UNIVERSITY OF PISA

Department of Biology

Degree in
BIOMOLECULAR SCIENCE AND TECHNOLOGY

Molecular description of *Holospora caryophila,* the highly infectious macronuclear endosymbiont of *Paramecium* spp.

Candidate:

Valerio Vitali

Supervisors:

Dr. Martina Schrallhammer

Dr. Giulio Petroni

This work is dedicated to Louise B. Preer and John R. Preer Jr., the American scientists that first described Holospora caryophila, formerly known as Alpha.

Content

1. Riassunto analitico

In questo studio, viene presentata la prima descrizione molecolare di *Holospora caryophila*, un simbionte macronucleare infettivo del *Paramecium*. In accordo con la descrizione originale, *H. caryophila* è affiliata al genere *Holospora* sulla base di morfologia, peculiarità del ciclo vitale e specificità d'ospite. Otto ceppi di *Paramecium* sono stati sottoposti ad analisi comparativa del marcatore molecolare 16S rRNA dei simbionti ospitati. Tutti i ceppi di *H. caryophila* investigati hanno mostrato valori di similarità del gene 16S rRNA pari od inferiori all' 87,8 % con i membri del genere *Holospora* finora descritti a livello molecolare, tra cui la "type species" *Holospora undulata*. Grazie al recente aumento di risoluzione nella filogenesi degli *Holospora-like* bacteria (HLBs), siamo adesso in grado di ricostruire un modello evolutivo del meccanismo di trasmissione orizzontale di questi parassiti. Viene infatti qui dimostrato, su base filogenetica, che la capacità delle forme infettive (IFs) di accumularsi nel "connecting piece" durante la divisione nucleare dell'ospite è un apomorfia delle olospore maggiormente specializzate nello stile di vita parassitario. Vista l'incapacità di *H. caryophila* di indurre il "connecting piece" durante il ciclo infettivo, e dati i modesti valori di similarità del 16S rRNA rispetto agl'altri HLBs, viene qui proposta la sua rimozione dal genere *Holospora*, il trasferimento ad un nuovo taxon *Preeria* gen. nov. e la combinazione *Preeria caryophila* comb. nov.. Infine, in un sottogruppo dei ceppi di *H. caryophila* investigati, è stato caratterizzato l' "Internal Transcribed Spacer" (ITS), e l'estremità 5' del gene 23S rRNA. L'analisi di questa regione ha rivelato una sorprendente "etereogeneità molecolare" tra i ceppi ospitati dalle due morfospecie di *Paramecium* naturalmente infettate da *H. caryophila*; il *Paramecium aurelia* complex ed il *Paramecium caudatum*. Anche se apparentemente identici, i ceppi di *H. caryophila* infettanti le due morfospecie ospite potrebbero essere geneticamente distinti.

2. Abstract

Protists belonging to phylum *Ciliophora* are frequently colonized by intracellular bacteria traditionally referred to as "endosymbionts". Among ciliates, *Paramecium* (*Eukaryota, Alveolata, Ciliophora, Intramacronucleata*) represents a suitable model for studying the biodiversity and phylogeny of these bacteria, as many as 60 prokaryotic organisms have been found to inhabit virtually all of its subcellular compartments. In this study is presented the first molecular description of *Holospora caryophila*, a highly infectious macronuclear symbiont of *Paramecium*. In agreement with the original description, *H. caryophila* is affiliated with the genus *Holospora* on the basis of morphology, life cycle peculiarities, and host specificity. Eight *Paramecium* strains (naturally infected with *H. caryophila*) were examined by comparative analysis of the 16S rRNA of the infecting symbionts and fluorescence *in situ* hybridization (FISH) experiments. Each investigated *H. caryophila* strain, including the type strain from *P. biaurelia* stock 562, showed 16S rRNA similarity values equal or lower than 87.8 % with the members of the genus *Holospora* accounted for in the analysis, including the type species *Holospora undulata*. As the phylogeny of the *Holospora*-like bacteria (HLBs) is increasing in resolution, the evolutionary pattern of the horizontal-transmission mechanism is revealed. Here, is shown on a molecular basis, that the ability of the infectious forms (IFs) to assemble in the connecting piece during the host nuclear division, is an apomorphy of the most specialized *Holosporas*, and has to be regarded as a highly advanced trait. According to the inability of *H. caryophila* to "induce" the connecting piece during the infective cycle, and the modest 16S rRNA similarity shared with others HLBs we propose its removal from the genus *Holospora*, its transfer to the new taxon *Preeria*, gen. nov., and the new combination *Preeria caryophila* comb. nov.. Furthermore, in a subgroup of the investigated *H. caryophila* strains, the Internal Transcribed Spacer (ITS) and the 5' end of the 23S rRNA gene have been characterized. The analysis of this region revealed an unexpected heterogeneous molecular composition between strains harbored by the two natural host morphospecies, namely the *Paramecium aurelia* complex and *Paramecium caudatum*.

Keywords: Endosymbionts, *Holospora*-like bacteria, *Holospora caryophila*, Infectious forms, Connecting piece, Macronucleus, Micronucleus, Internal Transcribed Spacer, Small sub-unit 16S ribosomal RNA, Fluorescence *in situ* hybridization, Full-cycle rRNA analysis.

3. Introduction

Protists belonging to the genus *Paramecium* represent a major opportunity for the studies of endosymbiosis in all its relevant aspects: from its establishment and maintenance, to its role in the evolution of the eukaryotic cell [1]. About as many as 60 endosymbiotic bacteria have been discovered so far in *Paramecium* spp. [2] showing different degrees of adaptation to the ciliate's host. In 1983, HD Görtz proposed a classification of the endocytobionts based on the depth of adaptation in the endocytobiont-host system [3, 4], rather than on the nature of its association (e.g. parasitic or mutualistic), that often remains elusive. Taking into account the degree of adaptation, maintenance, and infectivity, endocytobiosis can therefore be divided into three categories: accidental, permanent, and infectious. *Holospora* spp., highly infectious *Alphaproteobacteria* that multiply and grow in the nuclear apparatus of ciliates [5], belong to the last group. These endosymbionts (EBs) are specialized as a role for the infection of either the generative micronucleus (MI) or the somatic macronucleus (MA) [3, 6]. They were observed for the first time in the middle of the XIXth century (Müller 1856; Claparede and Lachmann 1858, 1861; Balbiani 1861; Bütschli 1876) and the first accurate description goes back to 1890, when MWV Hafkine, described and named three nuclear parasites of *Paramecium caudatum* [7] according to the Linnaean binary nomenclature. Since then, eight nuclear endosymbionts showing the same unique morphology and complex life cycle have been further described in *Paramecium* spp. [8-14], summing up to eleven *Holosporas* species to date. Among them, only four species, *Holospora undulata* [15], *Holospora obtusa* [15], *Holospora elegans* [16], and *Holospora caryophila* [16] were validly published [3, 17] and are thus far classified into the genus *Holospora* [17]. They are characterized by two morphologically and functionally distinct stages during their developmental cycle: rod, or spindle-shaped reproductive forms (RFs) and long specialized infectious forms (IFs) of various, species-specific morphology [3, 5, 14, 17]. These alternative forms were observed for the first time in 1969, when LB Preer investigated the ultrastructure of *H. caryophila*, the macronuclear symbiont of *P. biaurelia* (stock 562), using transmission electron microscopy [8]. The fundamental features of the *Holospora* spp., (namely host and nuclear specificity, infectivity and ultrastructure of the infectious form) were therefore outlined for the first time. Despite its bacterial nature, *H. caryophila* was originally named alpha, in conformance with the convention of that time according to which "cytoplasmic gene" were denoted

by Greek letters [15]. Soon after, alpha was named *Cytophaga caryophila* thanks to its resemblance to the gliding bacteria of the genus *Cytophaga* [18]. The name was then changed to *H. caryophila* [16], as a result of the growing knowledge about the biology of this group of infective EBs [3]. One of the most remarkable properties of *H. caryophila* is its ability to infect the newly developing macronuclear anlagen after the degradation of the old macronucleus during the sexual processes [18, 19], reflecting its adaptation to the *P. aurelia* species complex, that inevitably pass through autogamy at regular intervals [19]. This unique survival strategy distinguishes it from all the other EBs so far known [6]. Despite it being the first macronuclear EB of *Paramecium* investigated with the aid of modern techniques [3, 8, 20, 21], a molecular characterization of *H. caryophila* is still missing. On the other hand, the phylogenetic affiliation of one of the EBs discovered by Hafkine, *H. obtusa*, was already determined in 1991 [22]. The molecular phylogeny of two previously described endosymbionts of *P. bursaria*, *H. curviuscula* [23] and *H. acuminata* [24], and the novel EB of *P. jenningsi*, '*Candidatus* Gortzia infectiva', has been recently reconstructed. Moreover, in the latter study, the molecular characterization of the type species of the genus *Holospora*, *H. undulata*, has been carried out, and the systematics of the family *Holosporaceae* revised [14]. A provisional affiliation of *H. elegans* has been proposed as well [14, 23]. The molecular data are growing fast and they will most likely keep up with the available morphological descriptions in the near future. As the comprehension of the family *Holosporaceae* is becoming clearer, an evolutionary understanding of some aspects of their complex life cycle is now possible on a molecular basis. The infection and developmental cycles of *Holospora* spp. have been extensively studied and repeatedly reviewed [3, 5, 17, 25]. These bacteria have evolved a refined mechanism to invade the cytoplasm and the nucleus of the ciliate cell as well as several strategies to escape from them and infect a new host (horizontal transmission), fulfilling a major objective of parasitic bacteria. The IFs of *H. obtusa*, *H undulata*, *H. elegans*, *H. recta*, *H. acuminata* and *H. curviuscula* (**group I**), are collected in the connecting piece of the dividing nucleus while the RFs are segregated into the daughter nuclei [6, 13, 17], culminating in a quantitative separation of the two forms during the host nuclear division. The hypertrophic connecting piece, filled only with IFs, detaches into a vesicle that burst shortly after cytokinesis, releasing its content into the cytoplasm. IFs are then surrounded by one or two membranes and finally released in the external environment *via* the cytoproct by exocytosis [26]. Conversely, IFs of *H. bacillata*, *H. curvata*, *Holospora.* sp. (EB of *P. putrinum*), '*Candidatus* Gortzia infectiva' and *H. caryophila* (**group II**), do not assemble in the connecting piece, being randomly segregated along with RFs into the daughter nuclei [13, 14]. IFs leave the nucleus finding their way to the perinuclear space by crossing the inner membrane of the nuclear envelope, and then to the cytoplasm by vacuole formation from the outer

nuclear membrane [3, 6]. Moreover, while in *H. bacillata*, *H. curvata* and *Holospora*. sp., single IFs are packed in vacuoles that are later released from the cytoproct, *H. caryophila* IFs are packed and released in clusters [6]. On the basis of the abovementioned mechanisms of IFs' release from the host, an evolutionary model reflecting different degrees of adaptation has been proposed. This behavior would have evolved successively from releasing single to several IFs, to the "hijacking" of the host nuclear separation spindle. The ability or inability of the IFs to assemble in the connecting piece led to the outline of two groups of *Holosporas* and the hypothesis of the polyphyletic nature of the genus [13]. Here, for the first time, the molecular description and the phylogenetic affiliation of *Holospora caryophila* are presented, increasing the resolution of the recently reconstructed phylogeny of the family *Holosporaceae* [14, 23, 24]. In this context of growing molecular knowledge about the *Holospora*-like bacteria (HLBs), a more detailed model of the evolutionary relationships within the taxon is provided. This confirms on a molecular basis, previous interpretations on the evolution and classification of this remarkable group of endosymbionts. To achieve these tasks, the full-cycle rRNA approach [27] and the classical comparative analysis of the SSU 16S rRNA gene have been carried out. Fluorescent *in situ* hybridization (FISH) experiments have been performed in order to complement the molecular characterization of the symbiont. Finally, in a subgroup of the investigated *H. caryophila* strains, we have characterized an additional 700 bp downstream the 16S rRNA gene. The analysis of this region gave valuable information about the relationships between strains of *H. caryophila* that naturally infect *P. caudatum* [3, 13, 28] and several *Paramecium* spp. of the *P. aurelia* complex [13].

4. Materials & Methods

4.1 Investigated *Paramecium* strains

The stock 562 of *Paramecium biaurelia*, infected with the type strain of *Holospora caryophila*, was purchased from the American Type Culture Collection (ATCC30694). The type strain was originally collected by G.H. Beale in a stream near the airport at Milano, Italy in 1968 [19, 29]. The other *Paramecium* strains were kindly provided by colleagues. They belong to two morphospecies of *Paramecium*, namely either to the *P. aurelia* species complex or *P. caudatum*. They showed *Holospora*-like bacteria (HLBs) in the macronucleus (MA). According to host spectrum and morphology, these macronuclear bacteria were considered as *H. caryophila*. The investigated strains, their geographic origin and the original collectors are listed in **Table 1**.

Table 1. Investigated stocks of *Paramecium* spp. and proposed name for *H. caryophila* strains

Host	Stock	Origin	Endosymbionts
P. biaurelia	562	Milano, Italy, 1968 (G. Beale), ATCC 30694	*H. caryophila* 562
P. biaurelia	FGC3	Calabria, Italy (S. Galati)	*H. caryophila* FGC3
P. biaurelia	Hc+	Bot. Garden Münster, Germany (S. Fokin)	*H. caryophila* Hc+
P. biaurelia	Cs1	Carolinensiel, Germany	*H. caryophila* Cs1
P. dodecaurelia	UV1-3	Ukraine, 1998 (V. Yakovlev)	*H. caryophila* UV
P. octaurelia	GFg-1	Freiburg, Germany 1998 (A. Potekhin)	*H. caryophila* GFg
P. caudatum	94AB1-5	Boston, USA (I. Skoblo, 1994)	*H. caryophila* AB
P. caudatum	SH42	Süderfahrenstedt, Germany (H.-D. Görtz)	*H. caryophila* SH42

P. caudatum 94ABI-5 and *P. biaurelia* Cs1 have not been cultured, and their 16S rRNA gene sequences have been kindly provided by M. Schrallhammer.

4.2 Cultures screening

Each *Paramecium* culture has been routinely screened for macronuclear (MA) infection of *H. caryophila* either by light microscopy observation or by fluorescent in situ hybridization (FISH) using a high taxonomic level probe. Ciliate cells movement was reduced by storing the microscopic slides at 4 °C prior to *in vivo* observation and micrographs were taken using an Orthoplan Leitz microscope equipped with Nomarski optic and a Canon PowerShot S45 camera. Positive cultures were further analyzed by double hybridization experiments to rule out the occurrence of additional bacterial symbionts inhabiting the same host (see FISH section below for details).

4.3 Monoxenic batch cultures of *Paramecium*

Paramecia were cultured in Cerophyll Medium (CM). Therefore, wheat grass pellets were boiled in distilled water. Salt concentration and pH were adjusted with Balanced Salt Solution (NaCl, $MgSO_4 \cdot 7H_2O$,$MgCl_2 \cdot 6H_2O$, $CaCl_2$, KCl) and Cerophyll buffer (Na_2HPO_4, $NaH_2PO_4 \cdot 2\,H_2O$) respectively. The composition of final CM was as follows: 0.25 % wheat grass, 2.8 mM Na_2HPO_4, 0.85 mM $NaH_2PO_4 \cdot 2\,H_2O$, 1.8 mM NaCl, 1.6 mM $MgSO_4 \cdot 7H_2O$, 4.2 mM $MgCl_2 \cdot 6H_2O$, 0.12 mM $CaCl_2$, 0.31 mM KCl and 0.5µg/mL stigmasterol (modified from ATCC medium 802). The autoclaved medium was inoculated with *Raoultella planticola* DMSZ 3069 (previously known as *Klebsiella planticola*), and incubated overnight at 37°C for optimal microbial growth rate. Every *Paramecium* culture has been fed with bacterized CM under a biological safety cabinet class-II (BSC-II) for sterile handling purpose. Cultures were kept at 19°C in a thermostatic cell.

4.4 Prokaryotic cultures

The Gram-negative *R. planticola* has been cultured on Luria Bertani (LB) agar plates as well as in LB liquid cultures at 37°C. Uninoculated plates and tubes have always been incubated as a negative control. After incubation, the plates have been kept at 4°C and were renewed once a month to avoid nutrients depletion and subsequent cell death. In order to have a continuous supply of bacterial suspensions at the same concentration, stock aliquots were prepared as follows: optical density of

liquid culture was checked at 600 nm and LB medium removed by 5 min centrifugation of the samples at 8,000 RCF. Pelleted cells were then resuspended in $1/4^{th}$ of the initial volume of CM and mixed 1:1 with a sterile glycerol solution (50 % V/V in distilled water). Glycerol stocks were kept at -20°C to strongly reduce bacterial metabolism and preserve the ability to grow over time after medium inoculation. Finally, aliquots were centrifuged 2 times at 8,000 RCF for 5 min to remove glycerol and resuspended in sterile CM before the *inoculum.*

4.5 DNA isolation

Total cellular DNA extraction was performed using the cationic detergent Cetyl Tri-methyl Ammonium Bromide (CTAB) at high ionic strength (modified version of a *Doe Joint Genome Institute* protocol). 100-150 manually collected *Paramecium* cells were washed and then centrifuged 5 min at 8,000 RCF in a bench centrifuge (eppendorf, Centrifuge 5424 R). The pellet was resuspended in 340 µl TE buffer (10 mM Tris base, 1 mM EDTA and pH 8.0). After the subsequent addition of 400 µl buffer PL1 (NucleoSpin® Plant II Kit, Macherey-Nagel) cells were mechanically lysed by shearing. Following homogenization of the sample, 20 µl lysozyme (100 mg/ml), 40 µl 10 % (w/v) SDS and 8 µl Proteinase K (10 mg/ml) were added to the solution. The mixture was incubated for 1 h at 37 °C for optimal enzyme activity. Ionic strength was adjusted with 100 µl of 5 M NaCl, mixed with 100 µl of pre-heated (65 °C) CTAB and incubated 10 min at 65 °C. CTAB complexes and other lipophilic compounds were removed by liquid-liquid extraction using a mixture of organic solvents (500 µl of Phenol: chloroform: isoamylalcohol 25:24:1). The aqueous phase was collected after centrifugation (max speed for 10 min) and the organic phase discarded. 500 µl of Chloroform: isoamylalcohol were added to avoid carry-over of phenol in the freshly collected aqueous phase. Nucleic acids were then precipitated with 0.6 volume of cold (20 °C) 2-propanol for 30 min at 20 °C. DNA was collected by 15 min of centrifugation at maximum speed, washed with 25 µl of 70 % ethanol, air dried and resuspended in 99 µl of TE buffer. RNAse A (1 µl, 10 mg/ml) was added to the solution and the sample was stored at 4 °C.

4.6 Polymerase chain reaction and sequencing

Two regions of the symbionts' ribosomal operon have been targeted for DNA amplification and sequencing: the 16S rRNA gene and the ITS (Internal Transcribed Spacer), the latter located between the two routinely employed phylogenetic markers 16S rDNA and 23S rDNA. DNA was characterized as follows (standard workflow): Initial *in vitro* amplification using universal primers (or a combination of a universal primer and a taxon-specific one) followed by semi-Nested PCR, amplicon purification and direct-sequencing of the purified PCR product (pPCR). PCR was carried out in a C1000™ Thermal Cycler (BIO-RAD, Hercules, CA) using primers available from previous studies [14, 30, 31] (**Table 2**).

Table 2. List of primers used to characterize the 16S rRNA gene, the ITS and a portion of the 23S rRNA gene of the investigated *H. caryophila* strains

Primer Name	Sequence	Reference
16S F1256 NovHolo	5'-ATGCAATAGGGTGACCTGG-3'	This study
16S F1142 Holo	5′-GAGAACTTTAAGAAGACTGCC-3′	[14]
16S F114 HoloCaedi	5′-TGAGTAACGCGTGGGAATC-3′	[14]
16S F49 AlphaSym	5'-TAACACATGCAAGTCGAAC-3'	This study
16S F7a	5'-AGAGTTTGATCCTGGCTCAG-3'	[30]
16S F363 NovHolo	5'-GGACAATGGGGGAAACC-3'	This study
16S R537 NovHolo	5'-GAACAACGCTTGATCCCTTC-3'	This study
16S R660 NovHolo	5'-GAATTCCACTTTCCTCTCTC-3'	This study
16S R700 NovHolo	5'-GGTATTCCTCCTAATATCTGC-3'	This study
16S R1488 Holo	5′-TACCTTGTTACGACTTAACC-3′	[14]
16S R1328 HoloCaedi	5′-TAGCGATTCCAACTTCATG-3′	[14]
16S R1334 HoloCaedi	5'-GATTACTAGCGATTCCAAC-3'	This study
16S R1522a	5'-GGAGGTGATCCAGCCGCA-3'	This study
16S R1222b	5′-GGAGGTGATCCAACCGCA-3′	[31]
23S R457 WL992r	5'-CTTTCCCTCRCGGTACT-3'	This study

The Forward primer 16S F1256 NovHolo has been specifically designed for ITS characterization using the probe design tool of the ARB software [32] and NetPrimer (Copyright © 2009 by PREMIER Biosoft International). Amplifications were performed with the high fidelity Takara's Ex Taq™ DNA Polymerase (TaKara Bio Inc., Otsu, Japan). Touch-Down thermo-cycling profile [33]

and hot start [34, 35] have been routinely carried out to increase amplification specificity. PCR conditions used to characterize the ITS rDNA of *Holospora caryophila* 562 (Type Strain; ATCC 30694) are reported below. A fragment encompassing the ITS was first amplified using 16S F1142 Holo and 23S R457 WL992r primers pair. A second PCR was carried out on the pPCR from a previous amplification reaction using 16S F1256 NovHolo and 23S R457 WL992r primers pair (see Appendix, 10.2). The forward primer 16S F1256 NovHolo was used as a sequencing primer as well. PCR conditions were initially 10 min at 94 °C followed by 5 cycles for 30 s at 94 °C, 30 s at 60 °C (first annealing temperature; Ta) and 90 s at 72 °C, 10 cycles for 30 s at 94 °C, 30 s at 58 °C (second annealing temperature; Ta') and 90 s at 72 °C, 15 cycles for 30 s at 94 °C, 30 s at 55 °C (third annealing temperature; Ta"), 90 s at 72 °C and finally 10 min at 72 °C for synthesis completion. The Master Mix composition used is reported in detail in the Takara Ex Taq™ Polymerase specification sheet. PCR reaction products were purified applying chromatography columns (NucleoSpin® Extract II, Nucleic Acid and Protein Purification Kit, Macherey-Nagel) and eluted in distilled water. The DNA was send to either of the following sequencing companies: Macrogen Korea (Gasan-dong Geumchen-gu Seoul, Korea), Eurofins MWG Operon (Ebersberg, Germany). Samples were prepared as recommended by the specific company and service used. Retrieved electropherograms were manually inspected and carefully assembled.

4.7 Gel electrophoresis

Electrophoretic runs were performed using three slightly different set-ups and reagents, depending on the laboratory standards adopted. As running buffer was used TBE or TAE buffer. A 0.8-2.0 % (W/V) agarose gel was employed according to the resolving power needed. Either sucrose based solution or Glycerol solution, containing Bromphenol Blue and Xylene Cyanole were used as loading buffers and premixed with 5 µl DNA sample before loading of the gel. When Syber®Gold was used as DNA staining reagent, 3 µl of the dye were also premixed with the sample before running. As an alternative, 1 µl of staining dye was freshly added to the gel casting solution when GelRed™ was used. Finally, when Ethidium Bromide (EtBr) was used as a fluorescent dye, the gel was incubated for 30-45 min in a 5 µg/ml EtBr solution (in TBE) in order to stain DNA fragments following the electrophoretic run. The gels were exposed to 420-500 nm light under a non-UV Transilluminator (Dark Reader™) or 302 nm transmitting light under a BIO-RAD UV Transilluminator depending upon the DNA staining dye used. Gels were photographed with the integrated camera Olympus C-5060 (High quality lens). Electrophoretic runs have been prolonged

for 0.5-3 h according to amplicons size and carried out in a PEQLAB Biotechnologie GmbH (40-0708) electrophoretic cell at 50-120 V using the Gene Power Supply GPS 200/400 Pharmacia.

4.8 Fluorescent *in situ* hybridization (FISH)

Small volumes of medium were transferred from culture flasks to Sonneborn's slide. Cells were washed in volvic® water and dropped on a microscopic slide where FISH was carried out. Cells were fixed with 4 µl of 4 % paraformaldehyde in PBS (137 mM NaCl, 2.7 mM KCl, 30 mM Na2HPO4·2H2O, 6 mM KH2PO4, pH 7.4) and washed with 95 % ethanol. Permeabilization was achieved by performing a three step ethanol gradient (10 min each, 50-80-100 % V/V). Hybridization was performed according to Manz et al. [36] without formamide. Slides were exposed to the hybridization buffer (900 mM NaCl, 20 mM Tris-HCl, 0.01 % SDS) containing 0.5 pmol/µl of cy3 and FITC labeled probes, for 3 hours at 46 °C. After hybridization cells were washed for 20 min at 48 °C in washing buffer (900 mM NaCl, 20 mM Tris-HCl, 0.01 % SDS). FISH experiments were carried out using the *Holospora caryophila* specific probe NovHolo1257 and the universal eubacterial probe EUB338 [37]. For inspection, a Zeiss AxioVert 200M epifluorescence microscope (Zeiss Axioplan, Oberkochen, Germany) equipped with a Zeiss AxioCam MRm camera was used.

4.9 Oligonucleotide probe design

Two strain specific probes, HcSH42R1004 (matching the SSU rRNA of SH42\AB\GFg-1) and Hc562R1004 (matching 562\Hc+\CS1\UV) were designed to discriminate between this subgroups of *H. caryophila*. Probes were rationally designed according to the guideline of Yilmaz et al. [38] and Behrens et al. [39] using the probe design module of the ARB package [32]. Target selectivity was verified by querying GenBank and RDP [40, 41] with the chosen probe sequences. When FITC was used as a fluorophore for labeling the oligonucleotides at the 5'-end, guanosine (dG) and citidine (dC) were not placed as terminal residues to avoid fluorescence quenching after hybridization with the target [42].

4.10 Comparative analysis of the 16S rRNA gene

Phylogenetic reconstruction was based on the 16S rRNA marker. The dataset included the 16S rRNA of seven strains of *H. caryophila* (**Table 1**), 33 closely related sequences of bacteria belonging to the order Rickettsiales and 7 sequences belonging to the class *Alphaproteobacteria*, the latter used as outgroup. Sequences lengths have been reduced to the shortest present in the dataset. Multiple sequence alignment (MSAs) of rRNA sequences has been performed using the accurate high-throughput MSAs tool Silva Incremental Aligner (SINA) [43] available at SILVA, the comprehensive ribosomal RNA database. This tool, provided by the SILVA ribosomal RNA project [44], is specifically designed for rRNA alignment against a database (r115) containing around 500,000 high quality, nearly full length and aligned SSU rRNA sequences. The MSAs' accuracy evaluation was provided by SINA as an alignment score for each sequence. The alignment was manually checked and refined. Stationary test for homogeneity of base composition and evaluation of the amount of phylogenetic information by quartet puzzling of the input dataset were performed using TREE-PUZZLE [45] v.5.2. The input nucleotide matrix contained 1351 columns of which 49.3 % were constant sites and 50.7 % were site patterns (informative residues). The best fitting parametric nucleotide substitution model has been selected evaluating AIC (Akaike 1973), BIC (Schwarz 1978) and DT (Minin et al. 2003) information criteria analysis implemented in jModelTest [46] v.2.1.4. Molecular phylogeny has been inferred using two different reconstruction methods, the distance-based Neighbors Joining (NJ) [47] and the sequence-based Maximum likelihood (ML) [48] implemented in the PHYLIP package [49] v.3.695 and PHYML program [47] v.3.0 respectively. Distance matrix used as input for the NJ reconstruction has been calculated employing the F85 model, using custom parameters previously estimated by jModelTest v.2.1.4 (see Appendix, 10.3). The statistical significance of the inferred tree topology has been evaluated by three independent methods: non-parametric bootstrap analysis [50, 51, 52], Shimodaira and Hasegawa (SH) significance test [52], and Bayesian inference (posterior probabilities) [53], all of which are implemented in the PHYML v.3.0 program. The robustness of the ML-tree has been assessed with 1,000 pseudoreplicates and 10 jumbles (number of random starting trees). Newick format of the phylogenetic tree has been visualized and customized using iTOL [54].

4.11 ITS analysis

700bp of the ribosomal operon downstream the 16S rRNA gene have been sequenced in the *H. caryophila* strains 562, FGC3, Hc+, UV, GFg-1, and SH42. This region consisted of the Internal Transcribed Spacer (ITS) of 390-396 bp, and approximately 310 bp of the 5' end of the 23S rRNA gene. A similarity matrix was computed from multi-aligned sequences for the 16S gene, the ITS region and the full-length region of about 2,160bp (16S rRNA-ITS-5'23S rRNA). Computation was performed using dnadist implemented in the Felsenstein Package PHYLIP v.3.695. The average pairwise distance of each strain was plotted to visualize the information contained in the similarity matrices and highlight the relative distance between strains. Distances were computed for all the above mentioned similarity matrices and divided by the median of each distribution. Values were plotted in a "box and whiskers" graph in order to catch possible outliers (see Appendix, 10.5).

5. Results

5.1 Molecular characterization of *Holospora caryophila*

All investigated *Paramecium* strains were infected with bacteria that clearly showed the characteristic features (e.g. nuclear specificity, two distinct morphological forms) and the life cycle (see Introduction) typical of endosymbionts belonging to the family *Holosporaceae* [17]. According to the host spectrum (*P. aurelia* complex and *P. caudatum*), the macronuclear (MA) specificity and the crescent (sickle) shaped reproductive forms [29], these bacteria were identified as *H. caryophila* (**Figure 1**).

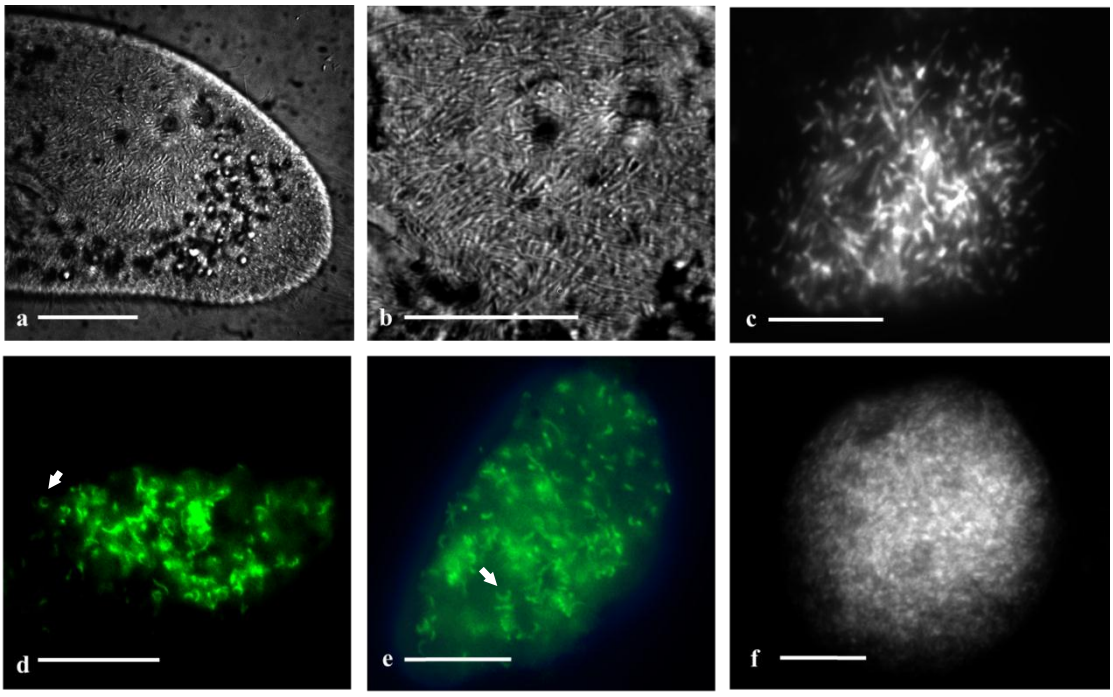


Figure 1. Micrographs of infected *Paramecium* cells. **a, b**) *In vivo* visualization of *H. caryophila* inside the macronucleus (MA) of *P. biaurelia* HC+. **a**) 400x magnification showing the bacteria inhabiting the MA. **b**) Higher magnification (1000x) of the same visual field showing long spiral-shaped infectious forms (IFs). **d, e**) Fluorescent *in situ* hybridization (FISH) of *P. biaurelia* FGC3 using Fluorescein-labeled EUB338 probe. Crescent shaped reproductive forms (RFs) are clearly distinguishable (arrows). **e**) Signal from EUB338-Fluos has been merged with that of DAPI to demonstrate the localization of the bacteria in the MA. **c, f**) FISH of *P. biaurelia* 562 using the same eubacteria-targeted probe showing the type-strain of *H. caryophila* in the MA. The cell in **c**) is infected by a mixed population of IFs and RFs while the MA in **f**) is prevalently infected by numerous RFs. Bars, 20 µm.

After morphological identification of the symbionts, the 16S rRNA gene sequences of the eight *H. caryophila* strains studied (*see* Table 1; 4.1), including the type strain from *P. biaurelia* stock 562 were characterized. On the basis of the obtained sequences, a complementary oligonucleotide probe (NovHolo 1257; **Table 3**, 5.3) was specifically designed to target the SSU rRNA of the endosymbionts and therefore unambiguously assign the newly retrieved sequence to *H. caryophila*. (*See* Appendix; 10.4**)**. The eubacterial probe EUB338 (Fluos-labeled) and the *H. caryophila* specific probe NovHolo1257 (cy3-labeled) were used in double hybridization experiments. Signal from both probes was detected and co-localized in the MA (**Figure 2**). Bacteria localized in the phagosomes were only positive for EUB338 but not for NovHolo1257, thus probably deriving from food bacteria. The probe NovHolo1257 labeled only *H. caryophila* in the MA, hence demonstrating its specificity during hybridization reactions (**Figure 2**; k, l).

Preliminary evaluation of the 16S rRNA gene similarity has been performed using the alignment tool BLAST®n [55] available at NCBI [56]. The analysis revealed only a similarity of 86.2 % between the type strain of *H. caryophila* (562) and *H. undulata* (GenBank: HE797906), the type species of the genus *Holospora*. The highest similarity reported by BLASTn was shared, however, with uncultured bacteria from environmental samples taken in a freshwater reservoir (Bantou) (Zhang,Y., Hu,A., Liu,L. and Yang,J. Bacterial community composition in four freshwater reservoirs; unpublished). These sequences could represent *Holospora*-like bacteria uncharacterized thus far.

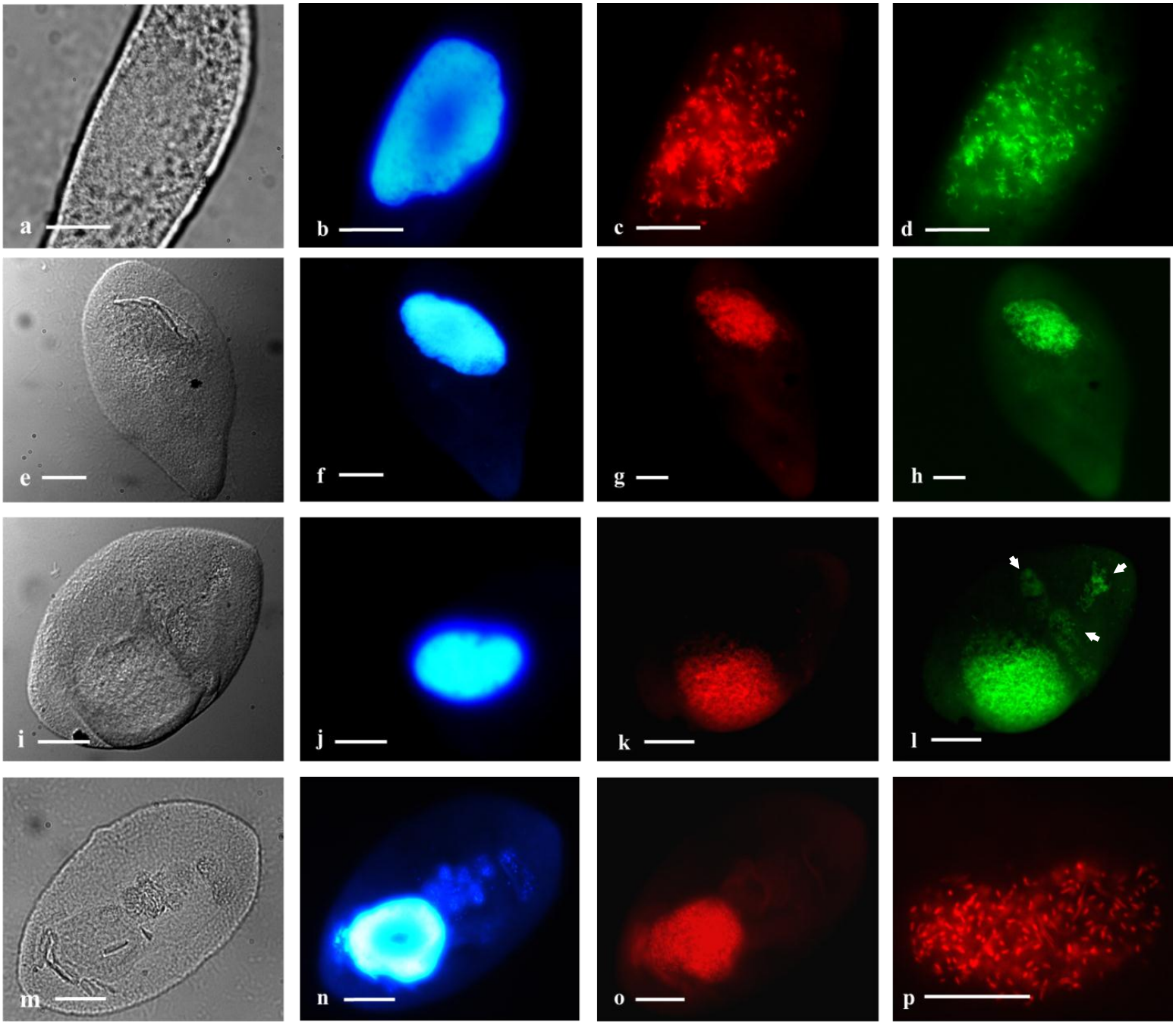


Figure 2. Fluorescence *in situ* hybridization (FISH) of *P. biaurelia* FGC3 (top row; **a-d**), *P. octaurelia* GFg-1 (Second row from the top; **e-h**), *P. biaurelia* Hc+ (third row from the top; **i-l**) and *P. dodecaurelia* UV (bottom row; **m-p)**. First column from the left (**a, e, i, m**): Differential Interference Contrast microscopy (DIC) showing the *Paramecium* cells. Second column from the left (**b, f, j, n**): DAPI staining highlighting the MA and food vacuoles containing bacteria. Third column from the left (**c, g, k, o**): Positive signal from NovHolo1257-cy3. In **d, h, l,**): Positive signal from EUB338-Fluos. **p**) Higher magnification of a different *P. dodecaurelia* UV cell showing a mixed population of IFs and RFs of *H. caryophila* in the MA. Bacteria contained in the food vacuoles were labeled only with the eubacterial probe EUB338 (arrows in **l**) but not with the *H. caryophila* specific probe NovHolo 1257 (**k**). Bars, 20 μm.

5.2 Relationships between *Holospora caryophila* and other *Holospora*-like bacteria

The stationary test for homogeneity of base composition (chi-square test, $\alpha = 0.05$) performed using TREE-PUZZLE v.5.2 revealed a statistically significant difference in base composition of *Holospora caryophila's* 16S rRNA from the other *Rickettsiales* included in the data set. However, the difference in G+C content of every analyzed sequence ranged within 5 % and did not require the use of a non-homogeneous nucleotides substitution model for unbiased phylogenetic inference [57]. Quartet puzzling of the input dataset showed that 93.5 % of the quartets had a well-defined topology and only 3.6 % were unresolved quartets, reflecting high phylogenetic information content. The best fitting parametric nucleotide substitution model selected by jModelTest v.2.1.4 was the GTR + I + G (*see* Appendix, 10.3). The sequence of *H. caryophila* FGC3 was identical to the type strain, therefore the two sequences have been labeled *H. caryophila* 562 | FGC3 although belonging to bacterial strains retrieved from different *Paramecium* stocks.

The molecular phylogenetic analysis based on the 16S rRNA gene sequences revealed the phylogenetic position of *H. caryophila*. It's affiliated to the order *Rickettsiales* within the *Alphaproteobacteria*. All here characterized strains cluster together and form a monophyletic group with high statistical support. This group is affiliated with one of the two major divisions within the order *Rickettsiales* (**Figure 3**). The overall topology of the inferred phylogenetic tree (**Figure 3**) is consistent with the previously published [14], besides minor differences.

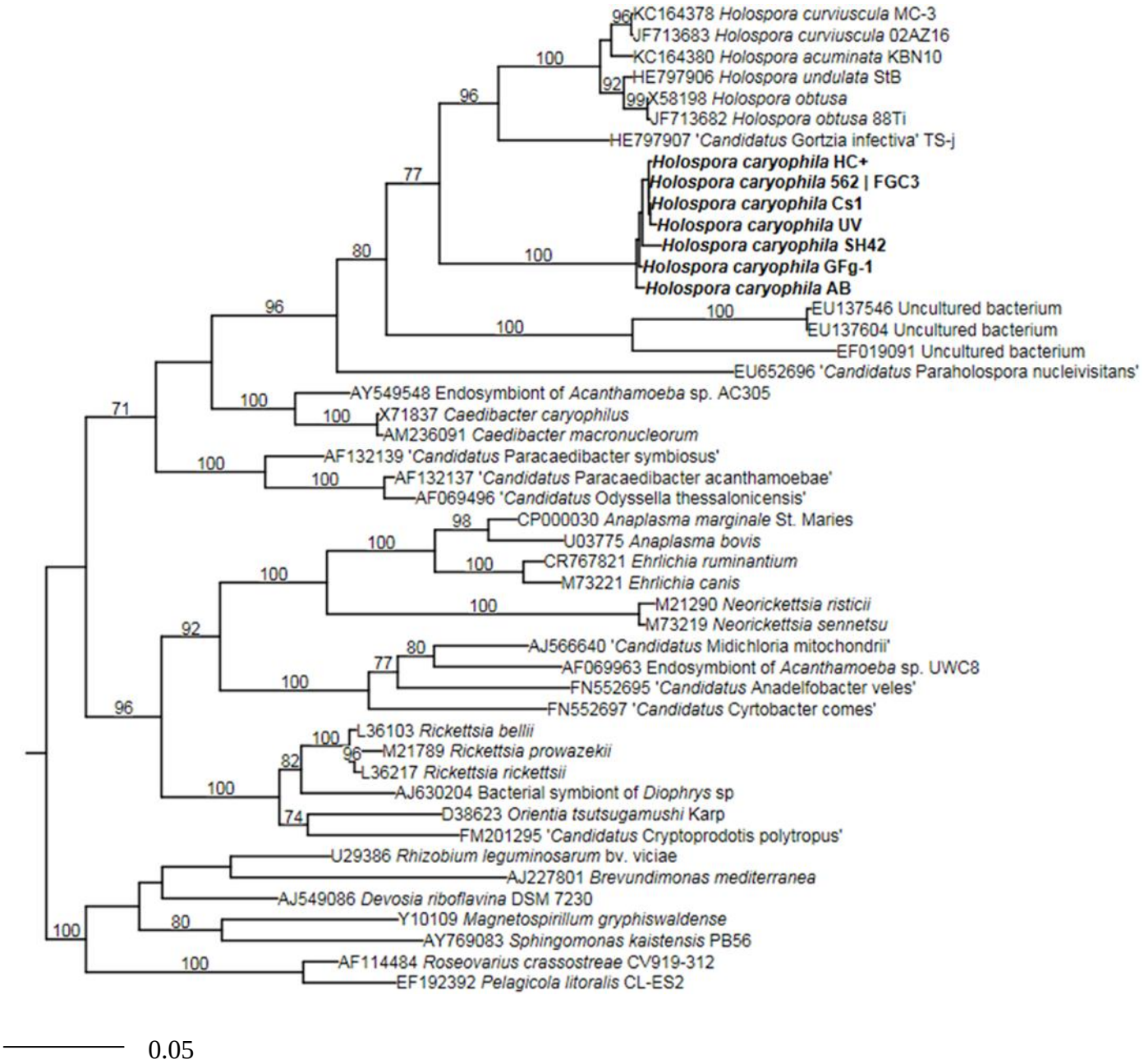


Figure 3. Maximum likelihood (**ML**) phylogram showing the phylogenetic position of *Holospora caryophila* within the order *Rickettsiales*. Bootstrap values higher than 70 % are displayed on the internal nodes. Newly characterized sequences are highlighted in *bold*. The bar represents sequence divergence expressed in nucleotide substitution per site.

Every highly supported node (bootstrap value higher than 75 %) in the maximum likelihood (**ML**) tree was recovered in the one inferred with the neighbor joining (**NJ**) method (data not shown) reflecting the robustness of the reconstructed global topology. All the recently characterized species of HLBs [14, 23, 24] have been included in the phylogenetic reconstruction to obtain a detailed and updated description of the evolutionary relationship within the family *Holosporaceae*. *H. caryophila* falls basally to the HLBs characterized so far resulting even more distantly related to

the "classical *Holosporas*" comparing to the macronuclear symbiont of *Paramecium jenningsi* "*Candidatus* Gortzia infectiva" (**Figure 4**).

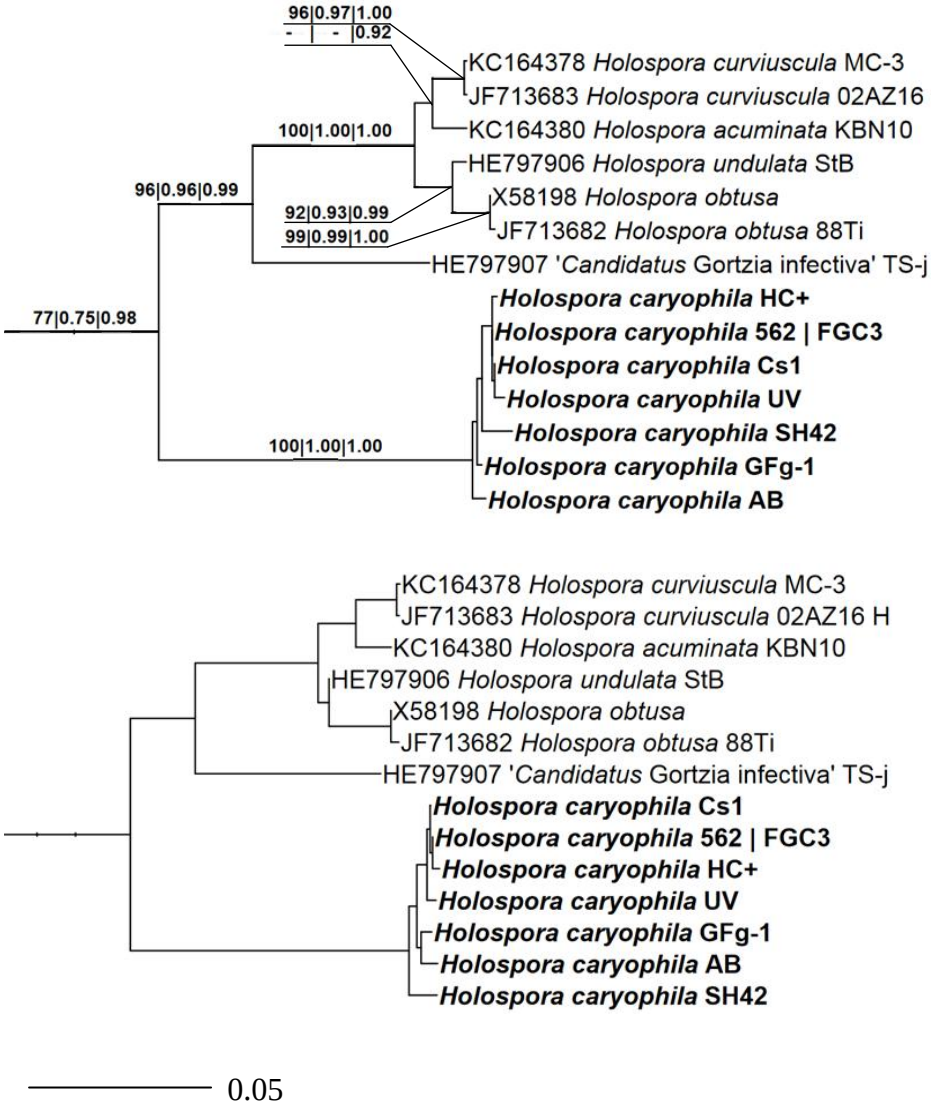


Figure 4.The HLBs clade (zoom-in) inferred using different reconstructions methods ML (top) and NJ (bottom) respectively. The cluster of *H. caryophila* strains branched basally to the HLBs previously described as a highly divergent species. In the ML tree the significance of the inferred topology has been assessed using three different methods and the resulting values are shown on the nodes of interest. Bootstrap value (higher than 70 %), SH test ($\alpha = 0.05$) probability and Bayesian support probability are shown from left to right respectively. The latter two values are more conservative, in other words are less likely to reject the inferred topology [52]. The bar represents sequence divergence expressed in nucleotide substitution per site.

The recovered local topology within the cluster of *H. caryophila* strains differs depending on the treeing method used and showed low bootstrap values in the ML tree, probably because the relative order of neighboring branches was determined by only a small number of differences (ranging from 1 to 22 over 1351 aligned character) [51]. Furthermore the 16S rRNA marker was shown in several cases to lack the resolution power to discriminate between closely related species (Fox et al., 1992; Stackebrandt and Goebel, 1994 [58, 59]) and it cannot therefore be expected to discriminate between the analyzed strains. Nevertheless both reconstructions suggested a divergence of *H. caryophila* strains SH42, GFg-1 and AB from the "type strain cluster" *H. caryophila* 562, Cs1, Hc+ and UV.

The similarity matrix (**Figure 5),** computed on the aligned input sequences is shown for the HLBs subset. Each *H. caryophila* strain shares similarity values equal or lower than 87.8 % with the members of the genus *Holospora* accounted for in the analysis, including the type species *H. undulata*. Furthermore they share similarity values equal or lower than 88.1 % with "*Candidatus* Gortzia infectiva". These observations, together with other evidences, would justify the removal of *Holospora caryophila* from the *Holospora* genus and its transfer to a new genus as it will be discussed later.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1 *H. caryophila* SH42	**100**													
2 *H. caryophila* Cs1	**98.8**	**100**												
3 *H. caryophila* AB	**98.4**	**99.2**	**100**											
4 *H. caryophila* GFg-1	**98.9**	**99.3**	**98.8**	**100**										
5 *H. caryophila* UV	**98.4**	**99.7**	**98.9**	**98.9**	**100**									
6 *H. caryophila* 562 | FGC3	**98.8**	**99.9**	**99.1**	**99.3**	**99.6**	**100**								
7 *H. caryophila* Hc+	**98.6**	**99.8**	**99.0**	**99.2**	**99.5**	**99.8**	**100**							
8 '*Cand.*Gortzia infectiva' TS-j	87.7	88.1	87.9	87.9	87.8	88.0	87.9	100						
9 *H. undulata* StB	87.5	87.5	87.5	87.8	87.3	87.5	87.3	91.5	100					
10 *H. obtusa* 88Ti	87.1	87.4	87.2	87.6	87.2	87.3	87.2	91.1	98.4	100				
11 *H. obtusa*	87.2	87.6	87.4	87.7	87.3	87.5	87.3	91.2	98.6	99.8	100			
12 *H. acuminata* KBN10-1	87.1	87.4	87.2	87.4	87.2	87.3	87.2	91.6	97.4	96.5	96.7	100		
13 *H. curviuscula* MC-3	87.6	87.6	87.4	87.6	87.4	87.5	87.4	91.3	97.5	96.7	96.8	98.0	100	
14 *H. curviuscula* 02AZ16	87.6	87.6	87.4	87.6	87.4	87.5	87.4	91.3	97.5	96.5	96.7	98.0	99.8	100

Figure 5. Similarity matrix (16S rRNA) of members of the family *Holosporaceae*. Numbers represent pairwise similarities expressed as a percentage. Highlighted in bold: similarity values between strains of *H. caryophila* (characterized sequences).

It's noteworthy, that *H. caryophila* SH42 infecting *P. caudatum* strain SH42, showed an average distance of the 16S rRNA sequence from all the other strains (Cs1, AB, GFg-1, UV, 562 | FGC3 and Hc+) of 98.65 %, higher than the informally proposed conservative threshold for the definition of a new species of 97 % (Stackebrandt and Goebel, 1994 [58]) but lower than the recently revised taxonomic parameter range of 98.7 − 99 % (Stackebrandt and Ebers, 2006 [59]).

5.3 Relationships between *H. caryophila* strains

To increase the resolution of the analysis within the "*H. caryophila* group", the complete Internal Transcribed Spacer (ITS) of 390-396 bp length and a fragment of ca. 310 bp at the 5' end of the 23S rRNA gene of strains SH42, GFg-1, 562, FGC3, UV and Hc+ were characterized. The type strain 562 showed 100 % similarity with *H. caryophila* FGC3 over the entire sequenced region (16S rRNA-ITS-5'23S rRNA) therefore the label 562 | FGC3 refers to the sequence of both strains retrieved from different *Paramecium* stocks. The similarity matrix computed on the aligned ITS sequences is shown in **Figure 6.**

Strain	1	2	3	4	5
1 *H. caryophila* SH42	100				
2 *H. caryophila* GFg-1	95.1	100			
3 *H. caryophila* UV	94.5	99.0	100		
4 *H. caryophila* 562 | FGC3	94.6	98.7	99.7	100	
5 *H. caryophila* Hc+	94.6	98.7	99.7	100	100

Figure 6. ITS similarity matrix of six *H. caryophila* strains.

The ITS of 562 and Hc+ showed 100 % sequence similarity despite slightly diverging (0.2 %) 16S rRNA sequences (**Figure 5**). The ITS of SH42 was the most diverging within the investigated group sharing a similarity equal or lower than 95.1 % with the other strains. The normalized average pairwise distance of each strain was plotted to visualize the information contained in the similarity matrices and highlight the relative distances (**Figure 7**).

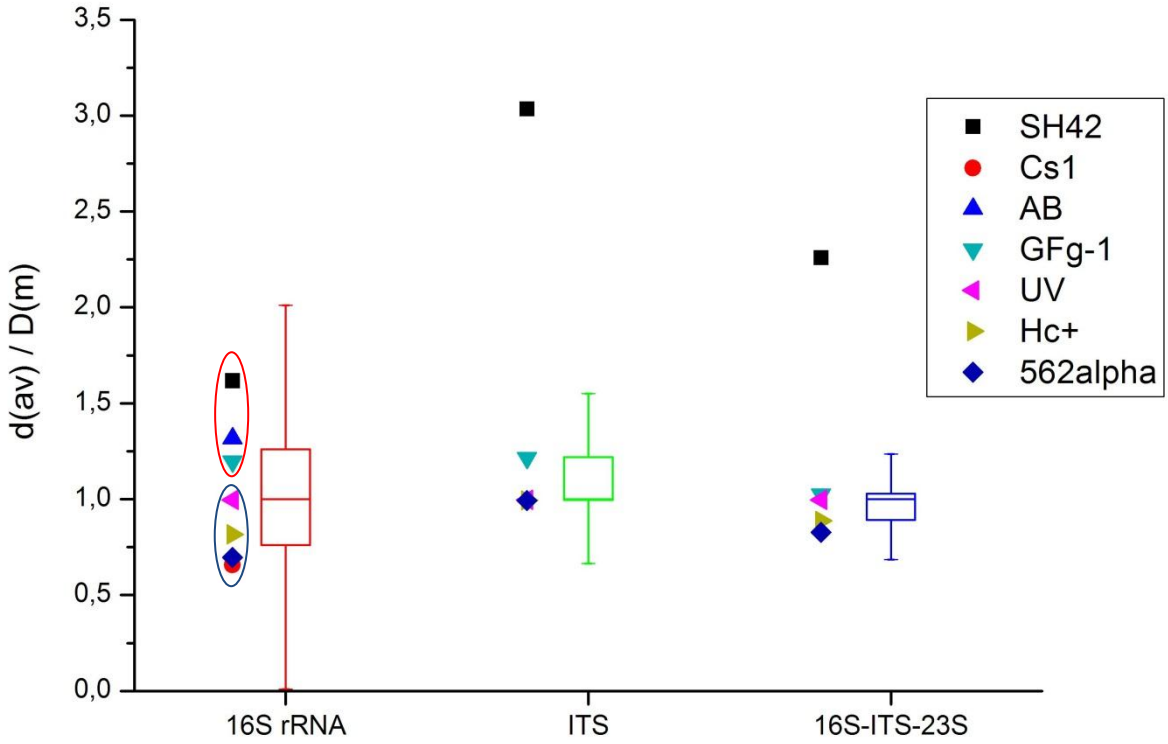


Figure 7. Average distances of each *H. caryophila* strain from the others. Data were taken from the 16S rRNA, ITS and 16S-ITS-5'23S similarity matrix respectively. Box border: lower quartile (Q_1) and upper quartile (Q_3). Whiskers length: $1.5 * r$ from Q where $r = Q_3 - Q_1$. Blue and red circles outline the two groups of *H. caryophila* strains discriminated by FISH (Figure 8). $\boldsymbol{d(av)}$ = average pairwise distance. $\boldsymbol{D(m)}$ = Median of each distribution. See Appendix (10.5) for a detailed description of the computation and interpretation of the plot.

The distribution of the average pairwise distances concerning the 16S rRNA region, highlighted three *H. caryophila* strains, namely SH42, AB and GFg-1 (red circle in **Figure 7**) showing different degree of divergence relative to the other strains. Moreover, *H. caryophila* SH42 was identified as an "outlier" taking into account the more variable region of the ITS and the complete sequenced region 16S rRNA-ITS-5'23S rRNA. In both distributions (**Figure 7**) SH42 fell outside the inner ($r \pm 1.5r$) and the outer fence ($r \pm 3r$; not show), reflecting its relatively strong divergence in these phylogenetic markers from all the other strains. The divergence in the 16S rRNA sequences of the two subsets of *H. caryophila* (highlighted in the "box and whiskers" plot [60], **Figure 7**) was used to design two oligonucleotide probes, able to further discriminate *H. caryophila* into two groups of strains, SH42|AB|GFg-1 and 562|FGC3|Hc+|Cs1|UV respectively.

H. caryophila FGC3 and SH42 were chosen as representatives of each group, and double hybridization FISH was carried out using the newly designed probes Hc562R1004 and HcSH42R1004 (**Figure 8**).

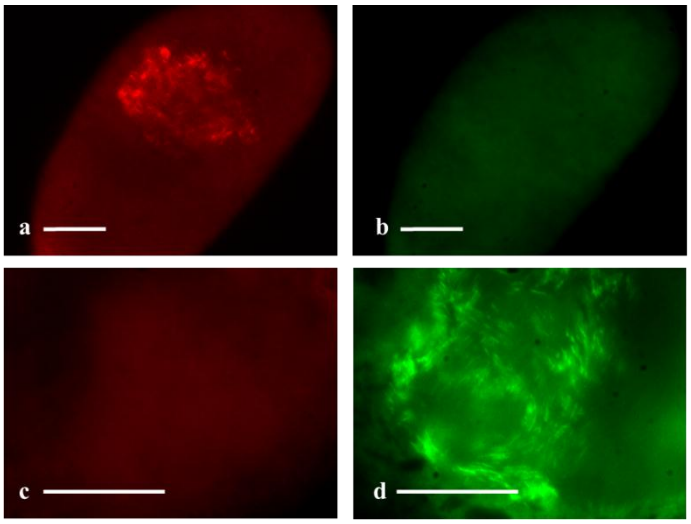


Figure 8. Fluorescent *in situ* hybridization (FISH) of *P. biaurelia* FGC3 (**a, b**) and *P. caudatum* SH42 (**c, d**). **a** and **b**: cohybridization with probes Hc562R1004 and HcSH42R1004; positive signal was detected with probe Hc562R1004-cy3 (**a**) but not with HcSH42R1004-Fluos (**b**) (see also Appendix, 10.7). **c** and **d**: cohybridization showing positive signal from probe HcSH42R1004-Fluos (**d**); signal from probe Hc562R1004-cy3 was not detected (**c**). Bars, 20 µm.

The probe Hc562R1004 (**Table 3**) labeled only *H. caryophila* FGC3 in the MA of *P. biaurelia* (a) but was not able to hybridize with the 16S rRNA of *H. caryophila* SH42 infecting the MA of *P. caudatum* (c). Complementary, probe HcSH42R1004 (**Table 3**) labeled only *H. caryophila* SH42 (d) but not *H. caryophila* FGC3 (b), thus demonstrating the possibility to discriminate between subgroups of *H. caryophila* strains by FISH (**Figure 8**).

Table 3. List of oligonucleotides probes used in FISH experiments[a]

Probe Name	Sequence	Reference
Hc562R1004[b]	5'-TGAAAAATCTAATCTCTTAAATTTAAAATC-3'	This study
HcSH42R1004[b]	5′-TGAAAGATCTAATCTCTTAAATCTAAA-3′	This study
NovHolo 1257	5′-CCAGGTCACCCTATTGCA-3′	This study
EUB338	5'-GCTGCCTCCCGTAGGAGT-3'	[37]

[a] Every FISH in this study have been performed without formamide in the hybridization buffer
[b] Nucleotides differences between subgroup-specific probes are highlighted in black

6. Discussion

The molecular characterization of *Holospora caryophila* and the phylogenetic reconstructions based on the 16S rRNA herein presented, confirm the affiliation of this endosymbiont with the family *Holosporaceae*. This family is part of the *Rickettsia*-like endosymbionts (RLE) group, one of the two major divisions of the order *Rickettsiales* [61]. At present, in agreement with the original description (ex Preer 1969, Preer and Jurand 1974, Preer and Preer 1982), *H. caryophila* is affiliated with the genus *Holospora* on the basis of the infectious forms' (IFs) ultrastructure, life cycle peculiarities, and nuclear specificity. It was however hypothesized long ago (according to the horizontal transmission mechanism of *H. caryophila*, *H. bacillata* and *Holospora* sp.) that a subset of *Holospora*-like bacteria (HLBs) could have been "ancestral", that is, less specialized for the parasitic life style comparing to the "classical *Holosporas*" [6, 13]. Considering their differential behavior during the host nuclear division, *Holosporas* were indeed divided into two groups, as emphasized in the introduction of this study. Additionally, in comparison with the strictly host specific Group I members, (regarded to be highly advanced parasites), the broader host spectrum of two species of the "primitive" group supported the abovementioned subdivision. In particular, *H. bacillata* was found in the macronucleus (MA) of *P. nephridiatum* [3, 6, 13] as well as in that of *P. calckinsi* [12], while *H. caryophila* was described as a natural parasite of *P. biaurelia* [8], *P. novaurelia* [13] and *P. caudatum* [28]. Here we report isolates of *P. dodecaurelia* and *P. octaurelia* naturally infected by *H. caryophila*, further widening its host spectrum. Moreover, Group II *Holosporas* could not be labeled by the *Holospora*-specific probe H16-23a in FISH experiments [13], corroborating the proposed classification. It should be noted, however, that FISH results have only a supportive "phylogenetic value", which can sometimes be misleading. An example of this is shown in the case of the negative response of *H. caryophila* to the *Alphaproteobacteria*-specific probe ALF1b, correctly reported by Fokin and co-authors [6, 13] (see Appendix, 10.6). Each of the eight investigated *H. caryophila* strains, including the type strain from *P. biaurelia* stock 562, showed 16S rRNA similarity values equal or lower than 87.8 % with the members of the genus *Holospora* described (on a molecular level) up to now, including the type species *Holospora undulata*. These values are quite far below the threshold for genus definition of 95 % informally proposed by *Ludwig* and co-workers [51]. Considering the aforementioned modest similarity values, together with the inability to "induce" the connecting piece (CP), *H. caryophila* should be removed from the *Holospora* genus and transferred to a "*genus novum"* (*see* Appendix, 10.1).

We therefore propose the establishment of a new taxon *Preeria* gen. nov. and the new combination *Preeria caryophila* (Preer and Preer 1982) comb. nov., in honor of Louise B. Preer and John R. Preer Jr., the American scientists that in 1969 described for the first time the ultrastructure and the life cycle of *H. caryophila* [8-29]. The phylogenetic affiliation of '*Candidatus* Görtzia infectiva' [14], already suggested that group I and group II *Holosporas* should not be joined together in a single genus. Furthermore, the phylogenetic placement of *'Candidatus* Görtzia infectiva', together with the inclusion of *H. caryophila* in the genus *Holospora*, would confer to the genus itself a paraphyletic character. In our analysis, *H. caryophila* falls basally as a fairly divergent sister species to the cluster of "classical *Holosporas*" and its closely associated *P. jennings*i endosymbiont, revealing itself as the most "ancient" species of the family *Holosporaceae* described thus far. The phylogenetic placement of *H. caryophila* has increased our understanding of the evolutionary relationships within the HLBs clade, confirming at the same time the monophyly of the taxon. Thanks to the accumulating phylogenetic evidence, it is now unequivocally clear that the ability of the IFs to assemble in the connecting piece is an apomorphy of the genus *Holospora "sensu stricto"* (*see* Appendix, 10.7). It is noteworthy that if *H. caryophila* had fallen as a sister species of '*Candidatus* Gortzia infectiva', it could have been equally likely to either lose or acquire the CP "induction capability" from the last common ancestor (LCA) of the "ancestral group" and the "classical *Holosporas*". Conversely, this result fits the evolutionary model of the IFs dissemination proposed by Fokin and colleagues (see Introduction, [13]), confirming on a molecular basis the derived nature of the character. The description of the genus *Holospora* should therefore differ from that of the family *Holosporaceae*, in that all the entitled members of the genus share this advanced trait. The 16S rRNA gene comparison has further revealed that the endosymbionts infecting *P. caudatum* (*H. caryophila* SH42) and the *P. aurelia* complex, despite their identical morphology, may have undergone a certain degree of adaptation to their hosts. It is in fact, shown here, that on the basis of the "borderline" reported values of 16S rRNA gene similarity (from 98.4 % to 98.9 %), *H. caryophila* infecting different *Paramecium* morphospecies could be considered either as closely related species or fairly differentiated strains depending on how conservative the adopted criterion for species discrimination is (see Results, 5.2). Furthermore, the characterization of the Internal Transcribed Spacer (ITS) and a fragment of the 23S rRNA gene have confirmed the divergence of *H. caryophila* SH42 from all the other strains accounted for in the analysis. Therefore, although seemingly identical, our data suggest that *H. caryophila* strains infecting different *Paramecium* morphospecies, may not be genetically alike. Despite the purification of nuclear symbionts DNA is technically challenging, DNA-DNA hybridization analysis or whole genome sequencing should be mandatory for assessing their genetic uniqueness.

In conclusion, once other group II HLBs (e.g. *H. bacillata* and *H. curvata*) will be characterized, a deeper understanding on the phylogeny of the family *Holosporaceae* and on the evolution of the horizontal transmission mechanism in this remarkable group of infective symbionts will be achieved.

7. Acknowledgments

I am indebted to Martina Schrallhammer for the accurate review of the manuscript and the devoted supervision during the entire course of the experimental activity. To Hans Dieter Görtz, Giulio Petroni, and the Erasmus program through the University of Pisa, I give my heartfelt thanks for allowing me to take 6 months traineeship at the department of Zoology at Stuttgart University. I am grateful to Hannah Moss for her efforts to facilitate the writing of the document in English. Sergei I. Fokin is acknowledged for providing a copy of his review on endocytobionts. I wish to thank Claudia Vannini and Giovanni Cercignani for valuable comments during the thesis evaluation. Finally, I would like to thank Stefano Galati for the sampling of *Paramecium biaurelia* FGC3 infected with *Holospora caryophila*.

8. References

[1] Fujishima M (ed) (2009) Endosymbionts in *Paramecium*. Microbiology Monograph, Münster

[2] Görtz HD, Fokin SI (2009) Diversity of endosymbiotic bacteria in *Paramecium*. In: Fujishima M (ed) Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 131–160

[3] Fokin SI, Görtz HD (2009) Diversity of *Holospora* bacteria in *Paramecium* and their characterization. In: Fujishima M (ed) Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 161–199

[4] Görtz HD (1988) Endocytobiosis. In: Görtz HD (ed) *Paramecium*. Springer-Verlag Berlin Heidelberg, pp 393-405

[5] Schweikert M, Fujishima M, Görtz HD (2013) Symbiotic association between ciliates and prokaryotes. In: Rosenberg et al. (eds) The prokaryotes – prokaryotic biology and symbiotic association, vol 1, 3rd edn. Springer-Verlag Berlin, Heidelberg, pp 427–463

[6] Fokin SI (2004) Bacterial endocytobionts of Ciliophora and their interactions with the host cell. Int Rev Cytol 236:181–249

[7] Hafkine MW (1890) Maladies infectieuses des paramécies. Ann Inst Pasteur, Paris 4:363–379

[8] Preer LB (1969) Alpha, an infectious macronuclear symbiont of *Paramecium aurelia*. J Protozool 16:570–578

[9] Ossipov DV, Skoblo II, Borchsenius ON, Rautian MS (1980) *Holospora acuminata* – a new species of symbiotic bacterium from the micronucleus of the ciliate *Paramecium bursaria* Foske. Cytologia 22:922-929. In Russian with English summary

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

[11] Fokin SI, (1991) *Holospora recta* sp. nov., a micronucleus-specific endobiont of the ciliate *Paramecium caudatum*. Cytologia 33: 135-141. In Russian with English summary

[12] Fokin SI, Sabaneyeva E, (1993) Bacterial endocytobionts of the ciliate *Paramecium calkinsi*. Europ J Protistol 29:390-395

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

[14] Boscaro V, Fokin SI, Schrallhammer M, Schweikert M, Petroni G (2012) Revised systematics of *Holospora*-like bacteria and characterization of '*Candidatus* Gortzia infectiva', a novel macronuclear symbiont of *Paramecium jenningsi*. Microb Ecol 65:255–267

[15] Gromov BV, Ossipov DV (1981) *Holospora* (ex Hafkine 1890) nom. rev., a genus of bacteria inhabiting the nuclei of paramecia. Int J Syst Bacteriol 31:348–352

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

[17] Görtz HD, Schmidt HJ (2005) *Holosporaceae* fam. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds) Bergey's manual of systematic bacteriology, 2nd edn, vol II, part C. Springer, New York, pp 146–160

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

[19] Beale GH, Preer JR (2008) Symbionts and mitochondria of *Paramecium*. In: Beale GH, Preer JR *Paramecium* genetics and epigenetics. Taylor & Francis Group, CRC press, pp 63–98

[20] Barhey K, Gibson I (1984) A study on the conditions for infection of *Holospora caryophila*, a macronuclear symbiont of *Paramecium biaurelia*. Micro Micros Acta 15:261-268

[21] Gibson I, Bedingfield G, Horne RW, Bird B (1986) Electron microscope study of the structure of *Holospora caryophila*. Micro Micros Acta 17:247-257

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

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

[24] Rautian MS, Wackerow-Kouzova ND (2013) Phylogenetic placement of two previously described intranuclear bacteria from the ciliate *Paramecium bursaria* (*Protozoa*, *Ciliophora*): '*Holospora acuminata*' and '*Holospora curviuscola*'. Int J Syst Evol Microbiol 63:1930–1933

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

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

[27] Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

[28] Görtz HD (1987) Different endocytobionts simultaneously colonizing ciliate cells. Ann N Y Acad Sci 503:561-568

[29] Beale GH, Jurand A, JR Preer (1969) The classes of endosymbiont in *Paramecium Aurelia*. J Cell Sci 5:65-91

[30] Vannini C, Rosati G, Verni F, Petroni G (2004) Identification of the bacterial endosymbionts of the marine ciliate *Euplotes magnicirratus* (Ciliophora, Hypotrichia) and proposal of ‘‘*Candidatus* Devosia euplotis’’. Int J Syst Evol Microbiol 54:1151–1156

[31] Schrallhammer M, Fokin SI, Schleifer KH, Petroni G (2006) Molecular characterization of the obligate endosymbiont “*Caedibacter macronucleorum*” Fokin and Görtz, 1993 and of its host *Paramecium duboscqui* strain Ku4-8. J Eukaryot Microbiol 53:499–506

[32] Ludwig W, Strunk O, Westram R, Richter L, Meier H, Kumar Y, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32:1363–1371

[33] Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) “Touchdown” PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008

[34] D’Aquila RT, Bechtel LJ, Videler JA, Eron JJ, Gorczyca P, Kaplan JC (1991) Maximizing sensitivity and specificity of PCR by pre-amplification heating. Nucleic Acids Res 19:3749

[35] Mullis KB (1991) The polymerase chain reaction in an anemic mode: how to avoid cold oligodeoxyribonuclear fusion. PCR Metodhs Appl 1:1-4

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

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

[38] Yilmaz LS, Ökten HE, Noguera DR (2006) Making All Parts of the 16S rRNA of Escherichia coli Accessible In Situ to Single DNA Oligonucleotides. Appl Environ Microbiol 72:733-744

[39] Behrens S, Rühland C, Inacio J, Huber H, Fonseca A, Spencer-Martins I, Fuchs BM, Amann RI (2003) *In situ* accessibility of small-subunit rRNA of members of the domains Bacteria, Archaea, and Eucarya to Cy3-labeled oligonucleotide probes. Appl Environ Microbiol 69:1748–1758

[40] Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR (1996) The Ribosomal Database Project (RDP). Nucleic Acids Res 24:82-85

[41] Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam- Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141–D145

[42] Nazarenko I, Pires R, Lowe B, Obaidy M, Rashtchian A (2002) Effect of primary and secondary structure of oligodeoxyribonucleotides on the fluorescent properties of conjugated dyes. Nucleic Acids Res 30:2089-2195

[43] Pruesse E, Peplies J, Glöckner FO (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823-1829

[44] Pruesse E, Quast C, Knittel K, Fuchs BM, 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 Res 35:7188–7196

[45] 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

[46] Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256

[47] Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

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

[49] Felsenstein J (1989) PHYLIP-phylogeny inference package (version 3.2). Cladistics 5:164–166

[50] Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791

[51] Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, Neumaier J, Bachleitner M, Schleifer KH (1998) Bacterial phylogeny based on comparative sequence analysis. Electrophoresis 19:554–568

[52] Whelan S, Liò P, Goldman N (2001) Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet 17:262-272

[53] Yang Z, Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: Markov chain Monte Carlo methods. Mol Biol Evol 14:717–724

[54] Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127-128

[55] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

[56] Woodsmall RM, Benson DA (1993) Information resources at the National Center for Biotechnology Information. Bull Med Libr Assoc 81:282-284

[57] Jermiin LS, Ho SYW, Ababneh F, Robinson J, Larkum AWD (2004) The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst Biol 53:638-643

[58] Brenner DJ, Staley JT, Krieg NR Classification of prokaryotic organisms and the concept of bacterial speciation. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds) Bergey's manual of systematic bacteriology, 2nd edn, vol II, part A. Springer, New York, pp 29

[59] Stackebrandt E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiology Today 33: 152–155

[60] Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading, Mass.

[61] Emelyanov VV (2001) Rickettsiaceae, *Rickettsia-like* endosymbionts, and the origin of mitochondria. Bioscience Reports 21:1-17

9. Bibliography

Wichterman R (1953) The biology of *Paramecium*. McGraw-Hill, New York

Görtz HD (ed) (1988) *Paramecium*. Springer-Verlag Berlin Heidelberg

Hausmann K, Bradbury PS (ed) (1996) Ciliates. Cells as organisms. Fisher, Stuttgart

Garrity GM, Brenner DJ, Krieg NR, Staley JT (ed) (2005) Bergey's manual of systematic bacteriology, 2nd edn, vol II. The Proteobacteria, Part A. Introductory Essay. 2nd ed, Springer

Beale GH, Preer JR (2008) PARAMECIUM genetics and epigenetics. Taylor & Francis Group, CRC press

Fujishima M (ed) (2009) Endosymbionts in *Paramecium*. Microbiology Monograph, Münster

Schweikert M, Fujishima M, Görtz HD (2013) Symbiotic association between ciliates and prokaryotes. In: Rosenberg et al. (ed) The prokaryotes – prokaryotic biology and symbiotic association, vol 1, 3rd edn. Springer-Verlag Berlin, Heidelberg, pp 427–463

10. Appendix

10.1 *Preeria caryophila* (Preer and Preer 1982) comb. nov.

Basis of the transfer of *Holospora caryophila* to the new genus *Preeria* gen. nov. (genus novum):

H. caryophila is not labeled by the *Holospora* "genus-specific" probe H16-23a in FISH experiments [13]. The 16S rRNA similarity values between the type strain 562 and the members of the genus *Holospora* are far below 95 % (see similarity matrix of 16S rRNA; Results, 5.2).

Diagnosis of the taxon: endosymbiont found in the macronucleus (MA) of several *Paramecium* species: *P. caudatum*, *P. biaurelia*, *P. octaurelia*, *P. novaurelia* [13] and *P. dodecurelia*. Reproductive forms (RFs) are rod or crescent-shaped, 0.3–0.5 x 1.0–3.0 µm. Infectious forms (IFs) are spiral, 0.2–0.3 x 5–6 µm, with both ends tapered [17]. *H. caryophila* is unable to assemble the IFs in the connecting piece of the dividing nuclei during the host division [13]. Its host spectrum is wider compairing with group I *Holosporas*.

Type species of the genus (single member of the proposed genus):

Preeria caryophila (Preer and Preer 1982) comb. nov. (combinatio nova).

Etimology of the new combination: *Preeria caryophila* (*Pree'ri.a ca.ry.o'phi.la*; N.L. fem. n. *Preeria* in honour of Dr. Louise Bertha Preer and Professor emeritus John R. Preer;. G. n. *caryum* nut, kernel, (in biology, nucleus); N.L. fem. adj. *phila* (from Gr. fem. adj. *philê*), friend, loving; N.L. fem. adj. *caryophila*, nucleus loving).

16S rRNA gene for the type strain (*P. caryophila* 562): characterized in this study.

Type strain: carried in strain 562 of *Paramecium biaurelia* (= ATCC 30694)*.

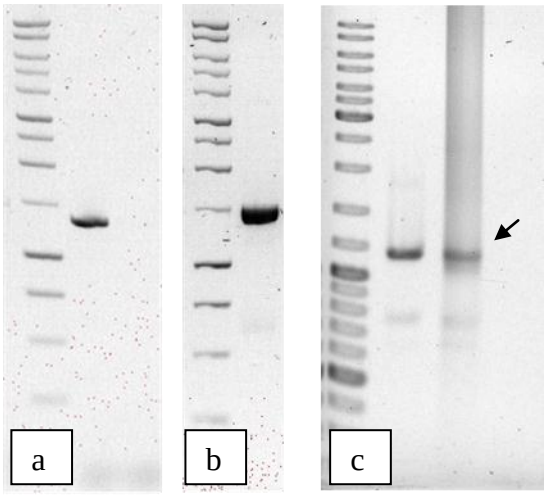
Basonym of *Preeria caryophila*: *Holospora caryophila* (ex Preer et al. 1974) Preer and Preer 1982, nom. rev., comb. nov.**

* The strain ATCC 30694 is no longer available in the ATCC.

**In violation of Rule 34a, Preer and Preer 1982 revived *Holospora caryophila* as a nomen novum.

10.2 PCR products

The PCR products of the 16S rDNA and the ITS-(5'23 S rDNA) are reported below for the type strain of *Holospora caryophila* (562). Details on amplification reactions are reported in the Matherials and Methods (4.6).



(a). PCR product of ~ 1.4 kb amplified using primers 16S F114 HoloCaedi and 16S R1488 Holo.

(b). PCR product of ~ 1.5 kb amplified using primers 16S F7 and 16S R1522a.

(c). Semi-Nested PCR product of ~ 0.9 kb (arrow) amplified using the primers 16S F1256 NovHolo and 23S R457WL. DNA sample (template) was a purified PCR products (pPCR) from a previous amplification reaction performed with primers 16S F1142 23S R457 WL992r.

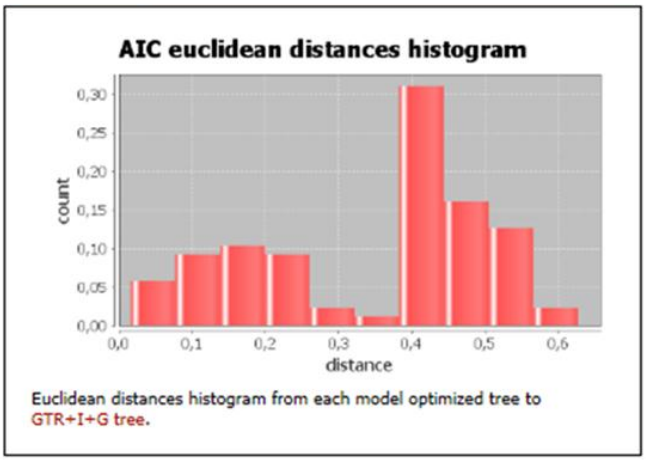
10.3 Nucleotide substitution model

Model selected

Model	GTR+I+G		
partition	012345		
-lnL	15173.4347		
K	102		
freqA	0.2651	**R(a)**	0.9821
freqC	0.1943	**R(b)**	4.3673
freqG	0.2820	**R(c)**	1.8473
freqT	0.2586	**R(d)**	0.7805
ti/tv	-	**R(e)**	5.7726
		R(f)	1.0000
p-inv	0.3520	**gamma**	0.6460

-lnL:	negative log likelihod
K:	number of estimated parameters
AIC:	Akaike Information Criterion
delta:	AIC difference
weight:	AIC weight
cumWeight:	cumulative AIC weight

	AIC	BIC	AICc	DT
RANK	0	12	-	12
Weight	0.9991	0.4540	-	0.0684
RF	0	6	-	6
AVG Distance	8.3152e-03	1.2669e-01	-	1.2669e-01
Distance VAR	6.9142e-05	4.8136e-07	-	4.8136e-07

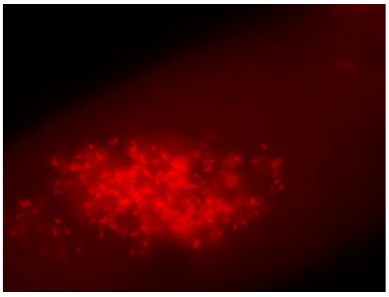


Output of jModelTest v. 2.1.4. Statistical selection of the best-fitting nucleotide substitution model (over 1624 models) using three different strategies. The AIC (Akaike Information Criterion) selected the GTR (General Time Reversible) + I (Invariable sites) + G (Rate variation among sites) model. On the upper left: parameters of the selected nucleotide substitution model. Lower left: legend. Upper right: ranking of the GTR + I + G model depending on the selection strategy used. Lower right: absolute model fit; Euclidean distances from the unconstrained multinomial model.

Detailed documentation on jModelTest can be found at http://code.google.com/p/jmodeltest2/

10.4 "Full Cycle rRNA" approach

This strategy was originally adopted for the characterization of *H. obtusa* [22], the first endosymbiont of *Paramecium* for which the phylogenetic affiliation was determined and is referred to be the "Full-Cycle rRNA analysis" [27]



In situ detection of the endosymbionts within their host cell (**FISH**)

Host cell culture → Detection of the endosymbiont

↓

Nucleic acid extraction

↓

16S rDNA amplification (PCR)

↓

Direct sequencing of the PCR products

← **Species specific probe design**

NovHolo1257 cy3

← **Direct match between EB and retrieved sequence**

↓

Comparative 16S rDNA sequence analysis

The "rRNA approach" adopted in this study. Note that the amplification products have not been cloned but directly sequenced instead. This speeded up the analyses and allowed the investigation of several strains of endosymbionts simultaneously.

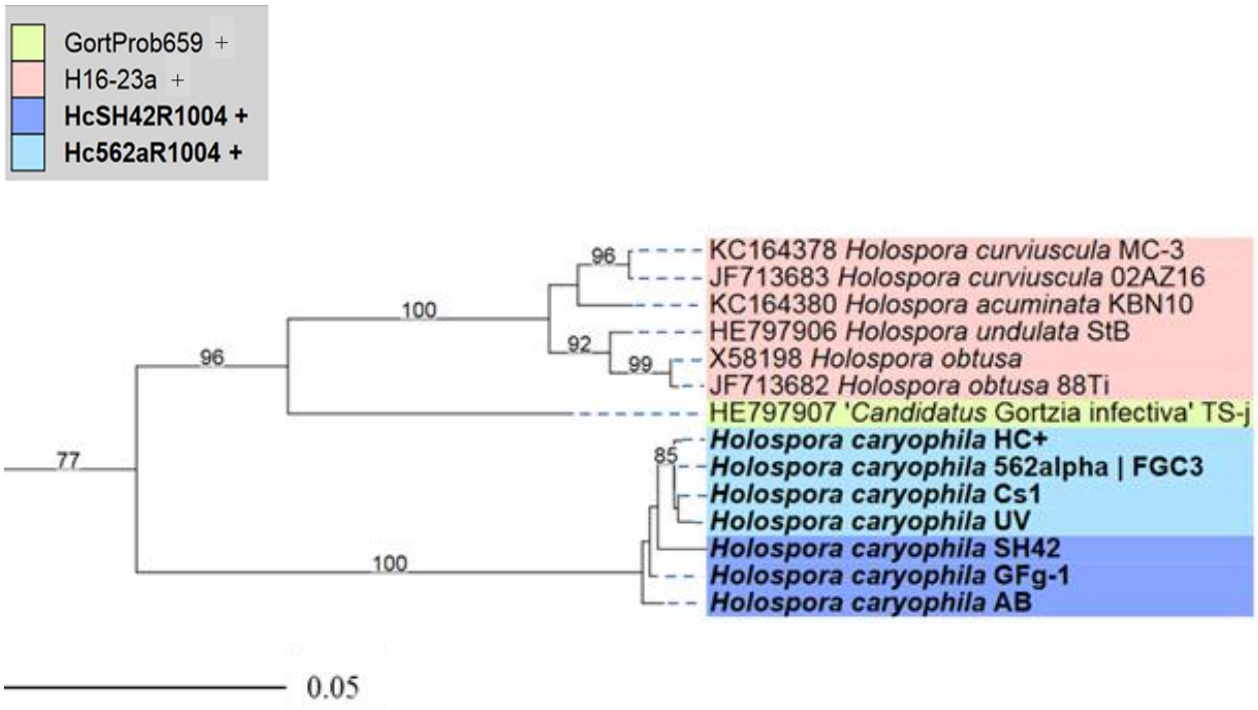
10.5 The pairwise distances plot

The graph is a "box and whiskers" plot build on "divergence values" expressed in percent equal to $100 - \mathrm{SV}\%$, where SV are similarity values between DNA sequences of *H. caryophila* strains. Next to each box the actual data distribution (dot plot) is shown for clarity. The plot is intended to "catch" putative outliers; in other words is a simple computation of the DNA similarity values listed in the similarity matrices to display the most diverging sequences within the cluster of investigated *H. caryophila* strains.

The average distance of the i^{th} strain from the others was computed as follows: $\boldsymbol{d(av)_i}=(\sum_{j=1}^{n} d_{i,i+j})/n$ where $d_{i,i+j}$ is the j^{th} pairwise comparison of the strain i with the others, and *n* is the number of comparisons (number of strains − 1). In simpler words, each value represents the average divergence of one strain from the investigated group. Calculated distances were divided by the median $\boldsymbol{D(m)}$ of each distribution, so that $\boldsymbol{d(av)_i}$ values close to the median fall around 1. $\boldsymbol{D(m)}$ was equal to 0.83, 1.75 and 1.14 for the 16S rRNA , ITS and 5'23S rRNA respectively. After normalization, any value falling above 1 is more divergent than the average relative divergence. Conversely, values below 1 are less divergent than the average. In conclusion, values falling outside the "upper whisker" are noteworthy and could represent endosymbionts less closely related to the investigated group, as discussed in the text.

10.6 FISH response within the family *Holosporaceae*

FISH response of the HLBs to probes NovHolo1257, HcSH42R1004, Hc562R1004, H16-23a [22] and GortProb659 [14], is summarized below.

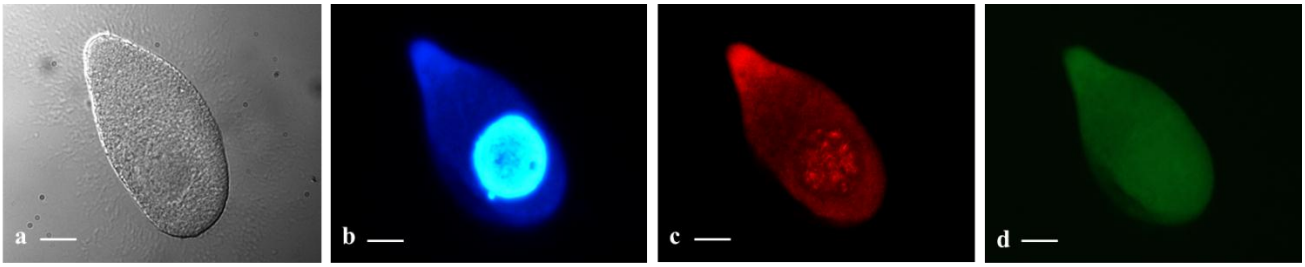


FISH-outcome of the *Holosporaceae* to probes acting at different taxonomic levels. H16-23a: specific for the genus *Holospora*. NovHolo1257: specific for *H. caryophila* (light and dark blue). HcSH42R1004 and Hc562R1004: *H. caryophila* strain-specific probes. GortProb659: specific for "*Candidatus* Gortzia infectiva".

H. caryophila was shown to test negative [6] to the *Alphaproteobacteria*-specific probe ALF1b [36]. Other authors, using different FISH protocols, reported the labeling of *H. caryophila* using ALF1b [6]. This probe hybridizes at the very beginning of the 16S rRNA marker with a stretch of 17 nt. Over the 8 nt covered in our sequences, we report one mismatch.

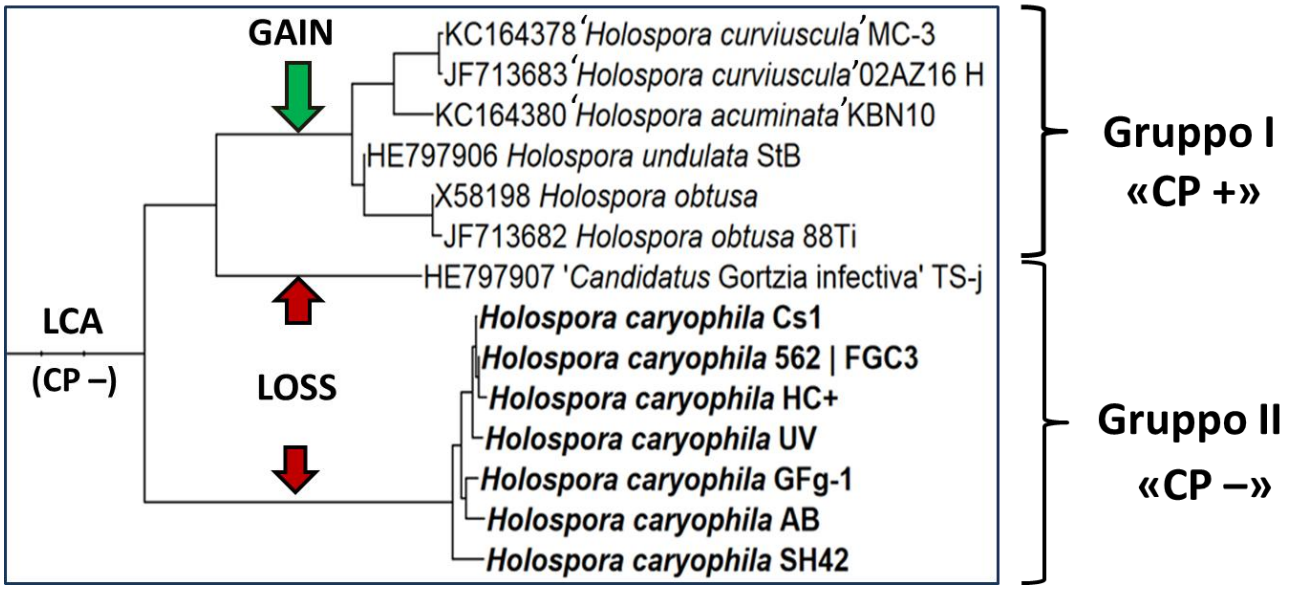
Note. Several FISH outcomes are inferred from DNA sequences and have not been tested experimentally.

10.7 FISH: Additional micrographs



Paramecium biaurelia FGC3 infected with *Holospora caryophila*. From left to right: DIC, DAPI, Hc562R1004-cy3, HcSH42R1004-FITC. Bars, 20 μm

10.8 Evolutionary pattern of the "connecting piece" trait



During the host nuclear division *H. caryophila* and '*Candidatus* Görtzia infectiva' are both unable to assemble the infectious forms (IFs) in the connecting piece («CP −»). The more parsimonious reconstruction is that in which the last common ancestor (LCA) was «CP −» and the trait had been acquired (green arrow) in the branch underling the group I clade («CP+»). The alternative scenario in which this trait would have been lost twice (red arrows) from a «CP+» common ancestor, and retained only in the group I clade is improbable. The CP "induction capability" is therefore likely to be an apomorphy of the genus *Holospora* "sensu stricto".