal pattern set and start the cycles

A.I.2. Protocol: Cloning and RFLP analysis

1. If necessary, add a poly-A single strand to the 3' end of the amplification product (see section 2.2.3)
2. Mix:
   - 4 μl of purified PCR product
   - 1 μl of plasmidic vector (pCR®2.1-TOPO®
   - 1 μl of saline solution (NaCl 1.2M, $MgCl_2$ 0.06M)
3. Incubate at room temperature for 15'
4. Place competent *E. coli* cells (Mach1-T1® stored at -80°C) in ice
5. Add the solution with the plasmidic vector on the competent cells
6. Incubate in ice for 20'
7. Apply a heat shock, placing for 30'' the bacterial cells at 42°C
8. Add 200 μl of S.O.C. medium
9. Incubate at 37°C for 1 hour
10. Plate *E. coli* cells on Petri dishes with LB-agar medium, ampicillin (25 μl of 50 mg/ml in acqueous solution) and X-gal (25μl of 50 mg/ ml solution in DMSO). It is suggested to employ three dishes and plate 20, 40 and 60 μl of the bacterized medium respectively
11. Incubate overnight at 37°C
12. Streak the bacteria from each white colony on a Master Plate subdivided and numbered, and incubate it until the bacteria grow
13. Perform the RFLP screening (steps 14-19)

14. Perform a control PCR on each bacterial clone growing on the Master Plate with primers M13F and M13R
15. Identify through electrophoretic run clones that contain the desired insert
16. For each useful clone, mix in Eppendhorf:
    - 6 μl of control PCR product
    - 0.8 μl of enzyme buffer
    - 0.2 μl of the enzyme BsuRI (Fermentas®)
17. Incubate overnight at 37°c
18. Identify through an electrophoretic run (on an agarose gel 2%) the digestion patterns
19. Once chosen the fit clones, inoculate them from the Master Plate into liquid medium (5 ml of LB medium + 5 μl of ampicillin 50 ng/μl)
20. Incubate at 37°C overnight
21. Extract the plasmidic DNA from an adequate volume of bacterized culture medium

A.I.3. Protocol: FISH

1. Pipet on the cells fixed on slide:
    - 40 μl of hybridization solution
    - 5 μl of probe solution (50 ng/μl)
2. Cover with cover slip
3. Incubate at 46°C for at least 3 hours in wet chamber (a close container with paper strips impregnated with hybridization solution

4. Remove the cover slip gently shaking the slide in a beaker with washing solution preheated at 48 °C
5. Place the slide in a jar for the washing steps:
    - 10' in washing solution at 48°C
    - 10' in washing solution at 48°C
    - 15' in distilled water at room temperature
6. Wait until the slide dries under chemical flow
7. Pipet on the cells 20 µl of SlowFade® Gold with DAPI (Invitrogen)
8. Cover with cover slip and fix it with nail polish
9. Wait until the nail polish dries and keep the slide at 4°C

WARNING: all steps should be performed under the lowest illumination possible

In the low stringency protocol I usually performed in this work, hybridization solution and washing solution are the same. 1 l of such solutions contains:

- 800 ml of distilled water
- 180 ml of NaCl 5M
- 20 ml of TrisHCl 1M, pH 8

A.II. Primers and probes

Primer name	Sequence	Reference
T7	5'-TAATACGACTCACTATAGGG-3'	-
M13R	5'-CAGGAAACAGCTATGAC-3'	-
18S F9	5'-CTGGTTGATCCTGCCAG-3'	Medlin et al., 1988
18S R1513 Hypo	5'-TGATCCTTCYGCAGGTTC-3'	Petroni et al., 2002
18S R536	5'-CTGGAATTACCGCGGCTG-3'	Rosati et al., 2004
18S F783	5'-GACGATCAGATACCGTC-3'	Rosati et al., 2004
18S R1052	5'-AACTAAGAACGGCCATGCA-3'	Rosati et al., 2004
18S F919	5'-ATTGACGGAAGGGCACCA-3'	Unpublished
RGD2	5'-GGTCCGTGTTTCAAGACGGG-3'	Unpublished
FG1400	5'-TTGYACACACCGCCCGTC-3'	Unpublished
cox1fT7_Cil_IMVFF	5'-TAATACGACTCACTATAGGGGTTGTTA CTGCTCATGGTTTTATHATGGTNTTYTT-3'	This study
cox1rM13R_Cil_HDT F	5'-CAGGAAACAGCTATGACATGGAAGT GTGCTACTACATARAANGTRTCRTG-3'	This study
cox1fT7_DVA(FY)	5'-TAATACGACTCACTATAGGGTAATACG ACTCACTATAGGGGATGTWGCNTWYCC-3'	This study
cox1_Rs_Olig_AGWT	5'-TTTGAATTGAAGGGTGT TATAAAAGTYCANCCNG-3'	This study
cox1_F543_SKmP	5'-TAATACGACTCACTATAGGGggttttgctaattttt taattccttatcatgTTGGWKCBAAAGATGTWGC-3'	This study; Strüder-Kypke & Lynn, 2010
cox1_R1321_SKmP	5'-CAGGAAACAGCTATGACgttggtataattataat aTADACYTCAGGGTGACCRAAAAATCA-3'	This study; Strüder-Kypke & Lynn, 2010
COIfT7_Oligo_INHK	5'-TAATACGACTCACTATAGGGTTTTTAA GAAGTATGTTTATACTATHAAYCAYAA-3'	This study
PdubR1088	5'-GTTCTTCTAGTAATT AATAAATTMGTAAAAG-3'	This study
PdubF922	5'-TCAAGAAGAMGAAAAATW TATTTTAGTAAATGTTC-3'	This study
1492R	5'-GGNWACCTTGTTACGACTT-3'	Lane, 1991
16S Alpha F19a	5'-CCTGGCTCAGAACGAACG-3'	Vannini et al., 2004
16S R1488 Holo	5'-TACCTTGTTACGACTTAACC-3'	Unpublished
16S R515ND	5'-ACCGCGGCTGCTGGCAC-3'	Vannini et al., 2004
16S F343ND	5'-TACGGGAGGCAGCAG-3'	Vannini et al., 2004
16S F785ND	5'-GGATTAGATACCCTGGTA-3'	Vannini et al., 2004
MLSU_F456Uni	5'-AGT ACC GTG AGG GAA AGG TGA A-3'	This study
MLSU_R2057Par	5'-AGTAAAGGTGCATAGGGTCTTTC-3'	This study
MLSU_F768Par	5'-TGTGATTAGGGGTGAAAGGCT-3'	This study

MLSU_F1824Par	5'-ATCTGACTCCTGCCCGGTGYT-3'	This study
MLSU_R1958Cili	5'-TCGTTACTCCATTCRTGCAGGAC-3'	This study
mtLSU_F670	5'-ACCCGAAGYYAAGTGATC-3'	This study
mtLSU_F1776	5'-GTTTAMTAAAAACATAAGATTTTGC-3'	This study
mtLSU_F548oligo	5'-TTGTNTTAACGTACCTTTTG-3'	This study
mtLSU_R2436cili	5'-TGTTATCCCTAGCGTACC-3'	This study
MLSU_R2595cili	5'-TAGGGACCAAACTGTCTCAC-3'	This study

Probe name	Sequence	Reference
Eub338_fluo	5'-GCTGCCTCCCGTAGGAGT-3'	Amann et al., 1990
ALF1b	5'-CGTTCGYTCTGAGCCAG-3'	Manz et al., 1992
Gortzia659_Cy3	5'-TTCCGTTTTCCTCTACCA-3'	This study
COL2_1249	5'-GGATTGCTAAGCCCATTGT-3'	This study

A.III. Sequences from trees

OTU	Reference	Accession
"Candidatus Anadelfobacter veles"	Vannini et al., 2010	FN552695
"Candidatus Captivus acidiprotistae"	Baker et al., 2003	AF533506
"Candidatus Cryptoprodotis polytropus"	Vannini et al., 2010	FM201295
"Candidatus Cyrtobacter comes"	Vannini et al., 2010	FN552697
"Candidatus Midichloria mitochondri"	Beninati et al., 2004	AJ566640
"Candidatus Odyssella thessalonicensis"	Birtles et al., 2000	AF069496
"Candidatus Paraholospora nucleivisitans"	Eschbach et al., 2009	EU652696
Acetobacter estunensis	Lisdyianti et al., 2000	AB032349
Acetobacter indonesiensis	Lisdyianti et al., 2000	AB032356
Acetobacter peroxydans	Lisdyianti et al., 2000	AB032352
Acidicaldus organivorans	Johnson et al., 2003	AY140238
Acidiphilium organivorum	Kishimoto et al.,1995	D30775
Acidiphilium rubrum	Kishimoto et al.,1995	D30776
Acidisoma sibiricum	Belova et al., 2009	AM947653
Acidisoma tundrae	Belova et al., 2009	AM947652
Acidisphaera rubrifaciens	Hiraishi et al., 2000	D86512
Acidocella aminolytica	Kishimoto et al.,1995	D30771
Acidocella facilis	Kishimoto et al.,1995	D30774
Acidomonas methanolica	Yamashita et al., 2004	AB110702
Ameyamaea chiangmaiensis	Yukphan et al., 2009	AB303366
Anaplasma marginale	Rurangirwa et al., unpublished	AF309866
Asaia bogorensis	Yamada et al., 2000	AB025928
Asaia krungthepensis	Yukphan et al., 2004	AB102953
Azospirillum irakense	Xia et al., 1994	Z29583
Azospirillum oryzae	Xie & Yokota, 2005	AB185396
Azospirillum rugosum	Young et al., 2008	AM419042
Belnapia moabensis	Gundlapally & Garcia-Pichel, 2006	AJ871428
Blepharisma americanum	Schmidt et al., 2007a	AM713182
Blepharisma elongatum	Schmidt et al., 2007a	AM713186
Blepharisma hyalinum	Schmidt et al., 2007a	AM713184
Blepharisma japonicum	Schmidt et al., 2007a	AM713185
Blepharisma steini	Schmidt et al., 2007a	AM713187
Blepharisma undulans	Schmidt et al., 2007a	AM713183
Brevundimonas mediterranea	Abraham et al., 1999	AJ227801
Caenispirillum bisanense	Yoon et al., 2007b	EF100694
Caulobacter henricii	Hamada et al., 1997	AB008532
Chattonidium setense	Modeo et al., 2006	AM295495
Climacostomum virens	Miao et al., 2009	EU583990
Condylostoma curva	Guo et al., 2008	EU379939
Condylostoma minutum	Guo et al., 2008	DQ822482
Condylostoma sp. Poe2.2	Modeo et al., 2006	AM295496
Condylostoma spatiosum	Guo et al., 2008	DQ822483

Condylostoma wangi	Miao & Song, unpublished	DQ445605
Condylostomides sp. SLS-200	Schmidt et al., 2007a	AM713188
Craurococcus roseus	Saitoh et al., 1998	D85828
Defluviicoccus vanus	Maszenan et al., 2005	AF179678
Devosia riboflavina	Vannini et al., 2004	AJ549086
Ehrlichia chaffeensis	Paddock et al., 1997	U86664
Elioraea tepidiphila	Albuquerque et al., 2007	EF519867
endosymbiont of Acanthamoeba sp. AC305	Hewett, unpublished	AY549548
endosymbiont of Acanthamoeba sp. EI3	Schmitz-Esser et al., 2008	AM408790
endosymbiont of Acanthamoeba sp. KA/E9	Xuan et al., unpublished	EF140635
endosymbiont of Acanthamoeba sp. S40	Matsuo et al., 2010	AB506682
endosymbiont of Acanthamoeba sp. UWC8	Fritsche et al., 1999	AF069963
endosymbiont of Acanthamoeba sp. UWET39	Horn et al, 1999	AF132139
Eufolliculina uhligi	Hammerschmidt et al., 1996	U47620
Fabrea salina	Miao et al., 2009	EU583991
Fodinicurvata fenggangensis	Wang et al., 2009	FJ357427
Fodinicurvata sediminis	Wang et al., 2009	FJ357426
Folliculina WWS-2008	Miao et al., 2009	EU583992
Geleia fossata	Droste & Epstein, unpublished	AY187925
Geleia simplex	Droste & Epstein, unpublished	AY187931
Geleia swedmarkii	Droste & Epstein, unpublished	AY187933
Gluconacetobacter intermedius	Boesch et al., unpublished	Y14694
Gluconacetobacter liquefaciens	Sievers et al., 1994	X75617
Gluconacetobacter saccharivorans	Trcek, 2002	AJ012466
Gluconobacter albidus	Takahashi et al., 2006	AB178392
Gluconobacter cerinus	Sievers et al., 1995	X80775
Granulibacter bethesdensis	Greenberg et al., 2007	CP000394
Gruberia sp.	Hirt et al., 1995	L31517
Holospora obtusa	Amann et al., 1991	X58198
Inquilinus limosus	Coenye et al., 2002	AY043374
Insolitispirillum peregrinus subsp. inte	Ding & Yokota, 2002	AB074521
Kentrophoros sp. QD061131	Gao et al., 2010	FJ467506
Kopriimonas byunsanensis	Kwon et al., unpublished	DQ167245
Kozakia baliensis	Lisdyianti et al., 2002	AB056321
Loxodes magnus	Hirt et al., 1995	L31519
Loxodes striatus	Andreoli et al., 2009	AM946031
Magnetospirillum gryphiswaldense	Schleifer et al., 1991	Y10109
Magnetospirillum magnetotacticum	Schleifer et al., 1991	Y10110
Marispirillum indicum	Lai et al., 2009	EU642410
Maristentor dinoferus	Miao et al., 2005	AY630405
Neorickettsia risticii	Weisburg et al., 1989	M21290
Neorickettsia sennetsu	Anderson et al.,1991	M73219
Neosaia chaingmaiensis	Yukphan et al., 2005	AB208549
Nisaea denitrificans	Urios et al., 2008	DQ665838
Nisaea nitritireducens	Urios et al., 2008	DQ665839
Novispirillum itersonii subsp. itersonii	Xia et al., 1994	Z29620
Oceanibaculum indicum	Lai et al., 2009 b	EU656113
Oceanibaculum pacificum	Dong et al., 2010	FJ463255
Orientsia tsutsugamushi	Ohashi et al., 1995	D38623
Paracraurococcus ruber	Saitoh et al., 1998	D85827
Parduczia orbis	Droste & Epstein, unpublished	AY187924

Pelagibius litoralis	Choi et al., 2009	DQ401091
Pelagicola litoralis	Kim et al., 2008	EF192392
Peritromus faurei	Miao et al., 2009	EU583993
Peritromus kahli	Rosati et al., 2004	AJ537427
Phaeospirillum chandramohanii	Anil Kumar et al., 2009	AM779061
Phaespirillum fulvum	Kawasaki et al., 1993b	D14433
Prosthecomicrobium consociatum	Yee et al., 2010	FJ560750
Remanella sp. R10	Andreoli et al., 2009	AM409181
Rhizobium leguminosarum	van Berkum et al., 1996	U29386
Rhodocista centenaria	Kawasaki et al., 1993a	D12701
Rhodocista pekingensis	Zhang et al., 2003	AF523824
Rhodopila globiformis	Hiraishi et al., 2000	D86513
Rhodospira trueperi	Pfenning et al., 1997	X99671
Rhodospirillum rubrum	Munk et al., 2011	CP000230
Rhodospirillum sulfurexigens	Anil Kumar et al., 2008	AM710622
Rhodovarius lipocyclicus	Kampfer et al., 2004	AJ633644
Rhodovibrio salinarum	Kawasaki et al., 1993b	D14432
Rickettsia bellii	Roux&Raoult, 1995	L36103
Rickettsia endosymbiont of Torix tukuban	Kikuchi & Fukatsu, 2005	AB113214
Rickettsia parkeri	Stothard & Fuerst, 1995	U12461
Rickettsia prowazekii	Weisburg et al., 1989	M21789
Rickettsia rhipicephali	Roux & Raoult, 1995	L36216
Roseococcus suduntuyensis	Boldareva et al., 2009	EU012448
Roseococcus thiosulfatophilus	Yurkov et al., 1994	X72908
Roseomonas aerilata	Yoo et al., 2008	EF661571
Roseomonas terrae	Yoon et al., 2007a	EF363716
Roseospira goensis	Kalyan Chakravarthy et al., 2007	AM283537
Roseospira navarrensis	Guyoneaud et al., 2002	AJ298880
Roseospirillum parvum	Glaeser & Overmann, 1999	AJ011919
Roseovarius crassostreae	Boettcher et al., 1999	AF114484
Rubritepida flocculans	Alarico et al., 2002	AF465832
Saccharibacter floricola	Jojima et al., 2004	AB110421
Skermanella aerolata	Weon et al., 2007	DQ672568
Skermanella xinjangensis	An et al., 2009	EU586202
Sphingobium scionense	Liang&Lloyd-Jones, 2010	EU009209
Sphingomonas kaistensis	Kim et al., 2007	AY769083
Spirostomum ambiguum	Schmidt et al., 2007b	AM398201
Stella umosa	Fritz et al., 2004	AJ535710
Stella vacuolata	Fritz et al., 2004	AJ535711
Stentor amethystinus	Di Giuseppe & Dini, unpublished	AY775566
Stentor cf. katashimai DB-201	Thamm et al., 2010	FN659818
Stentor coeruleus	Di Giuseppe & Dini, unpublished	DQ136037
Stentor elegans	Thamm et al., 2010	FN659817
Stentor muelleri	Thamm et al., 2010	FN659820
Stentor multiformis	Thamm et al., 2010	FN659821
Stentor polymorphus	Schmidt et al., 2007a	AM713190
Stentor roeseli	Gong et al., 2006	AF357913
Swaminathania salitolerans	Loganathan&Nair, 2004	AF459454
Tanticharoenia sakaeratensis	Takahashi & Nakagawa, unpublished	AB304087
Telmatospirillum siberiense	Sizova et al., 2007	AF524863
Terasakiella pusilla	Satomi et al., 1998	AB006768

Thalassobaculum litoreum	Zhang et al., 2008	EF203900
Thalassobaculum salexigens	Urios et al., 2010	EU008565
Thalassospira lucentensis	Swiderski, unpublished	AM294944
Thalassospira profundimaris	Liu et al.,2007	AY186195
Tistrella mobilis	Shi et al., 2002	AB071665
Trachelocerca ditis	Mazei et al., 2009	GQ167153
Tracheloraphis sp.	Hirt et al., 1995	L31520
uncultured alpha proteobacterium	Johnson et al., unpublished	FJ403068
uncultured alpha proteobacterium	Ye et al., unpublished	GQ250619
uncultured alpha proteobacterium	Loy et al., 1996	U65509
uncultured bacterium	Shimizu et al., 2007	AB294318
uncultured bacterium	Lesaulnier et al., 2008	EF019091
uncultured bacterium	Jones et al., 2008	EU137546
uncultured bacterium	Jones et al., 2008	EU137548
uncultured bacterium	Sunagawa et al., 2009	FJ202090
uncultured bacterium	Ashforth, unpublished	FJ716888
uncultured Trachelocercidae	Andreoli et al., 2009	AJ971527
uncultured Trachelocercidae	Andreoli et al., 2009	AJ971528
uncultured Trachelocercidae	Andreoli et al., 2009	AJ971531
uncultured Trachelocercidae	Andreoli et al., 2009	AJ971533
Wolbachia pipientis	Vandekerckhove et al., 1999	AF179630

Literature cited

- Abraham WR, Stroempl C, Meyer H, Lindholst S, Moore ERB, Christ R, Vancanneyt M, Tindall BJ, Bennasar A, Smit J, Tesar M (1999) **Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*.** *International Journal of Systematic and Evolutionary Microbiology* 49:1053-1073

- Alarico S, Rainey FA, Empadinhas N, Schumann P, Nobre MF, da Costa MS (2002) ***Rubritepida flocculans* gen. nov., sp. nov., a new slightly thermophilic member of the alpha-1 subclass of the *Proteobacteria*.** *Systematic and Applied Microbiology* 25:198-206

- Albuquerque L, Tiago I, Rainey FA, Taborda M, Nobre MF, Verissimo A, da Costa MS (2007) ***Salirhabdus euzebyi* gen. nov., sp. nov., a Gram-positive, halotolerant bacterium isolated from a sea salt evaporation pond.** *International Journal of Systematic and Evolutionary Microbiology* 57:1566-1571

- Amann RI, Binder BJ, Olson RJ, Chisolm SW, Devereux R, Stahl DA (1990) **Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations.** *Applied and Environmental Microbiology* 56:1919-1925

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

- An H, Zhang L, Tang Y, Luo X, Sun T, Li Y, Wang Y, Dai J, Fang C (2009) ***Skermanella xinjiangensis* sp. nov., isolated from the desert of Xinjiang, China.** *International Journal of Systematic and Evolutionary Microbiology* 59:1531-1534

- Anderson BE, Dawson JE, Jones DC, Wilson KH (1991) ***Ehrlichia chaffeensis*, a new species associated with human ehrlichiosis.** *Journal Clinical Microbiology* 29:2838-2842

- Andreoli I, Mangini L, Ferrantini F, Santangelo G, Verni F, Petroni G (2009) **Phylogeny of unculturable Karyorelictea (Alveolata, Ciliophora).** *Zoologica Scripta* 38:651-662

- Anil Kumar P, Aparna P, Srinivas TN, Sasikala Ch, Ramana ChV (2008) ***Rhodospirillum sulfurexigens* sp. nov., a phototrophic alphaproteobacterium requiring a reduced sulfur source for growth.** *International Journal of Systematic and Evolutionary Microbiology* 58:2917-2920

- Anil Kumar P, Srinivas TN, Takaichi S, Maoka T, Sasikala Ch, Ramana ChV (2009) ***Phaeospirillum chandramohanii* sp. nov., a phototrophic alphaproteobacterium with carotenoid glycosides.** *International Journal of Systematic and Evolutionary Microbiology* 59:2089-2093

- Aufderheide KJ, Daggett PM, Nerad TA (1983) ***Paramecium sonneborni* n. sp., a member of the *Paramecium aurelia* species complex.** *Journal of Protozoology* 30:128-131

- Baker BJ, Hugenholtz P, Dawson SC, Banfield JF (2003) **Extremely acidophilic protists from acid mine drainage host *Rickettsiales*-lineage endosymbionts that have intervening sequences in their 16S rRNA genes.** *Applied and Environmental Microbiology* 69 (9), 5512-5518

- 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

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

- Beier CL, Horn M, Michel R, Schweikert M, Görtz H-D, Wagner M (2002) **The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the *Rickettsiales* (*Alphaproteobacteria*) and to *Francisella tularensis* (*Gammaproteobacteria*).** *Applied and Environmental Microbiology* 68:6043-6050

- Belova SE, Pankratov TA, Detkova EN, Kaparullina EN Dedysh SN (2009) ***Acidisoma tundrae* gen. nov., sp. nov. and *Acidisoma sibiricum* sp. nov., two acidophilic, psychrotolerant members of the *Alphaproteobacteria* from acidic northern wetlands.** *International Journal of Systematic and Evolutionary Microbiology* 59:2283-2290

- 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*.** *Applied and Environmental Microbiology* 70:2596-2602

- Birtles RJ, Rowbotham TJ, Michel R, Pitcher DG, Lascola B, Alexiou-Daniel S, Raoult D (2000) **"*Candidatus* Odyssella thessalonicensis" gen. nov., sp. nov., an obligate intracellular parasite of *Acanthamoeba* species.** *International Journal of Systematic and Evolutionary Microbiology* 50:63-72

- Boettcher KJ, Barber BJ, Singer JT (1999) **Use of antibacterial agents to elucidate the etiology of juvenile oyster disease (JOD) in *Crassostrea virginica* and numerical dominance of an alpha-proteobacterium in JOD-affected animals.** *Applied and Environmental Microbiology* 65:2534-2539

- Boldareva EN, Tourova TP, Kolganova TV, Moskalenko AA, Mahneva ZK, Gorlenko VM (2009) ***Roseococcus suduntuyensis* sp. nov., a new aerobic**

bacteriochlorophyll a-containing bacterium isolated from a low-mineralized soda lake of Eastern Siberia. *Mikrobiologiya* 78:92-101

- Boscaro V, Fokin SI, Schweikert M, Petroni G (2011) **Preliminary characterization of "*Candidatus* Gortzia infectiva" (*Holosporaceae*, *Rickettsiales*, *Alphaproteobacteria*), a new *Holospora*-like bacterium from *Paramecium jenningsi*).** *Proceedings of the GDRE (Paramecium Genome Dynamics and Evolution) Pleniary Meeting, Roscoff (France)*

- Boscaro V, Petroni G, Verni F, Fokin SI, Vannini C (2010) **Molecular and morphological characterization of hosts and bacterial endosymbionts in a population of *Euplotes aediculatus* Pierson, 1943 reveals an unexpectedly diverse community of cytoplasmic prokaryotes.** *Proceedings of the XXVIII National Congress of the Italian Society of Protistology, Pisa (Italy)*

- Catania F, Wurmser F, Potekhin AA, Przybos E, Lynch M (2009) **Genetic diversity in the *Paramecium aurelia* species complex.** *Molecular Biology and Evolution* 26:421-431

- Chatton E, Brachon S (1933) **Sur un paramecia a deux races: *Paramecium duboscqui* n. sp.** *C.R. Société Biologique* 114:988-991

- Choi DH, Hwang CY, Cho BC (2009) ***Pelagibius litoralis* gen. nov., sp. nov., a marine bacterium in the family *Rhodospirillaceae* isolated from coastal seawater.** *International Journal of Systematic and Evolutionary Microbiology* 59:818-823

- Coenye T, Goris J, Spilker T, Vandamme P, LiPuma JJ (2002) **Characterization of unusual bacteria isolated from respiratory secretions of cystic fibrosis patients and description of *Inquilinus limosus* gen. nov., sp. nov.** *Journal of Clinical Microbiology* 40:2062-2069

- De Bary A (1879) **Die Erscheinungen der Symbiose.** Trübner, Strassbourg

- de Graaf RM, van Alen TA, Dutilh BE, Kuiper JWP, van Zoggel HJAA, Huhynh MB, Görtz H-D, Huynen MA, Hackstein JHP (2009**) The mitochondrial genomes of the ciliates *Euplotes minuta* and *Euplotes crassus*.** *BMC Genomics* 10

- Ding LX, Yokota A (2002) **Phylogenetic analysis of the genus *Aquaspirillum* based on 16S rRNA gene sequences.** FEMS Microbiology Letters 212-165-169

- Dong C, Lai Q, Chen L, Sun F, Shao Z, Yu Z (2010) ***Oceanibaculum pacificum* sp. nov., isolated from hydrothermal field sediment of the south-west Pacific Ocean.** *International Journal of Systematic and Evolutionary Microbiology* 60:219-222

- Edwards AWF, Cavalli-Sforza LL (1963) **The reconstruction of evolution.** *Annals of Human Genetics* 27:105-106

- Edwards AWF, Cavalli-Sforza LL (1964) **Reconstruction of evolutionary trees.** In: Heywood VH, McNeill J (Eds.) *Phenetic and Phylogenetic Classification.* Systematic Association Publ., London

- Eschbach E, Pfannkuchen M, Schweikert M, Drutschmann D, Brummer F, Fokin SI, Ludwig W, Görtz H-D (2006) **"*Candidatus* Paraholospora nucleivisitans", an intracellular bacterium in *Paramecium sexaurelia* shuttles between the cytoplasm and the nucleus of its host.** *Systematic and Applied Microbiology* 32:490-500

- Ewald PW (1987) **Transmission modes and evolution of the parasitism-mutualism continuum.** *Annals of the New York Academy of Sciences* 295-306

- Fels D, Kaltz O (2006) **Temperature-dependent transmission and latency of *Holospora undulata*, a micronucleus-specific parasite of the ciliate *Paramecium caudatum*.** *Proceedings of the Royal Society B.* 273:1031-1038

- Felsenstein J (1981) **Evolutionary trees from DNA sequences: a maximum likelihood approach**. *Journal of Molecular Evolution* 17:368-376

- Felsenstein J (2004) **Inferring Phylogenies.** Sinauer Associates, Sunderland, Massachussetts

- Ferrantini F, Fokin SI, Modeo L, Andreoli I, Dini F, Görtz H-D, Verni F, Petroni G (2009) **"*Candidatus* Cryptoprodotis polytropus", a novel *Rickettsia*-like organism in the ciliated protist *Pseudomicrothorax dubius* (Ciliophora, Nassophorea).** *Journal of Eukaryotic Microbiology* 56:119-129

- Finlay BJ, Fenchel T (1999) **Divergent perspectives on protist species richness.** *Protist* 150:229-233

- Foissner W (1999) **Protist diversity: Estimates of the near-imponderable.** *Protist* 150:363-368

- Fokin SI (1989) **Bacterial endobionts of the ciliate *Paramecium woodruffi*. I. Endobionts of the macronucleus.** *Cytology* 31:839-844

- Fokin SI (1991) ***Holospora recta* sp. nov. – a micronucleus-specific endobiont of the ciliate *Paramecium caudatum*.** *Cytology* 33:135-141

- Fokin SI (2011) **Frequency and biodiversity of symbionts in representatives of the main classes of Ciliophora.** *Proceedings of the VI European Congress of Protistology, Berlin (Germany)*

- Fokin SI, Sabaneyeva EV (1993) **Bacterial endocytobionts of the ciliate *Paramecium calkinsi*.** *European Journal of Protistology* 29:390-395

- Fokin SI, Boscaro V, Schweikert M, Petroni G (2011) **New *Holospora*-like bacterium from the *Paramecium* genus.** *Proceedings of the VI European Congress of Protistology, Berlin (Germany)*

- Fokin SI, Görtz H-D (2009) **Diversity of *Holospora* bacteria in *Paramecium* and their characterization.** In: Fujishima M (Ed.) *Endosymbionts in* Paramecium, Microbiology Monograph, Münster

- Fokin SI, Przybos E, Chivilev SM, Beier CL, Horn M, Skotarczak B, Wodecka B, Fujishima M (2004) **Morphological and molecular investigations of *Paramecium schewiakoffi* sp. nov. (Ciliophora, Oligohymenophorea) and current status of distribution and taxonomy of *Paramecium* spp.** *European Journal of Protistology* 40:225-243

- Fokin SI, Stoeck T, Schmidt H (1999) ***Paramecium duboscqui* Chatton, Brachon, 1933. Distribution, ecology and taxonomy.** *European Journal of Protistology* 35:161-167

- 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*.** *Applied and Environmental Microbiology* 65:206-212

- Fritz I, Strompl C, Abraham WR (2002) **Phylogenetic relationships of the genera *Stella*, *Labrys* and *Angulomicrobium* within the *'Alphaproteobacteria'* and description of *Angulomicrobium amanitiforme* sp. nov.** *International Journal of Systematic and Evolutionary Microbiology* 54:651-657

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

- Fujishima M, Kawai M, Yamamoto R (2005) ***Paramecium caudatum* acquires heat-shock resistence in ciliary movement by infection with the endonuclear symbiotic bacterium *Holospora obtusa*.** *FEMS Microbiology Letters* 243:101-105

- Gao S, Strüder-Kypke MC, Al-Rasheid KAS, Lin X, Song W (2010) **Molecular phylogeny of three ambiguous ciliate genera: *Kentrophoros*, *Trachelolophos* and *Trachelotractus* (Alveolata, Ciliophora).** *Zool. Scr.* 39:305-313

- Gentekaki E, Lynn DH (2010) **Evidence for cryptic speciation in *Charchesium polypinum* Linnaeus, 1758 (Ciliophora: Peritrichia) inferred from mitochondrial, nuclear, and morphological markers.** *Journal of Eukaryotic Microbiology* 57:508-519

- Glaeser J, Overmann J (1999) **Selective enrichment and characterization of *Roseospirillum parvum*, gen. nov. and sp. nov., a new purple nonsulfur bacterium with unusual light absorption properties.** *Archives of Microbiology* 171:405-416

- Gong YC, Yu YH, Zhu FY, Feng WS (2006) **Molecular phylogeny of *Stentor* (Ciliophora: Heterotrichea) based on small subunit ribosomal RNA sequences.** *Journal of Eukaryotic Microbiology* 54:45-48

- Görtz H-D (2006) **Symbiotic associations between ciliates and prokaryotes.** In Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds.) *The Prokaryotes*, 3rd edn, vol. 1, Springer, East Lansing

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

- Görtz H-D, Schmidt HJ (2005) **Holosporaceae *fam. nov.*** In: *Bergey's Manual of Systematic Bacteriology (2nd Edition)* Springer Press

- Greenberg DE, Porcella SF, Zelazny AM, Virtaneva K, Sturdevant DE, Kupko JJ, Barbian KD, Babar A, Dorward DW, Holland SM (2007) **Genome sequence analysis of the emerging human pathogenic acetic acid bacterium *Granulibacter bethesdensis*.** *Journal of Bacteriology* 189:8727-8736

- Gromov BV, Ossipov DV (1981) ***Holospora* (*ex* Hafkine 1890) nom, rev., a genus of bacteria inhabiting the nuclei of paramecia.** *International Journal of Systematic Bacteriology* 31:348-352

- Gundlapally SR, Garcia-Pichel F (2006) **The community and phylogenetic diversity of biological soil crusts in the Colorado Plateau studied by molecular fingerprinting and intensive cultivation.** *Microbial Ecology* 52:345-357

- Guo WB, Song WB, Al-Rasheid KAS, Shao C, Miao M, Al-Farraj SA, Al-Qurishy SA, Chen ZG, Yi ZZ, Gao S (2008) **Phylogenetic position of three *Condylostoma* species (Protozoa, Ciliophora, Heterotrichea) inferred from the small subunit rRNA gene sequence.** *Progress in Natural Science* 18:1089-1093

- Guyoneaud R, Moune S, Eatock C, Bothorel V, Hischler-Rea A, Willison J, Duran R, Liesack W, Herbert R, Matheron R, Caumette P (2002) **Characterization of three spiral-shaped purple nonsulfur bacteria isolated from coastal lagoon sediments, saline sulfur springs, and microbial mats: emended description of the genus *Roseospira* and description of *Roseospira marina* sp. nov., *Roseospira navarrensis* sp. nov., and *Roseospira thiosulfatophila* sp. nov.** *Archiv of Microbiology* 178:315-324

- Hafkine MW (1890) **Maladies infectieuses des paramécies.** *Annals of the Institute Pasteur* 4:363-379

- Hammerschmidt B, Schlegel M, Lynn DH, Leipe DD, Sogin ML, Raikov IB (1996) **Insights into the evolution of nuclear dualism in the ciliates revealed by phylogenetic analysis of rRNA sequences.** *Journal of Eukaryotic Microbiology* 43:225-230

- Hausmann K & Bradbury PC (eds.) (1996) *Ciliates. Cells as organisms.* Gustav Fischer, Stuttgart, New York

- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) **Biological identifications through DNA barcode.** *Proceedings of the Royal Society of London B* 270:313-321

- Heckmann K Schmidt HJ (1987) ***Polynucleobacter necessarius* gen. nov., sp. nov., an obligately endosymbiotic bacterium living in the cytoplasm of *Euplotes aediculatus*.** *International Journal of Systematic Bacteriology* 37:456-457

- Hentschel U, Steinert M, Hacker Jörg (2000) **Common molecular mechanisms of symbiosis and pathogenesis.** *Trends in Microbiology* 8:226-231

- Hiraishi A, Matsuzawa Y, Kanbe T, Wakao N (2000) ***Acidisphaera rubrifaciens* gen. nov., sp. nov., an aerobic bacteriochlorophyll-containing bacterium isolated from acidic environments.** *International Journal of Systematic and Evolutionary Microbiology* 50:1539-1546

- 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.** *Molecular and Phylogenetic Evolution* 4:77-87

- Hori M, Fuji K, Fujishima M (2008) **Micronucleus-specific bacterium *Holospora elegans* irreversibly enhances stress gene expression of the host *Paramecium caudatum*.** *Journal of Eukaryotic Microbiology* 55:515-521.

- 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*.** *Environmental Microbiology* 1:357-367

- Huelsenbeck JP, Ronquist F (2001) **MrBayes: Bayesian inference of phylogeny.** *Bioinformatics* 17:754-755

- International Society of Symbiosis (2003), 4th Congress, Halifax

- Johnson DB, Okibe N, Roberto FF (2004) **Novel thermo-acidophilic bacteria isolated from geothermal sites in Yellowstone National Park: physiological and phylogenetic characteristics.** *Archives of Microbiology* 180:60-68

- Jojima Y, Mihara Y, Suzuki S, Yokozeki K, Yamanaka S, Fudou R (2004) ***Saccharibacter floricola* gen. nov., sp. nov., a novel osmophilic acetic acid bacterium isolated from pollen.** *International Journal of Systematic and Evolutionary Microbiology* 54:2263-2267

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

- Jukes TH, Cantor CR (1969) **Evolution of protein molecules.** In: Munro MN (Ed.), *Mammalian Protein Metabolism* Academic Press, New York

- Kalyan Chakravarthy S, Srinivas TN, Anil Kumar P, Sasikala Ch, Ramana ChV (2007) ***Roseospira visakhapatnamensis* sp. nov. and *Roseospira goensis* sp. nov**. *International Journal of Systematic and Evolutionary Microbiology* 57:2453-2457

- Kampfer P, Busse HJ, Rossello-Mora R, Kjellin E, Falsen E (2004) ***Rhodovarius lipocyclicus* gen. nov. sp. nov., a new genus of the alpha-1 subclass of the Proteobacteria.** *Systematic and Applied Microbiology* 27:511-516

- Kawasaki H, Hoshino Y, Kuraishi H, Yamasato K (1993a) ***Rhodocista centenaria*** **gen. nov., sp. nov., a cyst-forming anoxygenic photosynthetic bacterium and its phylogenetic position in the proteobacteria alpha group.** *Journal of General and Applied Microbiology* 38:541-551

- Kawasaki H, Hoshino Y, Yamamoto K (1993b) **Phylogenetic diversity of phototrophic purple non-sulfur bacteria in the Proteobacteria alpha group.** FEMS *Microbiology Letters* 112:61-66

- Kikuchi Y, Fukatsu T (2005) ***Rickettsia*** **infection in natural leech populations.** *Microbial Ecology* 49:265-271

- Kim YG, Hwang CY, Cho BC (2008) ***Pelagicola litoralis*** **gen. nov., sp. nov., isolated from coastal water in Korea.** *International Journal of Systematic and Evolutionary Microbiology* 58:2102-2106

- Kim MK, Schubert K, Im WT, Kim KH, Lee ST, Overmann J (2007) ***Sphingomonas kaistensis*** **sp. nov., a novel alphaproteobacterium containing pufLM genes.** *International Journal of Systematic and Evolutionary Microbiology* 57:1527-1534

- Kishimoto N, Kosako Y, Wakao N, Tano T, Hiraishi A (1995) **Transfer of *Acidiphilium facilis* and *Acidiphilium aminolytica* to the genus *Acidocella* gen. nov., and emendation of the genus *Acidiphilium*.** *Systematic and Applied Microbiology* 18:85-91

- Kodama Y, Fujishima M (2009) **Infection of *Paramecium bursaria* by symbiotic *Chlorella* species.** In: Fujishima M (Ed.) *Endosymbionts in* Paramecium, Microbiology Monograph, Münster

- Lai Q, Yuan J, Gu L, Shao Z (2009a) ***Marispirillum indicum*** **gen. nov., sp. nov., isolated from a deep-sea environment.** *International Journal of Systematic and Evolutionary Microbiology* 59:1278-1281

- Lai Q, Yuan J, Wu C, Shao Z (2009b) ***Oceanibaculum indicum*** **gen. nov., sp. nov., isolated from deep seawater of the Indian Ocean.** *International Journal of Systematic and Evolutionary Microbiology* 59:1733-1737

- Lanave C, Preparata G, Saccone C, Serio G (1984) **A new method for calculating evolutionary substitution rates.** *Journal of Molecular Evolution* 20:86-93

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

- Leeuwenhoeck A (1674) **More observations from Mr. Leewenhook, in a letter of 7 September 1674, sent to the publisher.** *Philosophical Transactions of the Royal Society* 9:178-182

- Lemey P, Salemi M, Vandamme A-M (2009) *The Phylogenetic Handbook: a practical approach to phylogenetic analysis and hypothesis testing.* Cambridge University Press

- Lesaulnier C, Papamichail D, McCorkle S, Ollivier B, Skiena S, Taghavi S, Zak D, van der Lelie D (2008) **Elevated atmospheric CO2 affects soil microbial diversity associated with trembling aspen.** *Environmental Microbiology* 10:926-941

- Liang Q, Lloyd-Jones G (2010) ***Sphingobium scionense* sp. nov., an aromatic hydrocarbon-degrading bacterium isolated from contaminated sawmill soil.** *International Journal of Systematic and Evolutionary Microbiology* 60:413-416

- Lisdiyanti P, Kawasaki H, Seki T, Yamada Y, Uchimura T, Komagata K (2000) **Systematic study of the genus *Acetobacter* with descriptions of *Acetobacter indonesiensis* sp. nov., *Acetobacter tropicalis* sp. nov., *Acetobacter orleanensis* (Henneberg 1906) comb. nov., *Acetobacter lovaniensis* (Frateur 1950) comb. nov., and *Acetobacter estunensis* (Carr 1958) comb. nov.** *Journal of General and Applied Microbiology* 46:147-165

- Lisdiyanti P, Kawasaki H, Widyastuti Y, Saono S, Seki T, Yamada Y, Uchimura T, Komagata K (2002) ***Kozakia baliensis* gen. nov., sp. nov., a novel acetic acid bacterium in the alpha-proteobacteria.** *International Journal of Systematic and Evolutionary Microbiology* 52:813-818

- Liu C, Wu Y, Li L, Ma Y, Shao Z (2007) ***Thalassospira xiamenensis* sp. nov. and *Thalassospira profundimaris* sp. nov.** *International Journal of Systematic and Evolutionary Microbiology* 57:316-320

- Loganathan P, Nair S (2004) ***Swaminathania salitolerans* gen. nov., sp. nov., a salt-tolerant, nitrogen-fixing and phosphate-solubilizing bacterium from wild rice (*Porteresia coarctata* Tateoka).** *International Journal of Systematic and Evolutionary Microbiology* 54:1185-1190

- Loy JK, Dewhirst FE, Weber W, Frelier PF, Garbar TL, Tasca SI, Templeton JW (1996) **Molecular phylogeny and *in situ* detection of the etiologic agent of necrotizing hepatopancreatitis in shrimp.** *Applied and Environmental Microbiology* 62:3439-3445

- Lynn DH (2008) *The ciliated protozoa* (3rd edition). Springer

- Lynn DH, Strüder-Kypke M (2006) **Species of *Tetrahymena* identical by small subunit rRNA gene sequence are discriminated by mitochondrial cytochrome *c* oxidase I gene sequences.** *Journal of Eukaryotic Microbiology* 53:385-387

- Ludwig W et al. (2004) **ARB: a software environment for sequence data.** *Nucleic Acids Research* 32:1363-1371

- Maciejewska A (2006) **Relationships of new sibling species of *Paramecium jenningsi* based on sequences of the histone H4 gene fragment.** *European Journal of Protistology* 43:125-130

- Manz W, Amann R, Ludwig W, Wagner M, Schleifer KH (1992) **Phylogenetic oligodeoxynucleotide probes for the major subclasses of Proteobacteria: problems and solutions.** *Systematic and Applied Microbiology* 15:593-600

- Margulis, L (1981) *Symbiosis in cell evolution.* Freemannn, San Francisco

- Maszenan AM, Seviour RJ, Patel BK, Janssen PH, Wanner J (2005) ***Defluvicoccus vanus* gen. nov., sp. nov., a novel Gram-negative coccus/coccobacillus in the *Alphaproteobacteria* from activated sludge.** *International Journal of Systematic and Evolutionary Microbiology* 55:2105-2111

- Matsuo J, Kawaguchi K, Nakamura S, Hayashi Y, Yoshida M, Takahashi K, Mizutani Y, Yao T, Yamaguchi H (2010) **Survival and transfer ability of phylogenetically diverse bacterial endosymbionts in environmental *Acanthamoeba* isolates.** *Environmental Microbiology* 2:524-533

- Mazei Y, Gao S, Warren A, Li L, Li J, Song W, Esaulov A (2009) **A reinvestigation of the marine ciliate *Trachelocerca ditis* (Wright, 1982) Foissner and Dragesco, 1996 (Ciliophora, Karyorelictea) from the Yellow Sea and an Assessment of Its Phylogenetic Position Inferred from the Small Subunit rRNA Gene Sequence**. *Acta Protozool.* 48, 213-221

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

- Miao W, Simpson AG, Fu C, Lobban CS (2005) **The giant zooxanthellae-bearing ciliate *Maristentor dinoferus* (Heterotrichea) is closely related to folliculinidae.** *Journal of Eukaryotic Microbiology* 52:11-16

- Miao M, Song W, Clamp JC, Al-Rasheid KA, Al-Khedhairy AA, Al-Arifi 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.** *Journal Eukaryotic Microbiology* 56:244-250

- Modeo L, Dias RJP, Fernandes N, Tomei G, Boscaro V, Verni F, da Silva-Neto ID, Petroni G (2011) **Phylogenetic systematics of the genus *Condylostoma* (Ciliophora, Heterotrichea) by means of a multidisciplinary analytical study of some marine and brackish morphospecies.** *Proceedings of the VI European Congress of Protistology, Berlin (Germany)*

- 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).** *Proceedings of the Japan Academy B* 82:359-374

- Müller OF (1786) **Animalcula Infusoria fluviatilia et marina.** Havniae et Lipsiae, Copenhagen

- Munk AC et al., (2011) **Complete genome sequence of *Rhodospirillum rubrum* type strain (S1).** *Standards in Genomic Sciences* 4:293-302

- Murray RGE, Stackebrandt E (1995) **Implementation of the provisional status *Candidatus* for incompletely described prokaryotes.** *International Journal of Systematic Bacteriology* 45:186-187

- Ohashi N, Fukuhara M, Shimada M, Tamura A (1995) **Phylogenetic position of *Rickettsia tsutsugamushi* and the relationship among its antigenic variants by analyses of 16S rRNA gene sequences.** *FEMS Microbiology Letters* 125:299-304

- Paddock CD, Sumner JW, Shore GM, Bartley DC, Elie RC, McQuade JG, Martin CR, Goldsmith CS, Childs JE (1997) **Isolation and characterization of *Ehrlichia chaffeensis* strains from patients with fatal ehrlichiosis.** *Journal Clinical Microbiology* 35:2496-2502

- Petroni G, Dini F, Verni F, Rosati G (2002) **A molecular approach to the tangled intrageneric relationship underlying phylogeny in *Euplotes* (Ciliophora, Spirotrichea).** *Journal of Eukaryotic Microbiology* 51:49-59

- Pfennig N, Lunsdorf H, Suling J, Imhoff JF (1997) ***Rhodospira trueperi* gen. nov., spec. nov., a new phototrophic Proteobacterium of the alpha group.** *Archives of Microbiology* 168:39-45

- Posada D, Crandall KA (1998) **ModelTest: testing the model of DNA substitutions.** *Bioinformatics* 14:817-818

- Preer JR, Preer LB (1982) **Revival of names of protozoan endosymbionts and proposal of *Holospora caryophila* nom. nov.** *International Journal of Systematic Bacteriology* 32:140-141

- Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J, Glöckner FO (2007) **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.** *Nucleic Acids Research* 35:7188-7196

- Przybos E, Tarcz S, Fokin SI (2009) **Molecular Polymorphism of *Paramecium tetraurelia* (Ciliophora, Protozoa) in Strains Originating from Different Continents.** *Folia biologica (Krakow)* 57:57-63

- Przybos E, Tarcz S, Surmacz M (2008) **A molecular survey of *Paramecium dodecaurelia* (Ciliophora, Protozoa) strains from a single pond.** *Acta Protozoologica* 47:161-171

- Raultian M, Wackerov-Kousova N, Lebedeva N, Skoblo I (2011) **Intranuclear symbionts of *Paramecium*: genus *Holospora* revisited.** *Proceedings of the VI European Congress of Protistology, Berlin (Germany)*

- Rosati G (1999) **Epixenosomes: Symbionts of the hypotrich ciliate *Euplotidium itoi*.** *Symbiosis* 26:1-23

- Roux V, Raoult D (1995) **Phylogenetic analysis of the genus *Rickettsia* by 16S rDNA sequencing.** *Researches in Microbiology* 146:385-396

- Rypien KL, Ward JR, Azam F (2010) **Antagonistic interactions among coral-associated bacteria.** *Environmental Microbiology* 12:28-39

- Saitoh S, Suzuki T, Nishimura Y (1998) **Proposal of *Craurococcus roseus* gen. nov., sp. nov. and *Paracraurococcus ruber* gen. nov., sp. nov., novel aerobic bacteriochlorophyll a-containing bacteria from soil.** *International Journal of Systematic Bacteriology* 48 PT 3, 1043-1047

- Saitou N, Nei M (1987) **The neighbor-joining method: A new method for reconstructing phylogenetic trees.** *Molecular Biology and Evolution* 4:406-425

- Sambrook J, Russell D (2000) *Molecular cloning: a laboratory manual* (3rd Edition). Cold Spring Harbor Laboratory Press, USA

- Sapp J (2004) **The dynamics of symbiosis: an historical overview.** *Canadian Journal of Botany* 82:1046-1056

- Satomi M, Kimura B, Hayashi M, Shouzen Y, Okuzumi M, Fujii T (1998) ***Marinospirillum* gen. nov., with descriptions of *Marinospirillum megaterium* sp. nov., isolated from kusaya gravy, and transfer of *Oceanospirillum minutulum* to *Marinospirillum minutulum* comb. nov.** *International Journal of Systematic Bacteriology* 48:1341-1348

- Schleifer KH, Schueler D, Spring S, Weizenegger M, Ahmann R, Ludwig W, Koehler M (1991) **The genus *Magnetospirillum* gen. nov. Description of *Magnetospirillum gryphiswaldense* sp. nov. and transfer of *Aquaspirillum magnetotacticum* to *Magnetospirillum magnetotacticum* comb. nov.** *Systematic and Applied Microbiology* 14:379-385

- Schmidt SL, Foissner W, Schlegel M, Bernhard D (2007a) **Molecular phylogeny of the Heterotrichea (Ciliophora, Postciliodesmatophora) based on small subunit rRNA gene sequences.** *Journal of Eukaryotic Microbiology* 54:358-363

- 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

- 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

- Schmitz-Esser S, Toenshoff ER, Haider S, Heinz E, Hoenninger VM, Wagner M, Horn M (2008) **Diversity of bacterial endosymbionts of environmental acanthamoeba isolates.** *Applied and Environmental Microbiology* 74:5822-5831

- 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.** *Journal of Eukaryotic Microbiology* 53:499-506

- Schrallhammer M, Schweikert M (2009) **The killer effect of *Paramecium* and its causative agents.** In: Fujishima M (Ed.) *Endosymbionts in* Paramecium, Microbiology Monograph, Münster

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

- Schrallhammer M, Vitali V, Berendonk TU, Potehkin A, Gortz H-D, Petroni G (2011b) **First molecular description of the *Paramecium* parasite *Holospora caryophila*.** *Proceedings of the VI European Congress of Protistology, Berlin (Germany)*

- Shi BH, Arunpairojana V, Palakawong S, Yokota A (2002) ***Tistrella mobilis* gen. nov, sp. nov, a novel polyhydroxyalkanoate-producing bacterium belonging to alpha-Proteobacteria.** *Journal of General and Applied Microbiology* 48:335-343

- Shi X, Jin M, Liu G (1997) **Rediscovery of *Paramecium duboscqui* Chatton & Brachon, 1933 and a description of its characteristics.** *Journal of Eukaryotic Microbiology* 44:134-141

- Shimizu S, Akiyama M, Naganuma T, Fujioka M, Nako M, Ishijima Y (2007) **Molecular characterization of microbial communities in deep coal seam groundwater of northern Japan.** *Geobiology* 5:423-433

- Sievers M, Gaberthuel C, Boesch C, Ludwig W, Teuber M (1995) **Phylogenetic position of *Gluconobacter* species as a coherent cluster separated from all *Acetobacter* species on the basis of 16S ribosomal RNA sequences.** *FEMS Microbiology Letters* 126:123-126

- Sievers M, Ludwig W, Teuber M (1994) **Phylogenetic positioning of *Acetobacter*, *Gluconobacter*, *Rhodopila* and *Acidiphilium* species as a branch of acidophilic bacteria in the alpha-subclass of *Proteobacteria* based on 16S ribosomal DNA sequences.** *Systematic and Applied Microbiology* 17:189-196

- Sizova MV, Panikov NS, Tourova TP, Flanagan PW (2007) **Isolation and characterization of oligotrophic acido-tolerant methanogenic consortia from a *Sphagnum* peat bog.** *FEMS Microbiology and Ecology* 45:301-315

- Skovorodkin IN, Fokin SI, Fujishima M (2001) **Fates of the endonuclear symbiotic bacteria *Holospora obtusa* and *Holospora undulata* injected into the macronucleus of *Paramecium caudatum*.** *European Journal of Protistology* 237:137-145

- Song W, Warren A, Ji D, Wang M, Al-Rasheid KA (2003) **New contributions to two heterotrichous ciliates, *Folliculina simplex* (Dons, 1917), *Condylostoma curva* Burkovsky, 1070 and one licnophorid, *Licnophora lyngbycola* Fauré-Fremiet, 1937 (Protozoa, Ciliophora): descriptions of morphology and infraciliature.** *Journal of Eukaryotic. Microbiol*ogy 50:449-462.

- Sonneborn TM (1975) **The *Paramecium aurelia* complex of fourteen sibling species.** *Transactions of the American Microbiology Society* 94:155-178

- Spiegel A (1926) **Einige neue marine Ciliaten.** *Archiv fur Protistenkude* 55:184-190.

- Stackebrandt E, Ebers E (2006) **Taxonomic parameters revisited: tarnished gold standards.** *Microbiology Today*

- Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kampfer P, Maiden MCJ, Nesme X, Rossello-Mora R, Swings J, Truper HG, Vauterin L, Ward AC, Whitman WB (2002) **Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology.** *International Journal of Systematic and Evolutionary Microbiology* 52:1043-1047

- Stoeck T, Schmidt HJ (1998) **Fast and accurate identification of European species of the *Paramecium aurelia* complex by RAPD-fingerprints.** *Microbial Ecology* 35:311-317

- Stothard DR, Fuerst PA (1995) **Evolutionay analysis of the spotted fever and typhus groups of *Rickettsia* using 16S rRNA gene sequences.** *Systematic and Applied Microbiology* 18:52-61

- Strüder-Kypke M, Lynn DH (2003) **Sequence analyses of the small subunit rRNA gene confirm the paraphyly of oligotrich ciliates *sensu lato* and support the monophyly of the subclasses Oligotrichia and Choreotrichia (Ciliophora, Spirotrichea).** *Journal of Zoology* 260:87-97

- Strüder-Kypke M, 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.** *Systematics and Biodiversity* 8:131-148

- Strüder-Kypke M, Wright AG, Fokin SI, Lynn DH (2000) **Phylogenetic relationships of the genus Paramecium inferred from small subunit rRNA gene sequences.** *Molecular Phylogenetics and Evolution* 14:122-130

- Sunagawa S, DeSantis TZ, Piceno YM, Brodie EL, DeSalvo MK, Voolstra CR, Weil E, Andersen GL Medina M (2009) **Bacterial diversity and White Plague Disease-associated community changes in the Caribbean coral *Montastraea faveolata*.** *ISME Journal* 3:512-521

- Takahashi M, Yukphan P, Yamada Y, Suzuki K, Sakane T, Nakagawa Y (2006) **Intrageneric structure of the genus *Gluconobacter* analyzed by the 16S rRNA gene and 16S-23S rRNA gene internal transcribed spacer sequences.** *Journal of General and Applied Microbiology* 52:187-193

- Thamm M, Schmidt SL, Bernhard D (2010) **Insights into the phylogeny of the genus *Stentor* (Heterotrichea, Ciliophora) with special emphasis on the evolution of the macronucleus based on SSU rDNA data**. *Acta Protozoologica* 49:149-157

- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) **The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Research* 25:4876-4882

- Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P (2010) **Notes on the characterization of prokaryotes strains for taxonomic purposes.** *International Journal of Systematic and Evolutionary Microbiology* 60:249-266

- Tourancheau AB, Tsao N, Klobutcher LA, Pearlman RE, Adoutte A (1995) **Genetic-code deviations in the ciliates – evidence for multiple and independent events.** *EMBO Journal* 14:3262-3267

- Trcek J (2002) **Genotypic characteristics of *Gluconacetobacter hansenii* LMG 1582 suggest its reclassification to a new species.** *Zb. Bioteh. Fak. Univ. Ljubl. Kmet. Supl.* 79:19-26

- Urios L, Michotey V, Intertaglia L, Lesongeur F, Lebaron P (2008) ***Nisaea denitrificans* gen. nov., sp. nov. and *Nisaea nitritireducens* sp. nov., two novel members of the class *Alphaproteobacteria* from the Mediterranean Sea.** *International Journal of Systematic and Evolutionary Microbiology* 58 (PT 10), 2336-2341

- Urios L, Michotey V, Intertaglia L, Lesongeur F, Lebaron P (2010) ***Thalassobaculum salexigens* sp. nov., a new member of the family *Rhodospirillaceae* from the NW Mediterranean Sea, and emended description of the genus *Thalassobaculum*.** *International Journal of Systematic and Evolutionary Microbiology* 60:209-213

- van Berkum P, Beyene D, Eardly BD (1996) **Phylogenetic relationships among *Rhizobium* species nodulating the common bean (*Phaseolus vulgaris* L.).** *International Journal of Systematic Bacteriology* 46:240-244

- Vandekerckhove TT, Watteyne S, Willems A, Swings JG, Mertens J, Gillis M (1999) **Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium *Wolbachia* from the novel host *Folsomia candida* (Hexapoda, Collembola) and its implications for wolbachial taxonomy.** *FEMS Microbiology Letters* 180:279-286

- 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).** *Applied and Environmental Microbiology* 76:4047-4054

- Vannini C, Poeckl M, Petroni G, Wu QLL, Lang E, Stackebrandt E, Schrallhammer M, Richardson PM, Hahn MW (2007) **Endosymbiosis *in statu nascendi*: close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*).** *Environmental Microbiology* 9:347-359

- 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. *European Journal of Protistology* 39:481-485

- 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 Ecology* 49:434-442

- 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".** *International Journal of Systematic and Evolutionary Microbiology* 54:1151-1156

- Wang YX, Liu JH, Zhang XX, Chen YG, Wang ZG, Chen Y, Li QY, Peng Q, Cui XL (2009) ***Fodinicurvata sediminis* gen. nov., sp. nov. and *Fodinicurvata fenggangensis* sp. nov., poly-{beta}-hydroxybutyrate-producing bacteria in the family *Rhodospirillaceae*.** *International Journal of Systematic and Evolutionary Microbiology* 59:2575-2581

- Weon HY, Kim BY, Hong SB, Joa JH, Nam SS, Lee KH, Kwon SW (2007) ***Skermanella aerolata* sp. nov., isolated from air, and emended description of the genus *Skermanella*.** *International Journal of Systematic and Evolutionary Microbiology* 57:1539-1542

- Weisburg WG, Dobson ME, Samuel JE, Dasch GA, Mallavia L.P, Baca O, Mandelco L, Sechrest JE, Weiss E, Woese CR (1989) **Phylogenetic diversity of the Rickettsiae.** *Journal of Bacteriology* 171:4202– 4206.

- Wilson K, John W (2005) *Principles and Techniques of Biochemistry and Molecular Biology* (6th Edition). Cambridge University Press

- Woese CR, Kandler O, Wheelis ML (1990) **Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya.** *Proceedings of the National Acadamy of Sciences USA* 87:4576-4579

- Xia Y, Embley T O'Donnell A (1994) **Phylogenetic analysis of *Azospirillum* by direct sequencing of PCR amplified 16S rDNA.** *Systematic and Applied Microbiology* 17:197-201

- Xie CH, Yokota A (2005) ***Azospirillum oryzae* sp. nov., a nitrogen-fixing bacterium isolated from the roots of the rice plant *Oryza sativa*.** *International Journal of Systematic and Evolutionary Microbiology* 55:1435-1438

- Yamada Y, Katsura K, Kawasaki H, Widyastuti Y, Saono S, Seki T, UchimuraT, Komagata K (2000) ***Asaia bogorensis* gen. nov., sp. nov., an unusual acetic acid bacterium in the alpha-Proteobacteria.** *International Journal of Systematic and Evolutionary Microbiology* 50:823-829

- Yamashita S, Uchimura T, Komagata K (2004) **Emendation of the genus *Acidomonas* Urakami, Tamaoka, Suzuki and Komagata 1989.** *International Journal of Systematic and Evolutionary Microbiology* 54:865-870

- Yee B, Oertli GE, Fuerst JA, Stale JT (2010) **Reclassification of the polyphyletic genus *Prosthecomicrobium* to form two novel genera, *Vasilyevaea* gen. nov. and *Bauldia* gen. nov. with four new combinations: *Vasilyevaea enhydra* comb. nov., *Vasilyevaea mishustinii* comb. nov., *Bauldia consociata* comb. nov. and *Bauldia litoralis* comb. nov.** *International Journal of Systematic and Evolutionary Microbiology* 60:2960-2966

- Yoo SH, Weon HY, Noh HJ, Hong SB, Lee CM, Kim BY, Kwon SW, Go SJ (2008) ***Roseomonas aerilata* sp. nov., isolated from an air sample.** *International Journal of Systematic and Evolutionary Microbiology* 58:1482-1485

- Yoon JH, Kang SJ, Oh HW, Oh TK (2007a) ***Roseomonas terrae* sp. nov.** *International Journal of Systematic and Evolutionary Microbiology* 57:2485-2488

- Yoon JH, Kang SJ, Park S, Oh TK (2007b) ***Caenispirillum bisanense* gen. nov., sp. nov., isolated from sludge of a dye works.** *International Journal of Systematic and Evolutionary Microbiology* 57:1217-1221

- Young CC, Hupfer H, Siering C, Ho MJ, Arun AB, Lai WA, Rekha PD, Shen FT, Hung MH, Chen WM, Yassin AF (2008) ***Azospirillum rugosum* sp. nov., isolated from oil-contaminated soil.** *International Journal of Systematic and Evolutionary Microbiology* 58:959-963

- Yukphan P, Malimas T, Muramatsu Y, Takahashi M, Kaneyasu M, Potacharoen W, Tanasupawat S, Nakagawa Y, Hamana K, Tahara Y, Suzuki K, Tanticharoen M, Yamada Y (2009) ***Ameyamaea chiangmaiensis* gen. nov., sp. nov., an acetic acid bacterium in the alpha-Proteobacteria.** *Biosciences, Biotechnology, Biochemistry* 73:2156-2162

- Yukphan P, Malimas T, Potacharoen W, Tanasupawat S, Tanticharoen M, Yamada Y (2005) ***Neoasaia chiangmaiensis* gen. nov., sp. nov., a novel osmotolerant acetic acid bacterium in the alpha-Proteobacteria.** *Journal of General and Applied Microbiology* 51:301-311

- Yukphan P, Potacharoen W, Tanasupawat S, Tanticharoen M, Yamada Y (2004) ***Asaia krungthepensis* sp. nov., an acetic acid bacterium in the alpha-Proteobacteria.** *International Journal of Systematic and Evolutionary Microbiology* 54:313-316

- Yurkov V, Stackebrandt E, Holmes A, Fuerst JA, Hugenholtz P, Golecki J, Gad'on N, Gorlenko VM, Kompantseva EI, Drews G (1994) **Phylogenetic positions of novel aerobic, bacteriochlorophyll a-containing bacteria and description of *Roseococcus thiosulfatophilus* gen. nov., sp. nov., *Erythromicrobium ramosum* gen. nov., sp. nov., and *Erythrobacter litoralis* sp. nov.** *International Journal of Systematic Bacteriology* 44 (3), 427-434

- Zhang GI, Hwang CY, Cho BC (2008) ***Thalassobaculum litoreum* gen. nov., sp. nov., a member of the family *Rhodospirillaceae* isolated from coastal seawater.** *International Journal of Systematic and Evolutionary Microbiology* 58:479-485

- Zhang D, Yang H, Zhang W, Huang Z, Liu SJ (2003) ***Rhodocista pekingensis* sp. nov., a cyst-forming phototrophic bacterium from a municipal wastewater treatment plant.** *International Journal of Systematic and Evolutionary Microbiology* 53:1111-1114

- Zuckerkandl E, Pauling L (1965) **Evolutionary divergence and convergence in proteins.** In: Bryson V, Vogel HJ (eds.) *Evolving Genes and Proteins*. Academic Press, New York