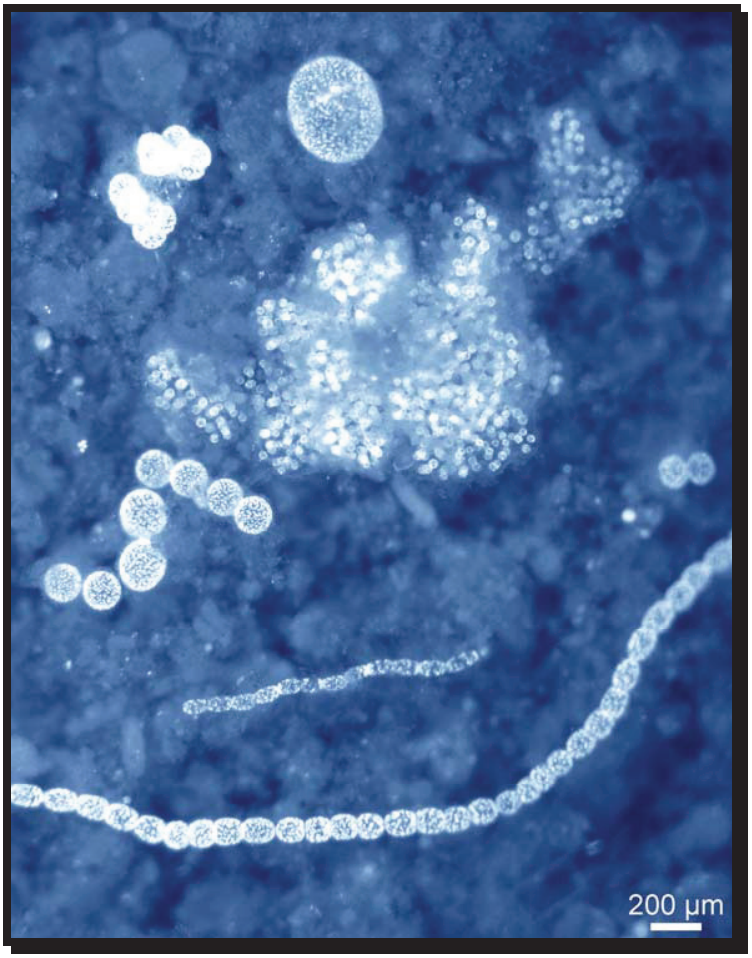
Diversity studies and molecular analyses with single cells and filaments of large, colorless sulfur bacteria



Verena Salman

Cover: Population of large sulfur bacteria in Namibian sediments.

ISSN 0723-2020

Diversity studies and molecular analyses with single cells and filaments of large, colorless sulfur bacteria

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
- Dr. rer. nat. -

dem Fachbereich Biologie / Chemie
der Universität Bremen
vorgelegt von

Verena Salman

Bremen, April 2011

Die vorliegende Doktorarbeit wurde in der Zeit von Januar 2008 bis April 2011 am Max-Planck-Institut für Marine Mikrobiologie in Bremen angefertigt.

Erste Gutachterin: Dr. Heide Schulz-Vogt
Zweiter Gutachter: Prof. Dr. Rudolf Amann

Tag des Promotionskolloquiums: 06. Juni 2011

Table of Contents

Summary

Large sulfur bacteria feature conspicuous morphologies that are usually visible with the naked eye. Most representatives were already described in the 19th and early 20th century and it need nearly another 100 years, until a new morphotype of large sulfur bacteria was discovered. This discovery initiated a perspective for revealing the presence of yet unknown types in this group of bacteria and indeed led to the finding of a new type in two marine seep settings. This novel morphotype is presented in **Chapter 2** and is the first non-filamentous members in the family *Beggiatoaceae* that shows a dimorphic life cycle, exhibiting alternation between sessile and free-living forms. In **Chapter 3**, another three novel morphotypes are presented, which were discovered in Namibian sediments.

The detection of these novel morphotypes of large sulfur bacteria led to the necessity for an improved phylogenetic classification of the entire group. Earlier attempts to sequence the 16S rRNA genes of the large sulfur bacteria resulted in only few nearly full-length sequences, whereas some genera were even represented by only partial sequences. Accordingly, differentiation of genera in this group is still based on morphological features. Now, **Chapter 3** presents the sequencing of 16S rRNA genes and ITS regions of more than 100 individual cells and filaments of large sulfur bacteria, revealing a major insight into the phylogeny of these extraordinary bacteria. It is demonstrated that the traditional, morphology-based classification does not correlate with the phylogeny derived from 16S rRNA gene sequences. Consequently, a reclassification of the entire family is proposed, being completely independent from former morphological categories.

Sequencing single cells and filaments of large sulfur bacteria furthermore revealed that they represent the first group of bacteria, which commonly contain large and numerous introns in their 16S rRNA genes. In **Chapter 4**, it is demonstrated that the introns are removed efficiently from the rRNA precursor, thereby ligating the two exons, and are not present in the native ribosome. Furthermore, it was experimentally verified that a commonly applied PCR approach introduces a length heterogeneity bias and systematically discriminates against enlarged genes and favors the amplication of shorter homologues. Consequently, this fact questions the universality of 16S rRNA-based clone libraries for diversity studies. At least the group of large sulfur bacteria is systematically discriminated against in such universal PCR approaches and it can be assumed that this PCR bias also affects a yet unknown amount of other microorganisms.

Zusammenfassung

Die großen Schwefelbakterien besitzen sehr auffällige Morphologien und können meist bereits mit bloßem Auge erkannt werden. Die meisten Vertreter dieser Gruppe wurden schon im 19. und Anfang des 20. Jahrhunderts beschrieben und es dauerte fast einhundert Jahre bis ein neuer Morphotyp von großen Schwefelbakterien entdeckt wurde. Diese Entdeckung eröffnete die Aussicht darauf, noch weitere, bisher unbekannte, Arten von großen Schwefelbakterien zu finden und tatsächlich wurde kürzlich ein neuer bakterieller Typ an zwei marinen Unterwasserquellen gefunden. Dieser neue Morphotyp wird in **Kapitel 2** beschrieben und ist der erste nicht-filamentöse Vertreter aus der Familie *Beggioatoaceae*, der einen dimorphen Lebenszyklus ausprägt und zwischen einem festsitzenden und einem frei lebenden Stadium alterniert. In **Kapitel 3** werden außerdem noch drei weitere neue Morphotypen beschrieben, die in namibischen Sedimenten gefunden wurden.

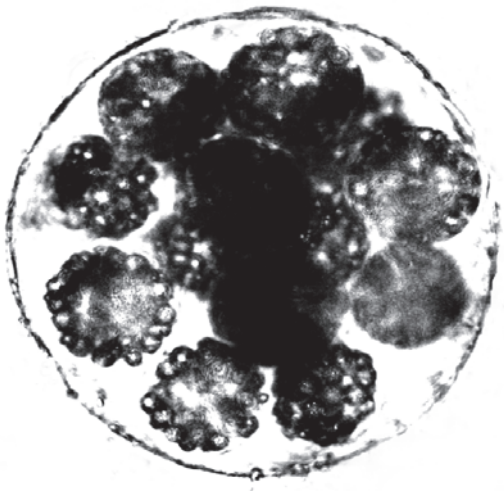
Die Entdeckung dieser neuen Morphotypen der großen Schwefelbakterien hatte zur Folge, dass nun eine verbesserte phylogenetische Klassifizierung der gesamten Gruppe benötigt wurde. Trotz vieler früherer Versuche, die 16S rRNA Gene der großen Schwefelbakterien zu sequenzieren, sind bis heute nur sehr wenige Vollängen-Sequenzen verfügbar, beziehungsweise von manchen Vertretern sind nur kurze, partielle Sequenzen vorhanden. Demnach basiert heute die Unterscheidung der Gattungen immer noch auf den morphologischen Charakteristika. In **Kapitel 3** wird das Sequenzieren von 16S rRNA Genen und ITS Regionen von über einhundert Einzelzellen und Einzelfilamenten der Gruppe von großen Schwefelbakterien präsentiert und zum ersten Mal kann eine tiefe Einsicht in die Abstammungsgeschichte dieser großartigen Bakterien erlangt werden. Es zeigt sich, dass die traditionelle, auf Morphologien basierende Klassifizierung, nicht der neuen, von 16S rRNA Sequenzen abgeleiteten Abstammung entspricht. Demzufolge wird eine Reklassifizierung der gesamten Familie vorgeschlagen, die nun gänzlich unabhängig von morphologischen Leitlinien besteht.

Das Sequenzieren von Einzelzellen und Einzelfilamenten der großen Schwefelbakterien zeigte außerdem, dass dies die erste Gruppe von Bakterien ist, die häufig große und zahlreiche Introns in ihren 16S rRNA Genen besitzt. In **Kapitel 4** konnte gezeigt werden, dass die Introns effizient von dem rRNA Vorläufermolekül entfernt werden,

wobei die zwei entstehenden Exons ligiert werden, und die Introns nicht Teil des nativen Ribosoms sind. Außerdem wurde experimentell bewiesen, dass ein allgemein gebräuchlicher PCR Ansatz einen auf Längenheterogenität begründeten Fehler einbringt und demnach längere Gene gegenüber ihren kürzeren Äquivalenten systematisch diskriminiert werden. Als Konsequenz daraus wird die Allgemeingültigkeit einer auf 16S rRNA Sequenzen basierenden Klonbank für Diversitätsstudien in Frage gestellt. Zumindest die Gruppe der großen Schwefelbakterien wird in solchen universellen PCR Ansätzen systematisch diskriminiert und es kann angenommen werden, dass sich dieser durch die PCR eingebrachte Fehler auch auf eine bisher unbekannte Anzahl von anderen Mikroorganismen auswirkt.

Chapter 1

General Introduction



Discoveries of large sulfur bacteria through the centuries

Central to this work is the group of large, colorless sulfur bacteria. This group comprises organisms, in which already the individual cell or the individual filament can be seen with the naked eye. This implies that at least one dimension of the individual cell or filament must be extraordinarily large in terms of microbiological scales. They are referred to as 'colorless', which distinguishes them from pigmented (red or green), sulfur-containing organisms that perform photosynthesis. Already Winogradsky introduced the term 'sulfur bacteria', referring to bacteria, which have a metabolism that in some ways involves sulfur (Winogradsky, 1887). The large, colorless sulfur bacteria can furthermore contain voluminous internal vacuoles (Hinze, 1903; Maier & Gallardo, 1984a; Nelson *et al.*, 1989; Schulz *et al.*, 1999), due to which they are sometimes also referred to as 'vacuolated, sulfide-oxidizing bacteria' (VSO, Kalanetra *et al.*, 2004; Kalanetra *et al.*, 2005). In the following, an overview over the history of these extraordinary bacteria is given, telling the story from the early discoveries on and reaching to the current state of knowledge obtained with modern analytical methods.

Early observations on *Beggiatoa* by Winogradsky

Sergei Nicholaevitch Winogradsky (1856−1953, Fig. 1A), a Russian microbiologist and ecologist, was the pioneer in developing the concept of chemosynthetic life. His work was initiated by the discovery of the first sulfur bacterium, *Beggiatoa alba* (renamed by Trevisan, 1842; first identified as *Oscillatoria alba*, Vaucher, 1803), which was a conspicuous organism forming long multicellular filaments visible to the naked eye and containing numerous sulfur inclusions (Trevisan, 1842). From this storage compound, Winogradsky deduced that also the physiology of these organisms is putatively aberrant from the known concepts of organotrophism and phototrophism. Subsequent to key experiments with enrichments of *Beggiatoa*-containing sediment and observations of isolated filaments in slide experiments, he concluded the concept of lithotrophism (Winogradsky, 1890; Winogradsky, 1887). In this physiological pathway, energy is gained by the oxidation of reduced inorganic compounds, like for example hydrogen sulfide, ammonium or iron.

Before Winogradsky published his revolutionary hypothesis, *Beggiatoa* and other sulfur-containing bacteria were considered to reduce sulfate to sulfide because the populated habitats usually showed concomitant increased amounts of sulfide (Cohn, 1875). Indeed, *Beggiatoa* preferentially inhabited sulfur springs and other habitats with

elevated sulfide concentrations, where they formed extensive, white mats. However, *Beggiatoa* died when oxygen was entirely removed (Hoppe-Seyler, 1886), which contradicted the theory of a sulfate-reducing respiration to sulfide.

Finally, it was Winogradsky, who revealed that the co-occurrence of reduced sulfur compounds in the environment and the sulfur globules inside the organisms were connected in a way that *Beggiatoa* were oxidizing sulfide to sulfur, which they then stored internally. In a next step, he observed the precipitation of $CaSO_4$ in the presence of $CaCO_4$, from which he concluded that *Beggiatoa* oxidized the internal sulfur to sulfate. In the absence of sulfide, the filaments rapidly lost their sulfur globules and died, so the organisms seemingly required this reduced sulfur compound for their living (Winogradsky, 1887). Winogradsky also laid the basis for the concept of gradient organisms (opposing gradients of electron donor and electron acceptor). The glass cylinders, in which Winogradsky enriched and grew *Beggiatoa* (Winogradsky, 1887), are the first described gradient cultures (sulfide from below and oxygen from above) and are nowadays known as 'Winogradsky column' (Fig. 1B).

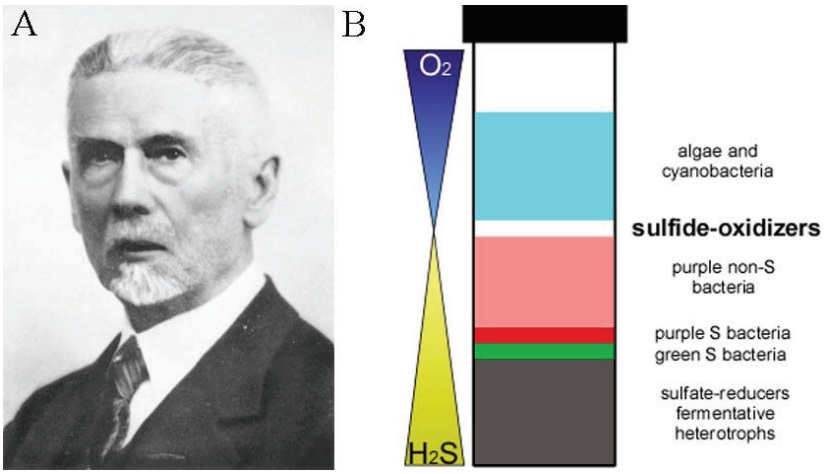


Figure 1. (A) Portrait of S. Winogradsky, who developed the concept of chemosynthesis. (Thornton, 1953) (B) Winogradsky column with opposing gradients and the enrichment of certain types of micro-organisms in different layers along the gradients.

Studying nitrifying bacteria, he furthermore revealed that lithotrophs were capable of growing with CO_2 as the only carbon source (Winogradsky, 1892), leading to the model of chemosynthesis in living organisms. This model describes the fixation of inorganic carbon compounds to form biomass coupled to the oxidation of inorganic compounds as energy source. Thus, organisms capable of this metabolism build the basis for life in those habitats, in which light is not present and photosynthesis (until then considered as the only pathway for primary production) cannot take place (Winogradsky, 1890).

Discoveries and morphological descriptions of large sulfur bacteria

Genus *Beggiatoa* (Vaucher 1803) Trevisan 1842

Species of the genus *Beggiatoa* were the earliest members of the large sulfur bacteria to be discovered. The name *Beggiatoa* was introduced by Trevisan in remembrance of the Italian physician and scientist Francesco S. Beggiato (Trevisan, 1842). The first specimens found by Beggiato and Trevisan originated from a hot spring (Terme Euganee) in Padova, Italy (Beggiato, 1833; Trevisan, 1842), but they were soon found in various freshwater and marine habitats and were thus considered as widespread organisms. During these early observations, filaments were described to always contain conspicuous sulfur inclusions, exhibit a gliding motility and vary in diameter between 1.5−55 µm (Fig. 2). Based on their filament diameters, the different species of *Beggiatoa* were distinguished, e.g. *B. alba* (2−5 µm, *alba* meaning white, Trevisan, 1842; Vaucher, 1803), *B. arachnoidea* (5–14 µm, *arachnoidea* meaning cobweb-like, Rabenhorst, 1865), *B. leptomitiformis* (1–2.5 µm, *Leptomitus* is a genus of water molds and *formis* meaning shape, Trevisan, 1842), *B. minima* (<1 µm, *minima* meaning smallest, Winogradsky, 1888), *B. gigantea* (26–55 µm, *gigantea* meaning gigantic, Hinze, 1901; Klas, 1937) and *B. mirabilis* (15–21.5 µm, *mirabilis* meaning marvelous, Cohn, 1865; Klas, 1937).

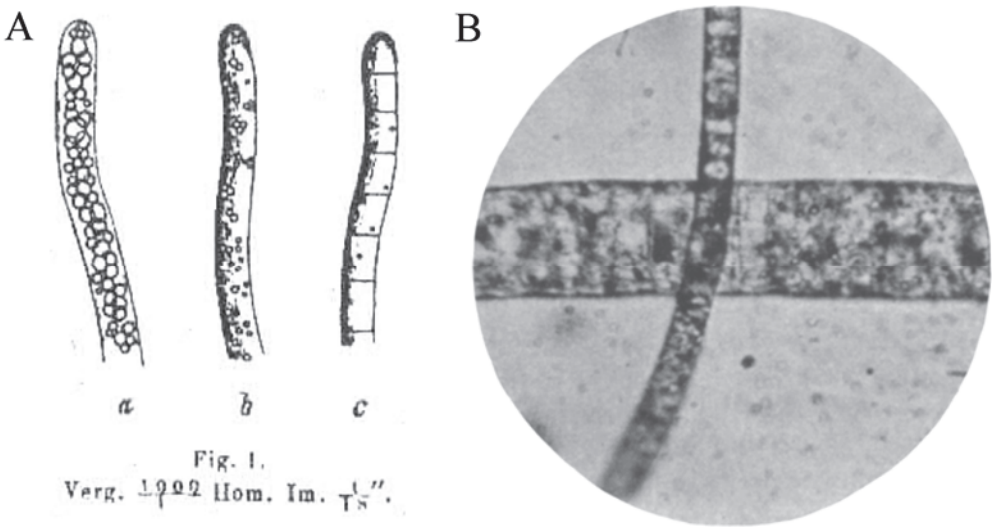


Figure 2. (A) Drawing of *Beggiatoa* filaments by Winogradsky, which he used in his famous studies published in 1887. He stated that a species definition based on filament diameter is not reliable, so he refused from specifying the organisms any further. Filament diameters were mostly 3 µm, but ranged between 1−5 µm. (Winogradsky, 1887) (B) Microscopic image of live *Beggiatoa* spp. The horizontal filament was identified as *Beggiatoa gigantea* and the vertical filament as *Beggiatoa mirabilis* by Klas. The magnification is about 500×. (Klas, 1937)

Some of the beggiatoas were cultivated in pure culture, however, only *Beggiatoa alba* remained as such, and is available today as *Beggiatoa alba* strain B15LD (Strohl & Larkin, 1978) and strain B18LD (Strohl *et al.*, 1981) from type culture collections. This lack of pure cultures of other species in the genus *Beggiatoa* is also the reason why *B. alba* is the only currently recognized species of the genus *Beggiatoa* according to the Approved Lists of Bacterial Names (Skerman *et al.*, 1980; Skerman *et al.*, 1989) and the current edition of Bergey´s Manual of Systematic Bacteriology (Strohl, 2005). Attempts for (re-)identification of *Beggiatoa* species in environmental samples based on filament diameter have shown to be difficult, as diameter ranges of observed populations overlap with traditional species categories (Jørgensen, 1977).

In the course of the 20th century, *Beggiatoa*-resembling filaments were discovered at numerous geographical sites and various enrichment and isolation experiments were performed. In the beginning of that century, diversity studies were predominantly performed, sometimes only mentioning the presence of *Beggiatoa* among many other organisms and sometimes giving a thorough description of the morphology and habitat (e.g. Hinze, 1901; Keil, 1912; Klas, 1937). Beginning in the late 1970s, also detailed physiological experiments were performed. By using mostly pure cultures, the knowledge already provided by Winogradsky was expanded, revealing e.g. the usage of thiosulfate as electron donor for the growth of a *Beggiatoa* strain (Nelson & Castenholz, 1981). Interestingly, besides lithotrophically growing strains, also strictly heterotrophically growing beggiatoas were obtained that did not store sulfur in the presence of hydrogen sulfide (Strohl & Larkin, 1978). Furthermore, N_2-fixation (Nelson *et al.*, 1982) as well as CO_2-fixation (Nelson & Jannasch, 1983) was shown for certain strains. In the absence of oxygen, reduction of nitrate (Nelson *et al.*, 1982) or internally stored sulfur (Nelson & Castenholz, 1981) was reported. Both, heterotrophically grown cultures (Strohl & Larkin, 1978), and strictly lithotrophic *Beggiatoa* spp. (Nelson & Jannasch, 1983) were usually microaerophilic. The characterization of the latter as gradient organisms was demonstrated in the environment (Jørgensen & Revsbech, 1983) and in the first marine pure culture of *Beggiatoa* (Nelson *et al.*, 1982).

Also occurring in the second half of the last century, the first bacterial mats, that were nearly exclusively composed of filamentous sulfur bacteria, were found at marine, organic-rich sites. They often spanned large areas at coastal regions of local upwelling or to surround hydrothermal vents and cold seeps (Fig. 3A, Jannasch *et al.*, 1989;

Jørgensen & Revsbech, 1983; Nelson *et al.*, 1989; Soutar & Crill, 1977; Williams & Reimers, 1983). These benthic populations usually consisted of several different sulfide-oxidizing bacteria, but major parts of the mats were often formed by single, gliding filaments that were always identified as *Beggiatoa* (Jannasch *et al.*, 1989; Nelson *et al.*, 1989; Soutar & Crill, 1977; Williams & Reimers, 1983). The habitats featured a sulfide flux emerging from below resulting from sulfate-reducing bacteria in deeper sediment layers and oxygen diffusing into the sediment from the overlying water. *Beggiatoa*, as motile gradient organisms, were considered to orient themselves in the preferred transition zone of oxygen and sulfide (Williams & Reimers, 1983).

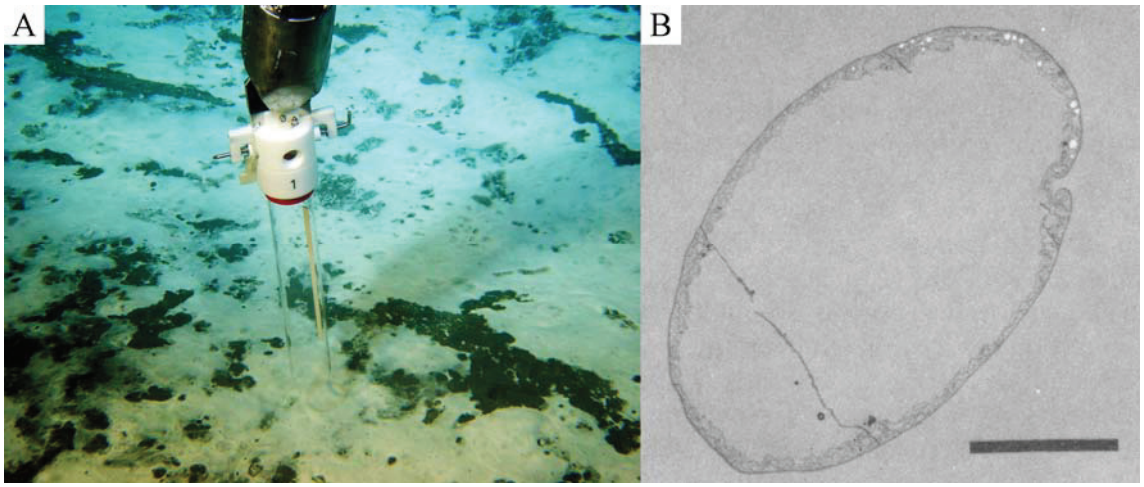


Figure 3. (A) White mats consisting of filamentous sulfur bacteria covering large areas surrounding the Haakon Mosby Mud Volcano in the Arctic deep sea. (www.mpi-bremen.de) (B) Transmission electron microscopic image of a cross-sectioned *Beggiatoa* filament. The majority of the biomass of a wide *Beggiatoa* cell mainly consists of a single intracellular vacuole. The cytoplasm forms only a thin layer at the rim of the cell. The scale bar represents 10 µm. (Nelson *et al.*, 1989)

Soon, the occurrence of extremely wide beggiatoas (116−122 µm in diameter) at Guaymas Basin vent sites in the Gulf of California was reported together with the first mentioning of large intracellular vacuoles in *Beggiatoa* (Fig. 3B, Nelson *et al.*, 1989). The importance of these vacuoles for the upkeep of diffusion into the cytoplasm was recognized (Nelson *et al.*, 1989), however, the accumulation of nitrate therein was found only seven years later (McHatton *et al.*, 1996). Detection of active dissimilatory nitrate reductase in cell lysates led to the conclusion that *Beggiatoa* can use internally stored nitrate as alternative electron acceptor (McHatton *et al.*, 1996). Nitrate is then supposedly reduced to ammonia (McHatton *et al.*, 1996; Sayama *et al.*, 2005; Schulz & Jørgensen, 2001) or denitrified (McHatton *et al.*, 1996; Mussmann *et al.*, 2007). Vacuolated *Beggiatoa* spp. were later also discovered at other marine locations

(Jørgensen *et al.*, 2010; Kalanetra *et al.*, 2004; Kojima & Fukui, 2003), as well as in brackish (Mussmann *et al.*, 2003) and hypersaline habitats (de Albuquerque *et al.*, 2010; Hinck *et al.*, 2007). Recently, it was shown that certain *Beggiatoa* strains can store polyphosphate in large amounts inside the vacuole and by releasing large amounts of phosphate into the sediment, they might have a major impact on the global phosphorous cycle (Brock, 2011).

Genus *Thiothrix* (Rabenhorst 1865) Winogradsky 1888 and genus *Leucothrix* Oerstedt 1844

One year after he published his famous work on *Beggiatoa*, Winogradsky (1888) classified a second type of filamentous sulfur bacteria. Like *Beggiatoa*, these multicellular filaments also populated hot sulfur springs and contained many sulfur inclusions, but in contrast to *Beggiatoa* these filaments were not motile (Winogradsky, 1888). In fact, they were attached with one end of the filament to either solid substrates of the habitat and/or to other filaments (Fig. 4A, Winogradsky, 1888), often forming rosettes. The organism was named *Thiothrix nivea*, *thium* meaning sulfur, *thrix* hair and *nivea* snow-white (first identified as *Beggiatoa nivea*, Rabenhorst, 1865; renamed by Winogradsky, 1888). In this genus, a dimorphic life cycle was detected. Non-motile, rosette-forming filaments were observed to produce rod-shaped gonidia from their free ends (Winogradsky, 1888). Sometimes, the entire apical region of the filament disintegrated to form gonidia (Fig. 4B, Harold & Stanier, 1955). These gonidia were motile by gliding, were subsequently attaching to a surface and grew out into a new filament. The diameter of rods and filaments was between 1−1.5 µm and lengths were 1−2.5 µm for rods and up to 200 µm for filaments (Winogradsky, 1888).

The genus *Thiothrix* is, besides few *Beggiatoa* freshwater strains, the only currently pure cultured among the large sulfur bacteria. Among the earliest isolated species are *T. nivea* (Larkin & Shinabarger, 1983) and *T. fructosivorans* (meaning fructose-eating, Howarth *et al.*, 1999), and among the latest are *T. caldifontis* (meaning from a hot spring) and *T. lacustris* (meaning from a lake, Chernousova *et al.*, 2009). *Thiothrix* spp. were isolated from sulfide-containing flowing water (Larkin & Shinabarger, 1983) or from activated sludge from waste water treatment plants (Williams & Unz, 1985). *Thiothrix* spp. were characterized as aerobic or microaerophilic, facultatively autotroph, chemoorganotroph or mixotroph (Chernousova *et al.*, 2009; Harold & Stanier, 1955; Howarth *et al.*, 1999; Larkin & Shinabarger, 1983; Williams & Unz, 1985). The only

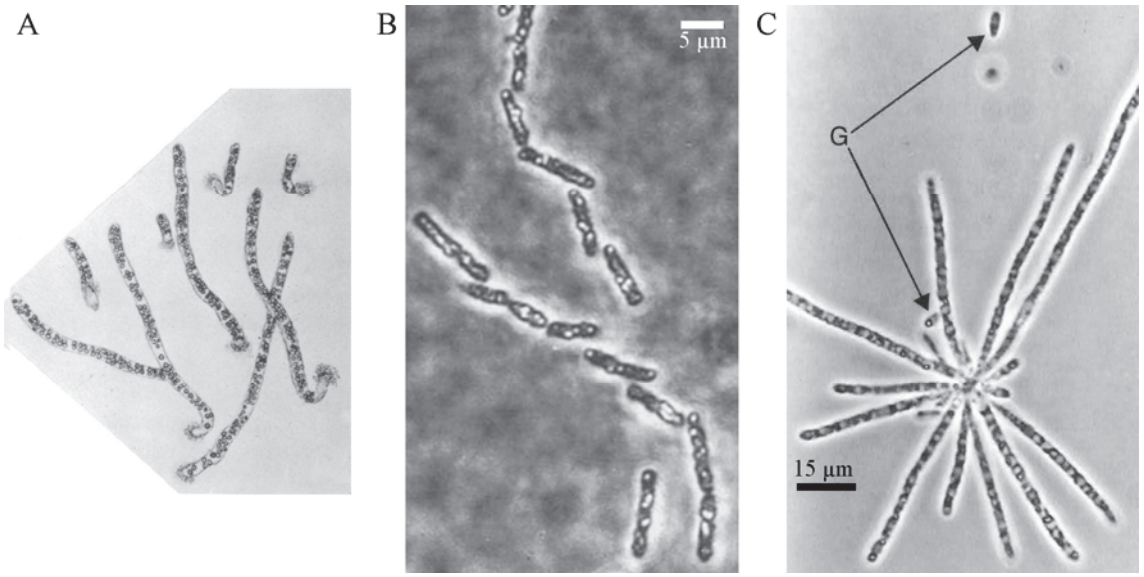


Figure 4. (A) Drawing of *Thiothrix* sp. by Winogradsky. The filaments have two poles, where one is the basal region, exhibiting a holdfast structure, and the distal end is the site of gonidia production. (Adapted from Winogradsky, 1888) (B) Microscopic image of rod-shaped fragments (gonidia) of a *Thiothrix nivea* filament. (Adapted from Bland & Staley, 1978) (C) Microscopic image of *Leucothrix mucor* forming a rosette. Single gonidia (G) are also visible. (Adapted from Bland & Brock, 2005)

lithoautotrophic species was *Thiothrix ramosa* (meaning much-branched, Odintsova *et al.*, 1993), however, it is not a recognized species, yet. In contrast to *Beggiatoa* spp., intracellular vacuoles or nitrate storage were never reported for *Thiothrix* spp. *Thiothrix nivea* is the only recognized species in the Approved Lists (Skerman *et al.*, 1980; Skerman *et al.*, 1989). However, most species were isolated after the Lists had been published, and the six additional species were recognized in the International Journal of Systematic and Evolutionary Microbiology and the current edition of Bergey´s Manual of Systematic Bacteriology (Unz & Head, 2005 and references therein).

A comparable morphotype to *Thiothrix* is *Leucothrix*, which is usually mentioned along with *Thiothrix*. However, it does not form intracellular sulfur inclusions, which actually excludes it from the group of large sulfur bacteria. Nevertheless, the genus is briefly listed at this point. *Leucothrix mucor* (meaning *leucus* clear, *thrix* hair and *mucor* is a genus of molds) is the only so far described (Oersted, 1844), isolated (Harold & Stanier, 1955) and validated (Skerman *et al.*, 1980; Skerman *et al.*, 1989) species of the genus and occurs at the littoral zones in marine environments (Harold & Stanier, 1955; Oersted, 1844). Filaments are also attached like *Thiothrix* spp. with attached ends ranging between 3−6 µm in diameter and the tips between 1.5−3 µm being somewhat tapered (Harold & Stanier, 1955; Pringsheim, 1957). Like *Thiothrix* spp., *Leucothrix mucor* forms rosettes (Fig. 4C) and produces motile gonidia (Harold & Stanier, 1955),

which often move by gliding (Harold & Stanier, 1955). *Leucothrix* is an aerobic, versatile heterotroph (Harold & Stanier, 1955; Pringsheim, 1957; Raj, 1977).

So far, *Thiothrix* spp. were considered freshwater organisms, however, in the 1980s and 90s, attached, *Thiothrix*-like filaments were discovered for the first time in microbial mats at hydrothermal vents in the deep-sea oceans (Guezennec *et al.*, 1998; Jacq *et al.*, 1989; Jannasch & Wirsen, 1981; Stein, 1984; Wirsen *et al.*, 1993) and as gut symbionts of marine invertebrates (Temara *et al.*, 1993). The sulfur-containing filaments were also attached (Fig. 5A and C) and formed rosettes on biotic and abiotic surfaces with filament diameters of a few micrometers. Like in *Thiothrix* spp., gonidia production of attached filaments and attachment of the rod-shaped gonidia was shown for these marine equivalents (Jacq *et al.*, 1989; Jannasch & Wirsen, 1981; Wirsen *et al.*, 1993). Recently, also extremely wide, attached, marine filaments were discovered (Heijs *et al.*, 2005; Kalanetra *et al.*, 2004; Kalanetra & Nelson, 2010). Like *Thiothrix* spp., they formed rosettes and gonidia (Kalanetra *et al.*, 2004; Kalanetra & Nelson, 2010). Filament diameters reached up to 112 µm and in these specimens, large intracellular vacuoles were detected (Fig. 5B), but putatively stored nitrate could not be measured. Instead, a transient storage of oxygen in the vacuoles was postulated (Kalanetra *et al.*, 2004; Kalanetra & Nelson, 2010). Besides acknowledging their resemblance to *Thiothrix* spp., the marine, attached filaments were never classified.

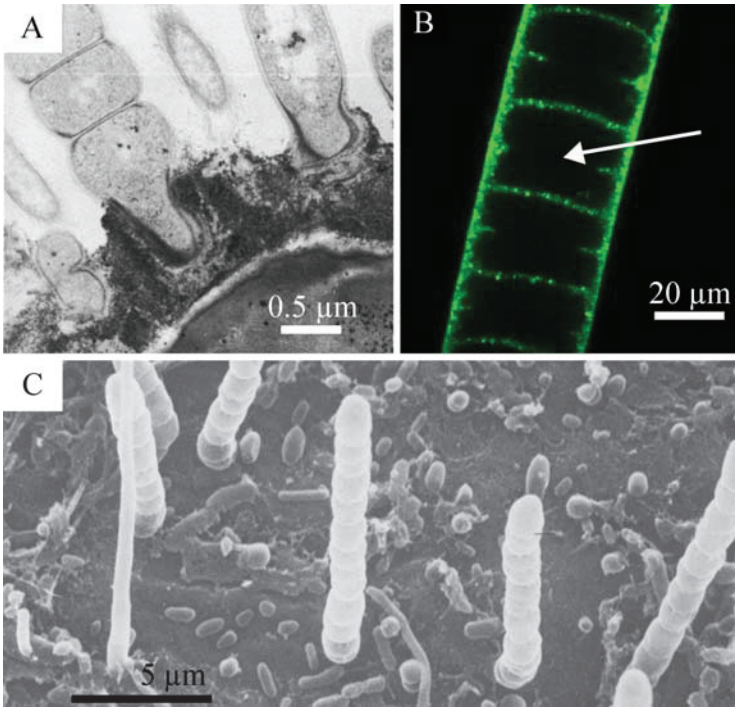


Figure 5. (A) Transmission electron microscopic image of cross-sectioned, attached filaments, anchored by a polysaccharide-type holdfast structure. (Adapted from Wirsen *et al.*, 1993) (B) Fluorescent microscopic image of a fluorescein isothiocyanate-stained attached filament, showing the stained cytoplasm as a thin rim surrounding the unstained central vacuoles. (Adapted from Kalanetra *et al.*, 2004) (C) Scanning electron microscopic image of *Thiothrix*-*Leucothrix*-like filaments from a vent-site. (Adapted from Jannasch & Wirsen, 1981)

Genus *Achromatium* Schewiakoff 1892

In the turn from the 19th to the 20th century, not only filamentous but also non-filamentous large sulfur bacteria were discovered. For the first time, large, sulfur-containing single cells of a spherical or ovoid shape were described (Fig. 6, Schewiakoff, 1892). They occurred in the sediments of turf and silt pits and a quarry pond along the river Rhine in central Germany. Besides storing sulfur, the organisms additionally contained other larger inclusions that filled the entire residual interior of the cell (Fig. 6.1.). These granules were proposed to consist of calcium oxalate. Hence, the bacteria were named *Achromatium oxaliferum*, *A*- meaning not, *chromatium* color and *oxaliferum* oxalate-containing (Schewiakoff, 1892). The diameters of the cells ranged from 9−22 µm and the ellipsoid cells, which were assumed to be subject to cell division, were up to 43 µm long. Occasionally, cells were slightly bent and had a vibroid form. They exhibited an inconspicuous, sporadic, jerky rolling motility on glass slides, sometimes twitching back and forth or rotating (Schewiakoff, 1892).

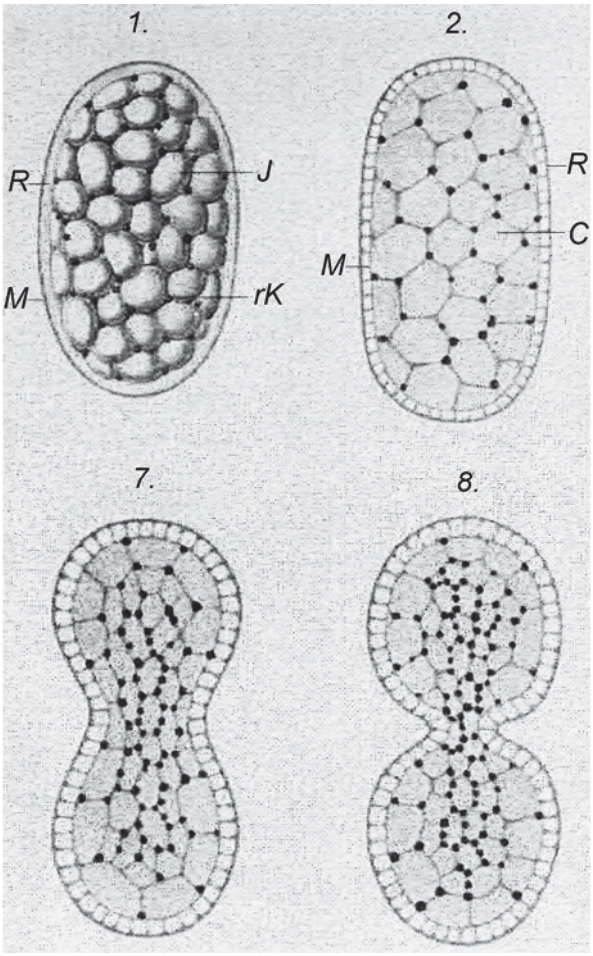


Figure 6. Drawings of *Achromatium oxaliferum* by Schewiakoff. The ovoid cells contain large inclusions (J) that are putatively calcium oxalate. They are assumed to be surrounded by a membrane and are distributed in the interior of a layer that has a network-structure like honey combs (R). This layer is supposed to lie beneath a thick membrane or sheath (M). Along the nodes and connections of the comb-like structure, numerous small inclusions were visible when stained with hematoxylin and are, therefore, assumed to be chromatin nodules (rK). Magnification is about 2200×. (Adapted from Schewiakoff, 1892)

In the beginning of the last century, spherical, sulfur and calcium oxalate-containing bacteria resembling *Achromatium* were found in many freshwater habitats, sometimes measuring up to 102 µm like *Achromatium gigas* (meaning gigantic, Nadson, 1913), but sometimes also named differently like e.g. *Hillhousia mirabilis* (named after W. Hillhouse and *mirabilis* meaning marvelous, West & Griffiths, 1909). None of the organisms was cultured, but a physiology similar to other lithotrophic sulfur bacteria was assumed, as the habitats always showed traces of hydrogen sulfide. The organisms occurred in the upper layers of the sediments, contained numerous sulfur inclusions and died when oxygen was absent (Nadson, 1913; Schewiakoff, 1892; West & Griffiths, 1909). Large liquid vacuoles were never described. Instead, the entire cell was always filled with large storage granules (Fig. 7). The initially characterized calcium oxalate granules were later corrected to consist of calcium carbonate (Bersa, 1920).

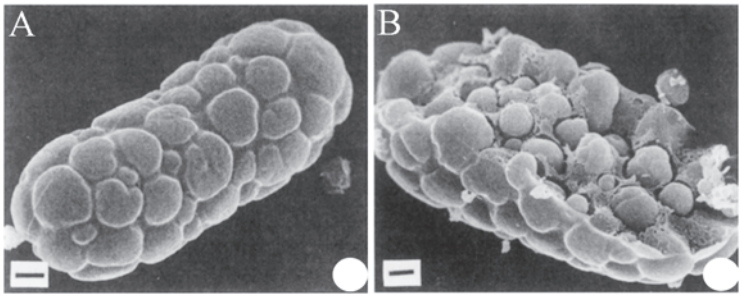


Figure 7. Scanning electron microscopic images of *Achromatium oxaliferum*. (A) An elongated cell is presented from outside and (B) from inside when the cell has been disrupted. In both cases, the larger calcium carbonate inclusions and the smaller sulfur globules can be seen. Apparently, the complete interior of the cell is filled with these granules. The cytoplasm is restricted to the interstitial space between the inclusion bodies. Scale bar represents 1.25 µm. (Adapted from Head *et al.*, 2000)

In 1948, van Niel proposed to combine all so far described morphologies of large, spherical sulfur bacteria into the genus *Achromatium* (Van Niel, 1948). Thereby, he rejected the names *Hillhousia*, *Thiophysa* (see below) and *Thiosphaerella* (introduced by Nadson, 1913, but not mentioned any further here), and maintained names of other smaller unicellular sulfur bacteria like *Thiovulum* and *Thiospira* (both names still existing today but not mentioned any further here). He differentiated two species: (1) *Achromatium oxaliferum* as freshwater organisms containing the conspicuous calcium carbonate inclusions and (2) *Achromatium volutans* as marine representatives containing exclusively sulfur inclusions and featuring a large central vacuole (see below *Thiophysa volutans*). Today, only the species *A. oxaliferum* is recognized (Skerman *et al.*, 1980; Skerman *et al.*, 1989) and is, although not in culture, the type species of the genus *Achromatium* (Garrity *et al.*, 2005).

Genus *Thiophysa* Hinze 1903

In the beginning of the last century, another type of spherical sulfur bacteria was discovered in marine coastal waters near Naples, which differed in more than habitat salinity from freshwater *Achromatium*. Though featuring similar cell sizes and a comparative motility, they did not contain any of the large calcium carbonate inclusions (Fig. 8, Hinze, 1903). Furthermore, these organisms were the first among the large sulfur bacteria, in which it was described that the cytoplasmic layer, which putatively contained the sulfur globules, surrounded a large intracellular vacuole. Also, cell division was slightly aberrant from regular binary fission. The division plane was flattened, so that the two emerging cells appeared like ball scrapers (Fig. 8, Hinze, 1903). Daughter cells eventually rounded subsequent to their release. These organisms were named *Thiophysa volutans*, meaning *thium* sulfur, *physa* bubble and *volutans* motile. Diameters of the cells ranged from 7−18 µm and the lengths in division stages were up to 29 µm (Hinze, 1903).

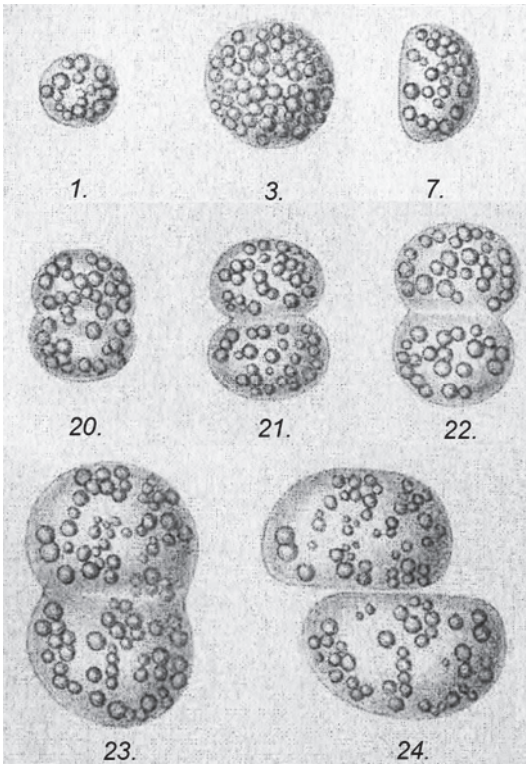


Figure 8. Drawing of *Thiophysa volutans* by Hinze. Cells were mostly spherical and contained many sulfur inclusions being assembled in the periphery and surrounding a large central vacuole (e.g. 1 and 3). When dividing, the division plane did not show a narrow constriction between the two emerging cells as in *Achromatium oxaliferum*, but the division plane was flat (e.g. 22 and 23) and the two emerging cells were leveled on one side (e.g. 7 and 24). Magnification is about 950×. (Adapted from Hinze, 1903)

The physiology of *Thiophysa* was considered to be similar to other lithotrophic large sulfur bacteria like *Beggiatoa*, because *Thiophysa*-populated habitats smelled of sulfide and intracellular sulfur globules disappeared when kept in sulfide-free water (Hinze, 1903). They required oxygen, but seemed to be rather microaerophilic, and were additionally sensitive to high sulfide concentrations (Nadson, 1913). Only one more

species was described, *Thiophysa macrophysa* (meaning large bubble), in which cell diameters reached up to 40 µm (Nadson, 1913). In one case, a gelatinous sheath was observed to contain up to 50 spherical, non-motile sulfur bacterial cells resembling *Thiophysa volutans*. Diameters of single cells were maximal 10 µm (Devide, 1952).

The genus name *Thiophysa* was later rejected and the species *Thiophysa volutans* was transferred to the genus *Achromatium* as the new species *Achromatium volutans* (see above, Van Niel, 1948). *A. volutans* were occasionally localized in different habitats (Dando *et al.*, 1995; Larkin & Henk, 1996). Eventually, neither the genus *Thiophysa*, nor the species *Achromatium volutans* have ever been recognized in the Approved Lists of Bacterial Names or considered as taxon in the current edition of Bergey´s Manual of Systematic Bacteriology. Thus, single, spherical, sulfur-containing cells with a large central vacuole have no valid name until today and remain uncultured.

Genus *Thioploca* Lauterborn 1907

By the beginning of the last century, a third large and filamentous sulfur bacterial morphotype was discovered during an investigation of the sediments of Lake Constance, Germany. Unlike *Beggiatoa* occurring as single filaments, this new morphotype exhibited trichomes that were bundled in a common sheath (Fig. 9A, Lauterborn, 1907). Within such sheaths, the trichomes were often twisted around each other and the organisms were accordingly named *Thioploca*, meaning *thium* sulfur and *ploca* braid. Lauterborn dedicated this discovery to his friend W. Schmidle, resulting in the name *Thioploca schmidlei*. Individual trichomes featured a gliding motility and diameters varied between 5−9 µm. The terminal cells of a trichome were often tapered. Depending on the amount of trichomes enclosed, the outer sheath exhibited dimensions of 50−160 µm in width and 3−4 cm in length (Lauterborn, 1907). Shortly after, a second species, *Thioploca ingrica*, was discovered and named after the place of discovery, Ingria, the old term for a district southwest of St. Petersburg (Fig. 9B, Wislouch, 1912). Being also freshwater organisms and sharing a similar morphology with *T. schmidlei*, a species differentiation of *T. ingrica* was again based on trichome diameters, which were between 2−4.5 µm, and bundle dimensions reaching up to 80 µm width and 1 cm length (Wislouch, 1912).

A few years later, another two species were described from Lake Constance, *Thioploca minima* (meaning smallest) and *Thioploca mixta* (meaning mixed) (Koppe, 1923). The

first comprised bundles of trichomes with extremely thin diameters (0.8−1.5 µm) and in the second species a single sheath contained trichomes of variable diameters (1 µm or between 6−8 µm, Koppe, 1923). It was proposed that the sheath is produced gradually by the *Thioploca* trichome itself aging over time, because single trichomes were observed to be surrounded by a soft, tight sheath and multiple trichomes by a rather stiff, loose sheath (Koppe, 1923). Multiplication of trichomes within sheaths were proposed to happen via a production of short trichome fragments, comparable to hormogonia production in Cyanobacteria, or by the release of single cells from a trichome, which then form another trichome within the sheath (Kolkwitz, 1955). In some cases, up to a few dozen trichomes were counted within a single sheath (Lauterborn, 1907). However, it was proposed that trichomes do not require their sheaths at all times (Koppe, 1923). Instead, it was observed that they occasionally left the sheath and even survived as free-living filaments for at least a week in a slide culture (Koppe, 1923). Also many empty sheaths were observed (Koppe, 1923). The re-entering of a free-living filament into an existing sheath was suggested impossible as trichome-inhabited sheath openings were considered to be too narrow (Kolkwitz, 1955). The development of *Thioploca mixta* was thus proposed to result from the joining of two filaments of different diameters that subsequently produce a common sheath. If this was true, of course, the validity of *T. mixta* as a taxon is questionable (Kolkwitz, 1955).

In contrast to *Beggiatoa*, which usually form thick mats on top and within the first few centimeters of the sediment, freshwater *Thioploca* spp. always spanned several tens of centimeters within the sediment and never occurred on the surface (Lauterborn, 1907). Usually, only a faint smell of hydrogen sulfide was detected in sediments, where they occurred (Kolkwitz, 1912; Kolkwitz, 1955; Lauterborn, 1907). This rather low amount of sulfide in the populated sediment was proposed to allow a deeper penetration of oxygen to support the growth of *Thioploca* spp. (Kolkwitz, 1912). In agreement with this hypothesis was the absence of an oxygen-consuming *Beggiatoa* mat directly on top of *Thioploca*-inhabited lake sediments (Kolkwitz, 1912). Nevertheless, it was proposed that *Thioploca* can survive short periods of anaerobic conditions as their habitats become occasionally oxygen-depleted due to fouling (Kolkwitz, 1955). Requirement of sulfide for the metabolism of *Thioploca* was deduced from slide culture experiments, in which free-living filaments revealed a positive chemotaxis towards hydrogen sulfide (Koppe, 1923). Sensitivity towards elevated sulfide and oxygen concentrations was shown later in enrichment cultures (Maier, 1980).

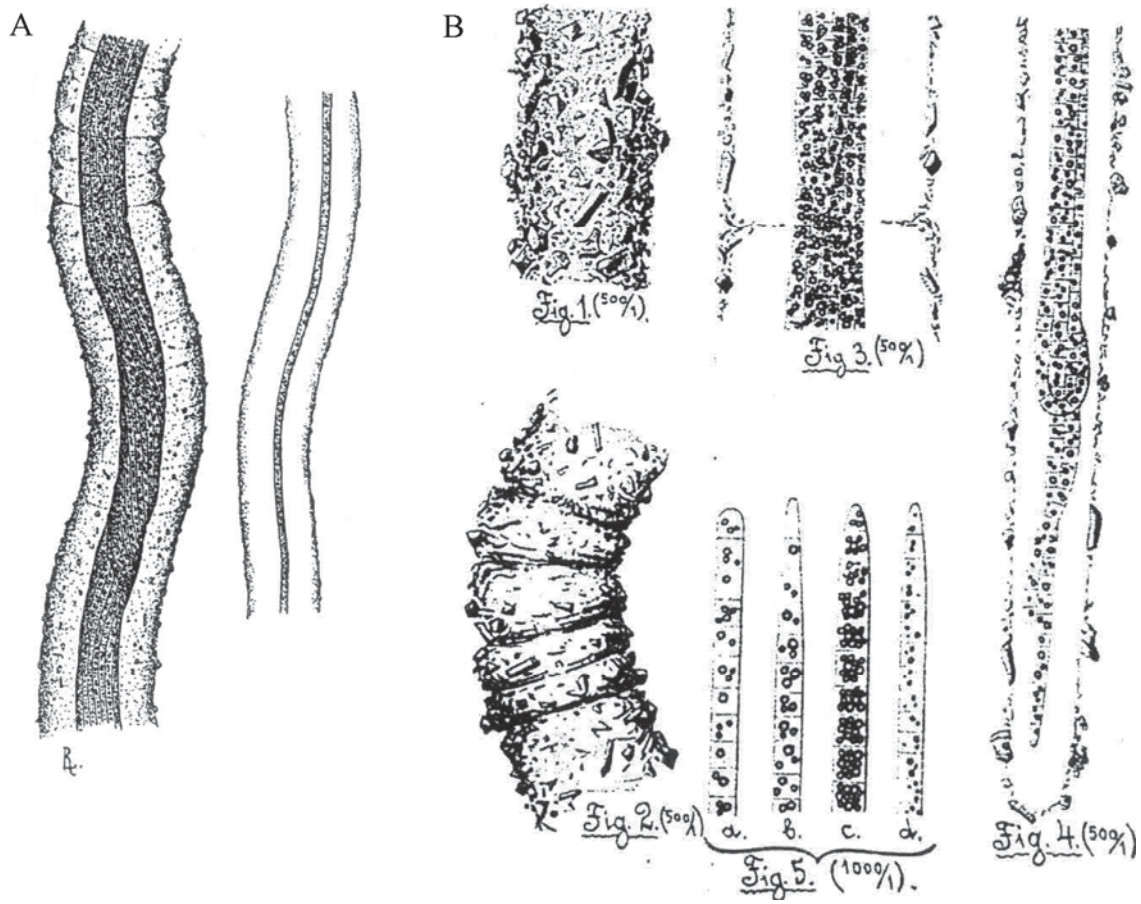


Figure 9. (A) Drawing of *Thioploca schmidlei* by Lauterborn, showing parts of a wide and narrow trichome, each residing inside a mucus sheath. Magnification is about 200×. (Adapted from Lauterborn, 1907) (B) Drawing of *Thioploca ingrica* by Wislouch, showing the outer appearance of a smooth mucus sheath (upper left) and a wrinkled sheath (lower left). Filaments of different widths and number of sulfur granules are shown as individuals (lower centre) and as cross-section within the mucus sheath (upper centre and right). Magnifications are indicated below each drawing. (Wislouch, 1912)

In the late 1970s, sheath-inhabiting trichomes were discovered for the first time in marine environments. Large communities of the bundle-forming organisms were found along the coastline off Peru and Chile, influenced by the oxygen-depleted but nutrient-rich Peru-Chile Subsurface Countercurrents (Gallardo, 1977). In this first description, an apparent morphological resemblance to *Thioploca* spp. was noted (Fig. 10A, Gallardo, 1977), also reporting a gliding of individual trichomes and tapered terminal cells. Diameters of individual trichomes formed three groups: 30−40 µm, 15−20 µm and 2.5−5 µm. Sheaths were 100−500 µm in diameter and several centimeters in length (Gallardo, 1977). Two species were distinguished based on trichome diameters (Maier & Gallardo, 1984a), *Thioploca araucae* (30−43 µm; named after Arauco, a city, gulf and province in Chile) and *Thioploca chileae* (12−20 µm; meaning of Chile). Again, it was reported that trichomes of different diameters can populate one individual sheath

(Maier & Gallardo, 1984a), and for these marine thioplocas the re-entering of a free-living filament into an inhabited sheath was considered as a likely process (Schulz *et al*., 1996). Marine thioplocas occasionally reached trichome diameters of 125 µm (Jørgensen & Gallardo, 1999), were microaerophilic, required low concentrations of sulfide (Maier & Gallardo, 1984a) and were probably mixotrophic (Maier & Gallardo, 1984b).

The two marine species represented the first members of the genus *Thioploca* to contain large central vacuoles (Maier & Gallardo, 1984a). At that time, this feature was already described for the unicellular relative *Thiophysa* (Hinze, 1903) and was about to be detected also in wide, marine *Beggiatoa* and *Thiomargarita* a few years later (Nelson *et al.*, 1989; Schulz *et al.*, 1999). Besides being the youngest genus among the large sulfur bacteria in the 1980s and being the first that was found to populate vast areas of marine sediments (Gallardo, 1977; about 10,000 km^2, Jørgensen & Gallardo, 1999), *Thioploca* spp. were also the first representatives, in which nitrate storage inside the voluminous vacuole was shown (Fig. 10B, Fossing *et al.*, 1995). Intracellular nitrate concentrations were as high as 500 mM and used as alternative electron acceptor (Fossing *et al.*, 1995), while being reduced to ammonia (Otte *et al.* 1999).

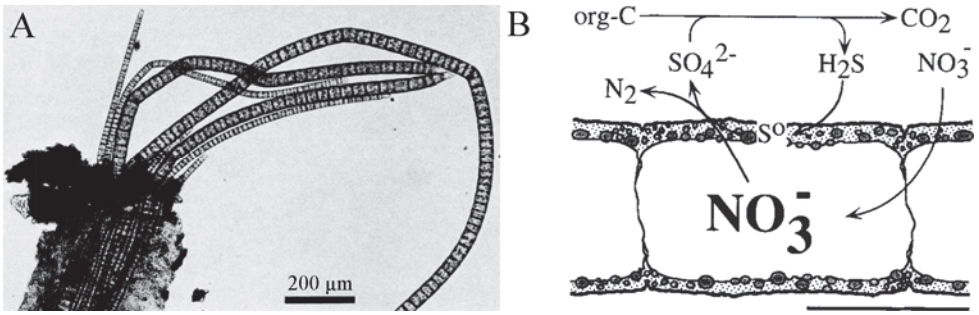


Figure 10. (A) Bundle of trichomes found in marine sediments off Chile. Except for the difference in diameter, the morphology of the organisms and the bundle-formation strongly resembles freshwater *Thioploca*. (Adapted from Maier & Gallardo, 1984a) (B) Each cell of a *Thioploca*-like trichome has a large central vacuole and a thin cytoplasmic, sulfur-containing rim. Within this vacuole, nitrate can be concentrated to about 20,000× above ambient seawater concentration. Nitrate can be used as alternative electron acceptor for sulfide/sulfur oxidation. Scale bar represents 40 µm. (Fossing *et al.*, 1995)

Thioploca schmidlei is the type species of the genus *Thioploca* (Skerman *et al.*, 1980; Skerman *et al.*, 1989), although never cultured or relocated since 1923 (Maier & Preissner, 1979). In 1984, *T. ingrica* was re-identified (Maier, 1984) and was, together with *T. araucae* and *T. chileae*, published in the International Journal of Systematic Bacteriology. These four species are currently recognized, but still remain uncultured.

Genus *Thiomargarita* Schulz *et al.* 1999

Remarkably, it needed more than ninety years until another genus of large, colorless sulfur bacteria was discovered. During a cruise in 1997, a fourth type of this group was detected in organic-rich, marine sediments, forming large populations along the Benguela upwelling area off Namibia (Schulz *et al.*, 1999). Being neither filamentous nor unicellular, this morphotype formed spherical or ovoid cells that were held together in a common sheath forming linear chains (Fig. 11A, Schulz *et al.*, 1999). Referring to their resemblance to a pearl necklace, they were named *Thiomargarita namibiensis*, meaning *thium* sulfur, *margarita* perl and *namibiensis* from Namibia. Most cells featured diameters between 100−300 µm, but maximum diameters of 750 µm rendered these organisms the largest prokaryotes ever (Schulz *et al.*, 1999). Corresponding to other sulfur bacteria, *Thiomargarita* also contained large internal vacuoles, which accounted for 98% of the cellular biovolume, and contained nitrate to concentrations of up to 800 mM (Schulz *et al.*, 1999).

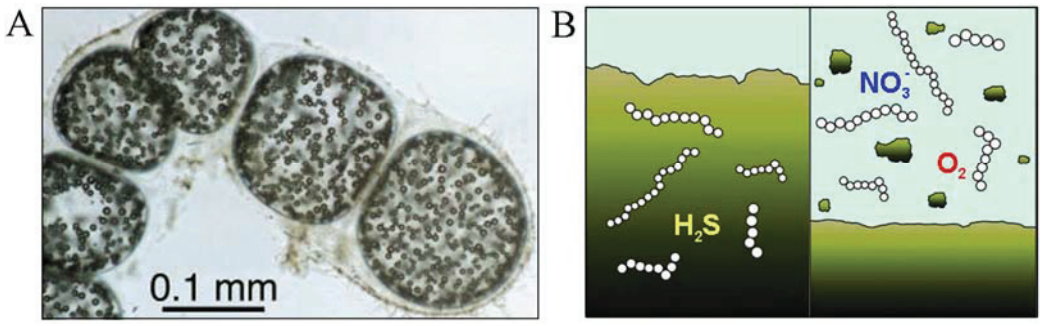


Figure 11. (A) Microscopic image of a *Thiomargarita namibiensis* chain. The round cells are not connected like they are e.g. in a *Beggiatoa* filament (only during cell division, two cells upper left). The interior of each cell appears hollow as it is filled with an aqueous vacuole. (Adapted from Schulz *et al.*, 1999) (B) Scheme showing the life strategy of *Thiomargarita namibiensis*. Non-motile chains are either buried in sulfidic sediments, oxidizing sulfide/sulfur with stored nitrate (left) or get resuspended with the overlying water containing oxygen and nitrate (right). (Adapted from Schulz, 2006)

In contrast to many of their relatives, motility in *Thiomargarita* was not observed. Whereas motile beggiatoas and thioplocas move actively between the gradients of sulfide and oxygen or nitrate, the life strategy of *Thiomargarita* was considered rather passive. Exposure to sulfide (within the sediment) and oxygen/nitrate (above the sediment) was suggested to be realized by temporary resuspension of the sediment with overlaying water, transporting also the *Thiomargarita* chains (Fig. 11B, right). Subsequent settling of the sediment again buried the *Thiomargarita* in sulfidic sediments (Fig. 11B, left). Probably due to this dependency, *Thiomargarita* seems to be

extremely tolerant towards both high sulfide and high oxygen concentrations (Schulz *et al.*, 1999). The usage of oxygen as electron acceptor (Schulz & de Beer, 2002) and the storage of large amounts of polyphosphate (Schulz & Schulz, 2005) was also reported. Like *Beggiatoa*, *Thiomargarita* can release phosphate in large amounts into the sediment, thereby allowing for the formation of apatite (Goldhammer *et al.*, 2010).

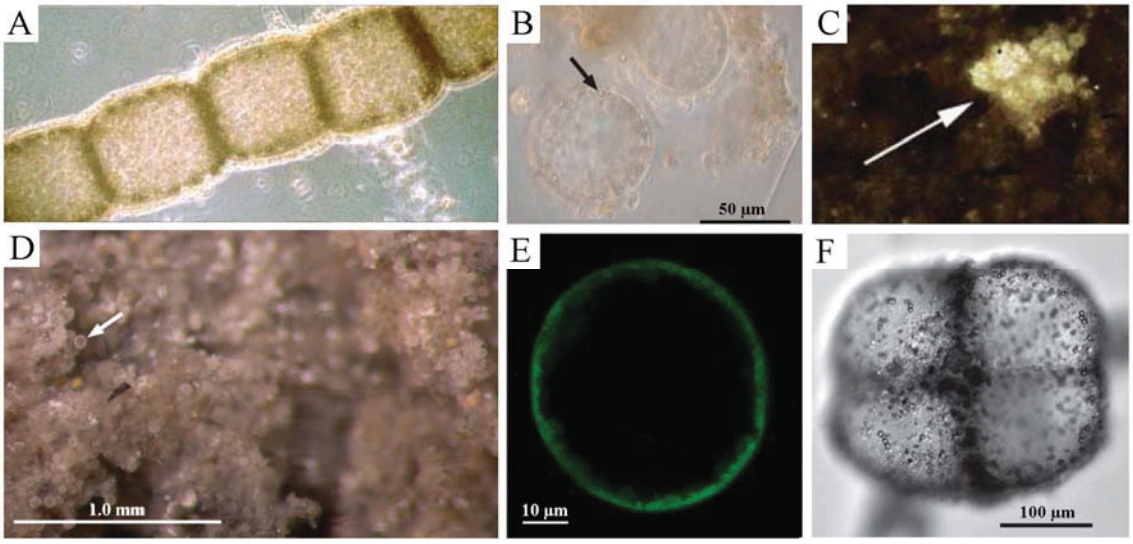


Figure 12. Images of different morphologies described for cells resembling *Thiomargarita*. (A) Another chain-forming type containing barrel-shaped cells was described from Namibian sediments (Schulz, 2006). (B) Also spherical single cells were observed (D) forming mats in the vicinity of the Amon mud volcano in the Mediterranean Sea (Girnth *et al.*, 2011). (C) Clump-forming cells were reported in Namibian sediments (Schulz, 2006) and (F) cube-like structures in sediments of the Gulf of Mexico (Kalanetra *et al.*, 2005). (E) Most cells were reporteded to be vacuolated (cytoplasma stained with fluorescein isothiocyanate, Girnth *et al.*, 2011)

In the following years, cells with a *Thiomargarita*-like morphology, but not forming chains, were detected in many other marine habitats, like the Gulf of Mexico (Kalanetra *et al.*, 2005), the Arctic Sea (de Beer *et al.*, 2006) or the Mediterranean Sea (Girnth *et al.*, 2011). Only one other chain-forming type was detected with cells having a rather barrel-shaped morphology (Fig. 12A, Schulz, 2006). Besides this, cells either formed large clumps (Fig. 12C, Schulz, 2006) or were regularly arranged in clusters resembling a cube (Fig. 12F, Kalanetra *et al.*, 2005). Single, spherical and vacuolated sulfur bacteria (Fig. 12B and E) were also found lately and referred to as *Thiomargarita*-like (de Beer *et al.*, 2006; Gallardo & Espinoza, 2007; Girnth *et al.*, 2011), without mentioning their morphological resemblance to *Achromatium volutans* (*Thiophysa volutans*). The cube-like forms (Fig. 12F) were proposed to undergo reductive cell division instead of regular binary fission observed for *Thiomargarita namibiensis*

(Kalanetra *et al.*, 2005). Although featuring aberrant morphologies than the originally described *T. namibiensis*, new species names were not introduced for the new cell types. *T. namibiensis* is the only recognized species according to the Validation List (1999) and is the type species of the genus *Thiomargarita* (Schulz & Jørgensen, 2005). All so far described thiomargaritas were marine, larger than 30 µm, vacuolated, considered as non-motile and have remained uncultured.

Conspicuous morphological diversity among large sulfur bacteria in Namibian sediments

Already during my diploma thesis in 2007, I examined the morphological diversity of large sulfur bacteria in Namibian sediments retrieved from the Benguela upwelling area (Gutzke, 2007). A co-occurrence of those five morphotypes already described for non-filamentous *Thiomargarita*-like sulfur bacteria (see above) was detected in these sediments (Fig. 13A−E). Additionally, a yet unrecognized morphotype was discovered, being one or more spherical cells contained in a rigid envelope (Fig. 13F). When this envelope was manipulated, the cells expelled as if the interior was aqueous. Namibian sediments apparently contained a much more diverse population of large sulfur bacteria than previously recognized (Gutzke, 2007).

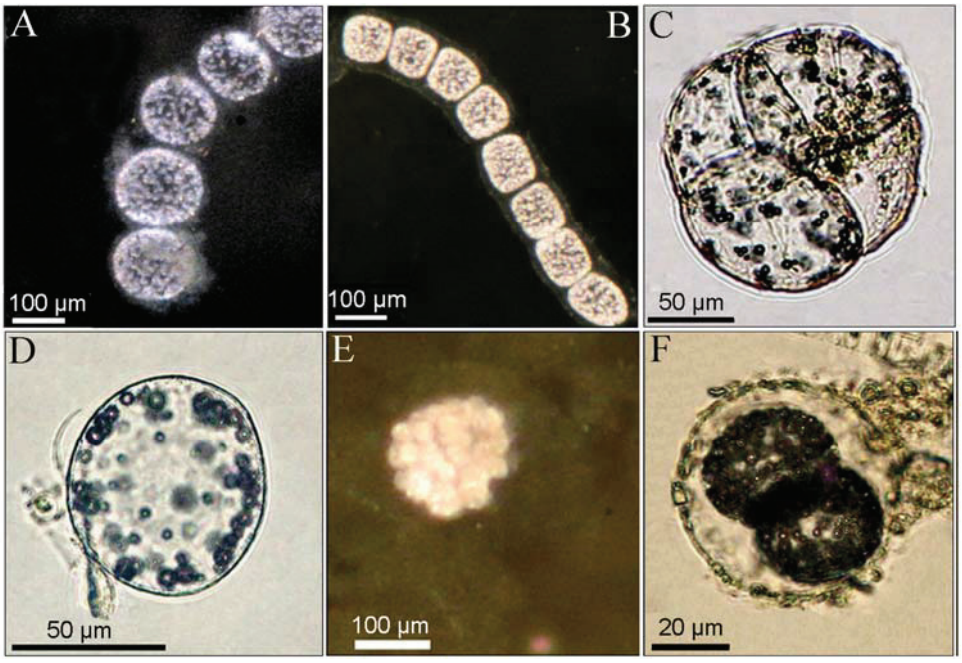


Figure 13. Morphological diversity of large sulfur bacteria in Namibian sediments: (A) typical round, chain-forming cells described for *T. namibiensis*, (B) chain-forming cells having a rather cylindrical or barrel-like shape, (C) cells that form a regularly arranged cube-like structure, similar to *Thiomargarita* sp. described from the Gulf of Mexico, (D) single spherical cells, (E) clusters of cells, like a cauliflower (Schulz, 2006), and (F) the newly discovered morphotype of enveloped spherical cells. (Gutzke, 2007)

All cells examined had a large central vacuole, which was visible as an unstained central compartment when stained with fluorescein isothiocyanate (FITC, Fig. 14), Acridine Orange or Nile Red. These dyes bind to different classes of molecules that are always situated in the cytoplasm of a cell. Thus, only a thin rim just beneath the cytoplasmic membrane was stained and visualized with a confocal laser scanning microscope (Fig. 14). Further studies on this cytoplasmic layer included scanning electron microscopy (SEM) analysis on chain-forming morphotypes. Cells were disrupted so that the interior of the cells could be visualized. Comparable to similarly performed examinations on *Beggiatoa* cells (Fig. 15B, Larkin & Henk, 1996), the cytoplasm is not a smooth layer surrounding the vacuole, but appears as a sponge-like network of strings and fibers, which are occasionally bulged by differently sized granules and inclusions, most of them being probably sulfur granules (Fig. 15A).

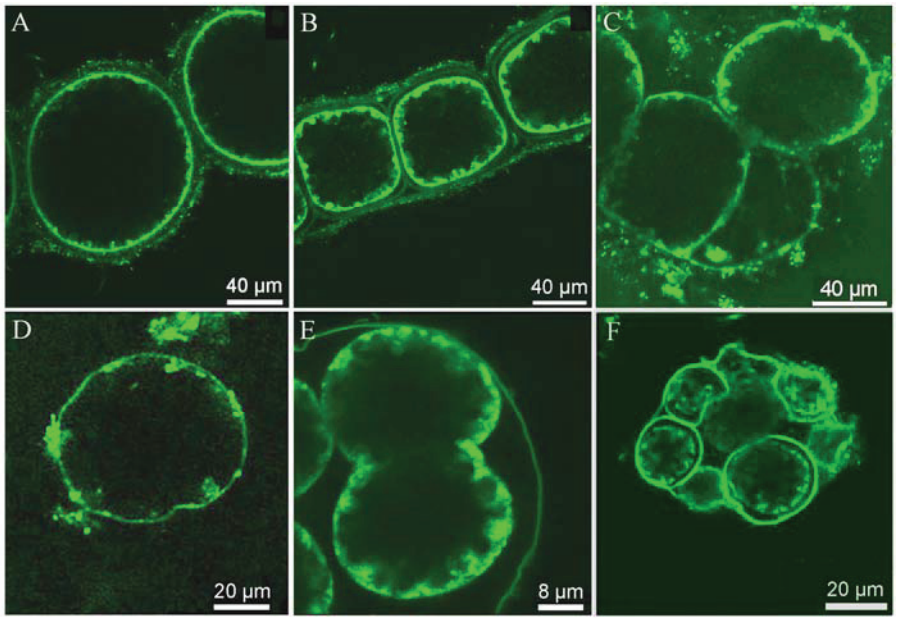


Figure 14. *Thiomargarita*-like morphotypes stained with fluorescein isothiocyanate (FITC). All cells contain a large central vacuole and the cytoplasm is reduced to a thin rim beneath the cytoplasmic membrane. (E) In enveloped cells, also the envelope is stained by this dye (also visible with the other dyes used), suggesting that it is composed of organic compounds, which are probably produced by the cells themselves. (F) In aggregated, clump-like cells, each cell seems to be surrounded by an individual layer similar to the envelope in (E). (Gutzke, 2007)

Furthermore, a FISH probe that was published as being specific to most sulfide-oxidizing bacteria (Kalanetra *et al.*, 2004) was tested on chain-forming morphotypes. With this probe, not only the typical round, chain-forming cells of *Thiomargarita namibiensis* could be stained (Fig. 15C), but also cylindrical and aberrantly shaped, chain-forming cells.

Finally, the process of regular binary fission in *Thiomargarita* spp. was opposed to the proposed reductive division in certain *Thiomargarita* sp. (Kalanetra *et al.*, 2005) and was suggested to be realized also in aggregated morphotypes. Therefore, several possible cycles with interchanging sizes and aggregate stages were postulated (Fig. 15D). Moreover, a morphological resemblance to certain fossils (Fig. 15E) was demonstrated.

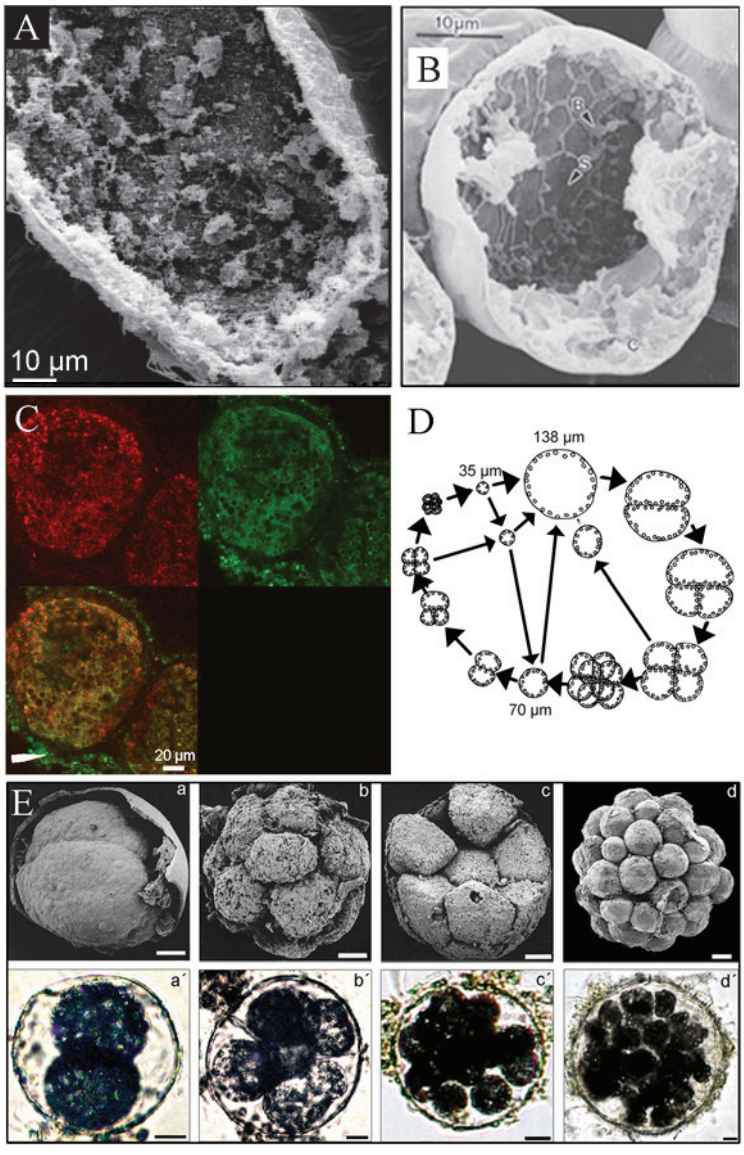


Figure 15. (A) SEM image of a disrupted *Thiomargarita*-like cell. (Gutzke, 2007). The cytoplasm in (A) appears similar to the cytoplasm in a disrupted *Beggiatoa* filament shown in (B) (Larkin & Henk, 1996). (C) FISH hybridizations with VSO673 (upper left panel) and EUB338 (upper right, overlay in lower left). Arrow points to the sheath being populated by other bacteria exclusively stained with EUB338 (Gutzke, 2007). (D) Putative cycle of reductively dividing *Thiomargarita*-like cells. (Adapted from Gutzke, 2007). (E) Comparison of fossils of the Doushantuo formation (upper row, scales are 100 µm) and contemporary *Thiomargarita*-like morphotypes (lower row, scales are 10 µm). (lower row, Gutzke, 2007; upper row, Xiao *et al.*, 2000)

Evolving phylogenetic data for the classification of organisms

Starting in the mid-1990s, the first phylogenetic data of large sulfur bacteria became available. Immediately, this brought about a major conflict between the traditional morphological identification and the modern molecular classification (Ahmad *et al.*, 1999; Ahmad *et al.*, 2006; Mussmann *et al.*, 2003). The large sulfur bacteria were not the only group, in which the modern phylogeny clashed with the former, morphology-derived classification. After the retrieval of large amounts of phylogenetic data, many taxa were eventually reclassified (e.g. the *Chromatiaceae*, Imhoff *et al.*, 1998; the genus *Thiobacillus*, Kelly & Wood, 2000; or the heterocytous cyanobacteria, Rajaniemi *et al.*, 2005).

In order to understand these sudden changes in the taxonomical concept, certain aspects about phylogeny need to be explained. The discipline of phylogeny is the overall study of evolutionary relatedness among organisms and it is necessary to specify, on which characteristics a phylogenetic study is based on. As introduced above, in former times the affiliation of an organism was mainly deduced from its morphological characteristics. Already then, the orderly arrangement of different phylogenetic groups was performed and a systematic ranking according to their distance in evolutionary relationship was established. This procedure is termed classification. With the help of a specific nomenclature, nowadays determined by the Bacteriological Code (Lapage *et al.*, 1992), the ranks of relatedness are defined. For example, there are specific endings for higher ranks, like *–eae* for families and *–ales* for orders. When it comes to species definition, a binominal nomenclature is used, which was already introduced by Carl Linnaeus (1735).

By studying their phylogeny, novel organisms can be either identified as one of the already classified organisms, called identification, or be newly positioned in the hierarchical system by the introduction of a new taxon, called classification. A taxon is a certain rank in the phylogenetic system with an assigned name to it. Taxa are therefore named entities or groups that share a certain relationship. Taxonomy includes all three disciplines: classification, nomenclature and identification (reviewed in Ludwig & Schleifer, 1994). Overall, our current understanding of the species concept is based on an evolutionary relationship of all living organisms that share a common descent (Darwin, 1859). This implies that all organisms of the same taxon share the same evolutionary relationship and are thus considered monophyletic.

Nowadays, phylogeny is, aside of phenotypic characteristics, predominantly based on molecular data (Ludwig & Schleifer, 1994). Already in the mid-1960s, the idea for the usage of sequences of macromolecules for reconstructing evolution and for determining phylogenetic relationships between living organisms evolved (Zuckerland & Pauling, 1965). For this purpose, molecules like DNA, RNA and polypeptides among others were proposed to be suitable (Zuckerland & Pauling, 1965). Shortly after, the quality of ribosomal RNA (rRNA) genes was tested on pure culture strains for qualification as phylogenetic markers (e.g. Fox *et al.*, 1977a; Fox *et al.*, 1977b; Woese *et al.*, 1976; Zablen *et al.*, 1975a). The rRNA genes were considered appropriate because they are universally present in every living organism and have a conserved function. They are not subject to horizontal genetic transfer and at least two of the three rRNA genes have a sufficient number of nucleotides to reflect evolutionary distance (reviewed in Pace *et al.*, 1986). Soon, the opinion emerged that the 16S rRNA gene was the most suitable because the 5S rRNA gene was too short to contain sufficient information and the 23S rRNA gene was too long for technical handling at that time with regard to sequencing (Fox *et al.*, 1977b). The entire length of the 16S rRNA gene was categorized into variable and conserved regions in order to view the potential of the molecule as molecular chronometer (Woese *et al.*, 1975; Woese *et al.*, 1980).

Information gained by 16S rRNA gene sequencing and comparison revolutionized the entire systematic of living organisms. The universal tree of life was subsequently divided into three lineages (Woese & Fox, 1977), namely the three domains of life: Bacteria, Archaea and Eucarya (Fig. 16, Woese *et al.*, 1990) – a systematic which is still the most accepted among the scientific community today. Other novel possibilities evolved, e.g. investigation of the ribosomal genes in chloroplasts revealed bacterial 16S rRNA molecules, which were opposed to nuclear 18S rRNA, and thus supported the endosymbiotic theory (Zablen *et al.*, 1975b). Furthermore, unculturable bacteria and many other yet undetected organisms could be investigated, starting with their phylogenetic affiliation when cloning and sequencing their rRNA gene sequences (Pace *et al.*, 1986). After nearly two decades, ribosomal RNA sequence data was accepted to be used for phylogenetic analyses and the study of bacterial evolution (Woese, 1987). In this time, a great invention enabled the specific amplification of genes several hundred thousand fold, called polymerase-chain-reaction (PCR), using oligonucleotides (Mullis *et al.*, 1986; Saiki *et al.*, 1985) and a thermostable polymerase (Saiki *et al.*, 1988). Soon, PCR was applied on universal regions in bacterial 16S rRNA genes (Wilson *et al.*,

1990), thereby revolutionizing the procedure for phylogenetic analyses and giving way to a tremendous gain in 16S rRNA gene sequence data for diversity studies. Nowadays, studying the diversity of microorganisms of a given habitat is commonly proceded as follows: extraction of bulk genomic DNA, application of PCR with universal (domain-specific) primers, separation of the amplicons in clone libraries and sequencing of all clones. The ribosomal RNA database project (RDP) was initiated to provide access to the tremendous amount of obtained rRNA gene sequences and to alignments and tools to operate on the sequences (Olsen *et al.*, 1991).

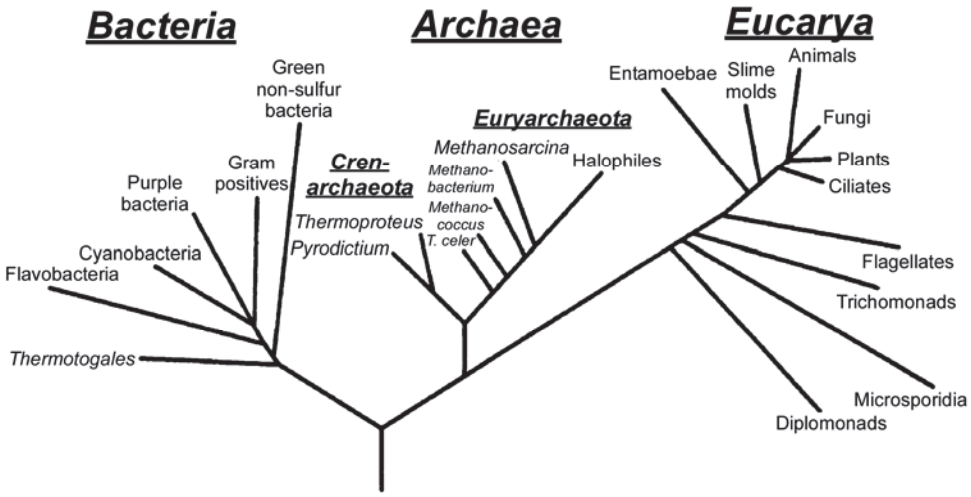


Figure 16. The universal tree of life based on sequence comparison of the 16S rRNA genes available in 1990. The term prokaryotes did not suffice anymore for a specific domain as prokaryotes are separated into two domains, namely *Bacteria* and *Archaea*. The latter is phylogenetically closer related to the third domain, the *Eucarya*, because the lowest bifurcation separates the branch containing current *Bacteria* and and the branch from which modern *Archaea* and *Eucarya*. (Adapted from Woese *et al.*, 1990)

According to the Darwinian theory of evolution, organisms of the same taxon must share a monophyletic relationship. When analyzing gene sequences for phylogenetic studies, like the 16S rRNA gene, phylogenetic trees can be reconstructed based on sequence alignment and comparative sequence analysis, revealing the relative affiliation of organisms. The monophyly of taxa in an evolutionary sense is interpreted from a tree as those branches emerging from the same node. In those cases, where the exact relationship of phylogenetic clusters cannot be unambiguously resolved by different treeing methods, a multifurcation is introduced, which means that the given branches emerge from the same baseline (Ludwig & Schleifer, 1994; Peplies *et al.*, 2008).

At present, it is standard practice to identify microorganisms according to their 16S rRNA gene sequences, which are compared to already available sequence data, at least

down to the genus level. Also classification of the higher ranks, beyond genus level, is today exclusively based on 16S rRNA gene sequence comparison (Table 1). However, species or strain differentiation based on 16S rRNA genes is disputable as the molecule (length of maximal 1600 nucleotides) does not provide enough evolutionary information (Fox *et al.*, 1977b; Fox *et al.*, 1992; Stackebrandt & Goebel, 1994; Tindall *et al.*, 2010). In this case, a multiphasic approach to an accurate identification or classification is suggested, including physiological traits, lipid analysis, G+C content, DNA-DNA-hybridization, multilocus sequence typing and sequencing of intergenic transcribed 16S−23S rDNA spacer (ITS1) region (Feil, 2004; Gevers *et al.*, 2005; Stackebrandt & Goebel, 1994; Stackebrandt *et al.*, 2002; Tindall *et al.*, 2010).

Table 1. Currently suggested cut-off levels for taxa based on 16S rRNA gene sequence identity. Values for the different ranks are sometimes either not interconnecting or even overlapping, implying that the values are only a reference and that the actual determination of the taxonomic rank needs to be verified by means of the multiphasic approach (Feil, 2004; Tindall *et al.*, 2010).

taxonomic level	**16S rDNA gene sequence identity**	**reference**
one species	98.7−99%	(Stackebrandt & Ebers, 2006)
separate species	<97%	(Amann *et al.*, 1992; Collins *et al.*, 1991; Fox *et al.*, 1992; Martinez-Murcia *et al.*, 1992; Stackebrandt & Goebel, 1994; Tindall *et al.*, 2010)
one genus	95−98%	(Tindall *et al.*, 2010; Yarza *et al.*, 2008)
separate genera	<95%	(Tindall *et al.*, 2010; Yarza *et al.*, 2008)
one family	86−93%	(Yarza *et al.*, 2008)
separate families	<86%	(Yarza *et al.*, 2008)

In order to eventually classify a novel organism or strain, standards are currently very specific and complex (outlined in Tindall *et al.*, 2010). Most importantly, an organism can only be legitimately classified when obtained in pure culture and deposited in at least two type culture collections in different countries (Tindall *et al.*, 2006; Tindall *et al.*, 2010). A new name is only validly published when listed in the Approved Lists of Bacterial Names (Skerman *et al.*, 1980; Skerman *et al.*, 1989), published in the International Journal of Systematic and Evolutionary Microbiology (IJSEM) and being conform with the Bacteriological Code (Lapage *et al.*, 1992), or published elsewhere but subsequently listed in one of the Validation Lists. In any case, a name can only be considered as validly published when meeting the requirements of the rules laid down in the Bacteriological Code (Tindall *et al.*, 2006). Regarding organisms, of which more than a 16S rRNA gene sequence is available but which yet remain uncultured, a *pro forma* classification can be suggested by introducing a *Candidatus* taxon (Murray & Schleifer, 1994; Murray & Stackebrandt, 1995; Stackebrandt *et al.*, 2002).

Recent changes in the classification of large sulfur bacteria

As introduced above, there emerged a conflict between the traditional classification of the large sulfur bacteria based on morphological characteristics and the 16S rRNA-based phylogeny. Traditionally, identical or similar morphotypes, like a *Beggiatoa*-like filament, were named as a species of the same genus (*Beggiatoa* sp., e.g. Kalanetra *et al*. 2004; Mussmann *et al.*, 2003; Hinck *et al.*, 2007). However, 16S rRNA gene sequencing of several such filaments revealed that they do not share an evolutionary monophyletic relationship (Fig. 17, Ahmad *et al.*, 1999; Ahmad *et al.*, 2006; Mussmann *et al.*, 2003). The same is true for the bundled *Thioploca*-like morphotypes (Fig. 17, Ahmad *et al.*, 1999; Ahmad *et al.*, 2006). Only the distantly related genera of filamentous *Thiothrix* spp. and *Leucothrix* sp. and of the unicellular *Achromatium* seem to be correctly classified according to the Bacteriological Code (Lapage *et al.*, 1992) as they form distantly related, separate branches (Fig. 17).

In the past decade, numerous attempts to cultivate large sulfur bacteria were performed. Until today, no representative of these bacteria, which has a large central vacuole and stores nitrate, has been successfully cultured, not to mention achieving a pure culture. Besides several non-vacuolated freshwater *Beggiatoa* strains (see above), there existed two non-vacuolated marine pure culture *Beggiatoa* strains (MS-81-1c and MS-81-6, Nelson *et al*., 1982), which, however, are not available anymore. For some organisms of this group, not only culturing is a major challenge for studying the bacteria, but also sequencing of the entire 16S rRNA gene seems difficult. For the wide, marine *Thioploca*-like organisms and the novel genus *Thiomargarita*, only partial (less than 800 nucleotides) sequences were available prior to this thesis (Kalanetra *et al.*, 2005; Schulz *et al.*, 1999; Teske *et al.*, 1995). Furthermore, attempts to identify large sulfur bacteria in their natural habitats using universal clone libraries often failed as well (Angert *et al.*, 1998; Edgcomb *et al.*, 2002; Gillan *et al.*, 1998; Lopez-Garcia *et al.*, 2003; Sekar *et al.*, 2006; Stevens & Ulloa, 2008).

These major hurdles, namely the frequent failure in culturing the majority of large sulfur bacteria and the common difficulty to retrieve 16S rRNA gene sequence data, has surely prevented from a thorough reclassification of the organisms so far.

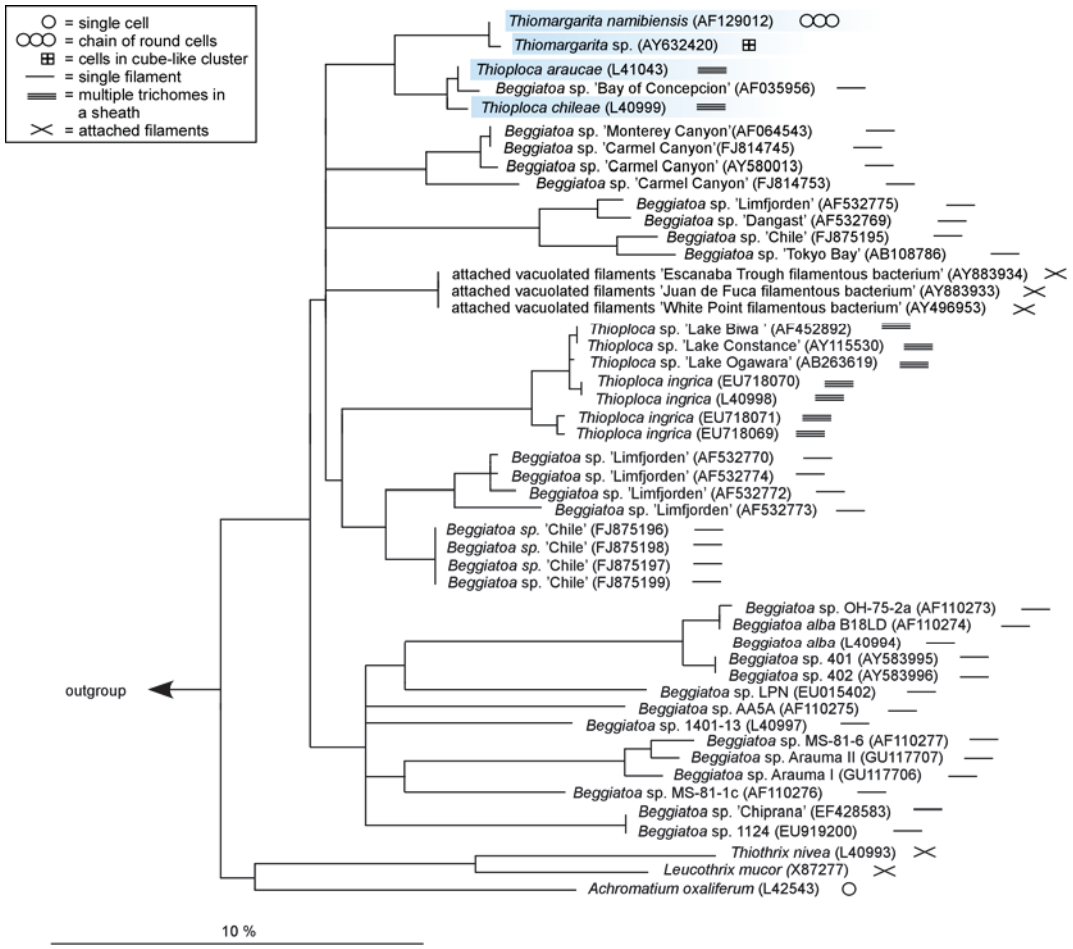


Figure 17. Multifurcation tree based on nearly full-length 16S rRNA gene sequences of large sulfur bacteria available prior to this thesis. Highlighted in blue are partial sequences (less than 800 nt). This nomenclature is not following the guidelines for modern classification based on the relative identity of 16S rRNA gene sequences and the concept of monophyletic taxa. Instead, genus names like *Thioploca* exhibit a paraphyletic and *Beggiatoa* even a polyphyletic relationship. Some sequences were not referred to with a name, yet. Besides nine *Beggiatoa* strains, which are all in the cluster including *Beggiatoa alba*, all other sequenced organisms are not in pure culture.

Among the large sulfur bacteria, not only the classification of genera is an issue of debate, but also that of the families. The family name *Beggiatoaceae* was introduced by Migula (1894) and all large sulfur bacteria, which were discovered since then, were assigned to this family. In the latest edition of Bergey´s Manual of Determinative Bacteriology published in 1974, the family included the genera *Beggiatoa*, *Thioploca* and *Vitreoscilla* (the latter not being treated any further here, Leadbetter, 1974). The genera *Thiothrix* and *Leucothrix* were assigned to a separate family, the *Leucotrichaceae*, and the genus *Achromatium* was likewise assigned to a different family, the *Achromatiaceae* (Buchanan & Gibbons, 1974). Since 1974, there was no further edition published, but the Approved Lists of Bacterial Names were released in

1980, and amended in 1989 (Skerman *et al.*, 1980; Skerman *et al.*, 1989). These Lists also recognized the three separate families. In 1984–89, the first edition of the novel book series Bergey´s Manual of Sytematic Bacteriology was released and again the three families *Beggiatoaceae*, *Leucotrichaceae* and *Achromatiaceae* were listed (Staley *et al.*, 1989). This time, however, the *Beggiatoaceae* included the genera *Beggiatoa*, *Thioploca* and *Thiothrix* (Strohl, 1989).

In the second edition of this book series and for the first time ever, phylogenetic data (16S rRNA gene sequences) were included for classification. Surprisingly, the authors did not reclassify the disputable genera or accepted only sequenced or cultured organisms. Instead, they reclassified the family and recognized a mix of cultured and uncultured species, even listing uncultured organisms as type species (Garrity *et al.*, 2005). The reclassification combined all genera of large, colorless sulfur bacteria, and even some other, smaller sulfur bacteria, into one family with the new name *Thiotrichaceae*. This family henceforth included eight genera: *Beggiatoa*, *Thioploca*, *Thiothrix*, *Leucothrix*, *Achromatium*, *Thiomargarita*, *Thiobacterium* and *Thiospira*. However, this classification is against rule 51b of the Bacteriological Code (Lapage *et al.*, 1992). This rule includes one paragraph explaining when a name is illegitimate: “If the taxon to which the name was applied, as circumscribed by the author, included the nomenclatural type of a name which the author ought to have adopted under one or more of the Rules. Example: If an author circumscribes a genus to include *Bacillus subtilis*, the type species of the genus *Bacillus*, then the circumscribed genus must be named *Bacillus*.”

Accordingly, the name *Thiotrichaceae* as a taxon is illegimate, because it includes three type genera of separate families (*Beggiatoa* of the *Beggiatoaceae*, *Leucothrix* of the *Leucotrichaceae* and *Achromatium* of the *Achromatiaceae*). In spite of this conflict, the family name *Thiotrichaceae* (and the respective order name *Thiotrichales*) was recognized as valid names in Validation List no. 106 (Euzéby, 2005).

Aims and Objectives

In my PhD thesis, the analysis sulfur bacterial diversity in Namibian sediments was continued, revealing an extensive morphological diversity among these bacteria. Considering in particular the non-filamentous sulfur bacteria, which usually exhibited a *Thiomargarita*-like morphology, this novel diversity was in no way corresponding to the phylogenetic diversity known at that time because only two partial sequences (less than 800 nucleotides each) for the genus *Thiomargarita* were available.

The immediate next step, was to resolve the putative underlying phylogenetic diversity among the different morphotypes detected in Namibian sediments. Therefore, specific primers were designed, according to the partial sequences, in order to obtain nearly full-length 16S rRNA gene sequences and additional sequence information of the intergenic transcribed spacer (ITS) region. In a second step, a novel single-cell PCR method was developed, for a direct correlation of the obtained sequence data and the morphology of the selected cell. This was particularly important as all organisms examined were neither enriched nor pure cultured (besides *Beggiatoa alba*). In many former studies, phylogenetic analysis of large sulfur bacteria were hindered by the problem that obtained sequences could not be clearly assigned to the source organism or that sequence data could not be retrieved when generating clone libraries, especially when using bulk DNA extracts.

The novel, in-depth phylogenetic analysis of non-filamentous large sulfur bacteria revealed a major phylogenetic diversity and led to addressing the currently contradicting phylogeny among the large sulfur bacteria. Therefore, four additional geographic sites, which were inhabited by large sulfur bacteria, were sampled. Both their morphological and phylogenetic diversity was included in the survey, this time also involving filamentous types. Eventually, a profound taxonomic revision of the group was aspired, addressing disputable family and genus names and involving yet unclassified organisms.

Chapter 2

Dimorphism in methane seep-dwelling ecotypes of the largest known bacteria

Jake V. Bailey[a,b], **Verena Salman**[c], Gregory W. Rouse[d], Heide N. Schulz-Vogt[c], Lisa A. Levin[d], Victoria J. Orphan[b]

[a] Department of Geology and Geophysics, University of Minnesota – Twin Cities, 310 Pillsbury Drive SE, Minneapolis, MN 55455, USA

[b] Division of Geological and Planetary Sciences, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

[c] Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

[d] Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093, USA

Manuscript submitted for publication to The ISME Journal

Relevant contributions:

The concept of the study was developed by J. Bailey.

I designed specific primers and performed 16S rRNA gene sequencing and phylogenetic tree reconstruction.

I assisted in writing the manuscript, particularly regarding microbiological issues and the biology and classification of large sulfur bacteria.

Abstract

We present evidence for a dimorphic life cycle in the vacuolate sulfide-oxidizing bacteria that appears to involve the attachment of a spherical *Thiomargarita*-like cell to the exteriors of invertebrate integuments and other benthic substrates at methane seeps. The attached cell elongates to produce a stalk-like form before budding off spherical daughter cells resembling free-living *Thiomargarita* that are abundant in surrounding sulfidic seep sediments. The relationship between the attached parent cell and free-living daughter cell is reminiscent of the dimorphic life modes of the prosthecate Alphaproteobacteria, but on a grand scale, with individual elongate cells reaching nearly a millimeter in length. Abundant growth of attached *Thiomargarita*-like bacteria on the integuments of gastropods and other seep fauna provides not only a novel ecological niche for these giant bacteria, but also for animals that may benefit from epibiont colonization.

Introduction

The bacterial domain includes representatives with a variety of cell shapes (Young, 2006) and types of cell division (Angert, 2005). Bacterial cell morphologies can provide selective advantages ranging from increased motility, to decreased predation susceptibility (Young, 2006). The enormous size of the sulfide-oxidizing Gamma-proteobacterium, *Thiomargarita namibiensis*, is thought to provide it with the ability to store nitrate for months, persisting in environments where availability of the primary terminal electron acceptor oxygen, or the alternative acceptor nitrate, can be temporally and spatially variable (Schulz, 2006; Schulz and Jørgensen, 2001). However, the influence of cell morphology on the ecophysiology of these giant sulfur bacteria may not be limited to their large size.

Thiomargarita was first described in 1999 from diatom-rich oozes in the Benguela upwelling system off West Africa (Schulz *et al*., 1999). The originally described species *T. namibiensis* occurs primarily as linear chains of spherical cells enclosed in a mucous sheath (Schulz *et al*., 1999), though other morphotypes have since been described (Gutzke 2007; Schulz, 2006). In 2005, *Thiomargarita*-like bacteria that undergo reductive cell division resulting in symmetrical clusters of up to 16 cells were discovered at hydrocarbon seeps from the Gulf of Mexico (Kalanetra *et al.*, 2005; Bailey *et al*., 2007). Other *Thiomargarita*-like bacteria have also been described from mud volcanoes in the Barents Sea (de Beer *et al.*, 2006) and the Eastern Mediterranean Sea (Girnth *et al*., 2011). During a 2009 cruise aboard the R/V *Atlantis,* we observed a population of *Thiomargarita*-like cells in cores and on carbonate rock samples collected with the DSV *Alvin* at methane seeps associated with fault scarps, landslides and mud mounds (Bohrmann *et al.*, 2002) at water depths of ~990-1600 m along the Pacific margin of Costa Rica. Within the *Thiomargarita*-like population present at Costa Rica margin seeps are canonical free-living, solitary and reductively-dividing forms, as well as cells exhibiting an attached, presumptive dimorphic life cycle that we describe here. Similar populations of free-living and attached *Thiomargarita*-like cells were observed in Hydrate Ridge (off Oregon, USA) seep settings in 2010. The observation of dimorphic forms at multiple localities suggests that a dynamic morphology plays an important role in the ecology of vacuolate sulfur bacteria in seep settings.

Methods

Sample collection and microscopy

Samples were obtained during a research expedition to the Pacific margin of Costa Rica aboard the R/V *Atlantis* in Feb.-Mar. 2009 (AT15-44) and during a follow-up cruise to the same sites in Jan. 2010 (AT15-59). Push cores, samples of authigenic carbonate and associated macro- and megafaunal animals were collected from methane seeps associated with mud mounds and fault scarps at Mound 11 (1000-1025 m, 8^o 55.4'N, 84^o 18.24'W), Mound 12 (990-1100 m, 8^o 55.8'N, 84^o 18.75'W), Jaco Scarp (1775-1850 m, 9^o 7.00'N 84^o 50.51'W), and Mound Quepos (1030-1179 m, 8^o 58.09'N, 84^o 37.95'W) using the DSV *Alvin*. Additional collections of living gastropods with attached sulfur bacteria were made during AT 15-68 at Hydrate Ridge North (44^o 40 N 125 5.8 W) and Hydrate Ridge South (44^o 34.1 N 125^o 9.1 W).

During the research expedition, a Leica S8 APO stereo-microscope coupled with a Canon Powershot 9 was used to observe and image freshly collected free-living cells, as well as those cells attached to byssal threads, shells and other benthic substrates. Cells that were budding at the time of collection were followed for the ten-day duration of the cruise and imaged on a daily basis in order to document bud detachment. In the laboratory, an Olympus SZX16 stereo-microscope and an Olympus BX51 compound microscope equipped with an Olympus DP72 color camera were used to obtain higher resolution images of these cells and attachment substrates.

Electron microscopy

Gastropods covered with attached *Thiomargarita*-like sulfur bacteria were preserved in Karnovsky's fixative (Karnovsky, 1965) before undergoing ethanol series dehydration (25, 50, 75, 90, 99%) in artificial seawater/ethanol mixtures for 30 minutes each. The bacteria-covered gastropod shells were then dried in a Tousimis Model 780A Critical-Point Dryer before being imaged uncoated with a Hitachi TM-1000 tabletop scanning electron microscope.

For TEM imaging, samples were fixed with Karnosky's fixative solution and after three times washing with artificial seawater, the samples were post-fixed with 1% OsO_4 for 1 hour. Following three times rinsing with H_2O, the samples were dehydrated in a

graded series of ethanol (see above). The samples were embedded in Epon 812 resin. Ultrathin sections (65 nm) were produced using a Reichert UltraCut S Ultramicrotome and stained with uranyl acetate and lead citrate before being examined with a JEOL 1200EX electron microscope.

16S rRNA gene sequencing

Free-living *Thiomargarita*-like cells from Costa Rica and Namibia were removed from the sediment (Salman *et al.*, *in press*). The novel attached cells from Costa Rica were carefully scraped off byssal threads with two needles. Individual cells, i.e. free-living cells and formerly attached cells sometimes showing distal budding, were washed in artificial seawater, photographed under the microscope and subsequently used for performing single-cell 16S rRNA PCR methods described elsewhere (Salman *et al*., *in press*). The specific primer for *Thiomargarita*-related organisms, VSO233F, 5'-CCTA TGCCGGATTAGCTTG-3' (*E. coli* 16S rRNA position 233-251, Salman *et al.*, *in press*), was combined with the universal primer ITSeubR, 5´-GCCAAGGC ATCCACC-3´ (*E. coli* 23S rRNA position 23-38, Cardinale *et al*., 2004). Some attached cells did not reveal any PCR product with these primers, thus another specific forward primer was developed based on *Thiomargarita* sequences from the public databases: VSO360F, 5´-GGAATATTGGACAATGGGC-3´ (*E. coli* 16S rRNA positions 361-379, developed in this study), which was used in combination with primer VSO1300R, 5´-ATCCGGACTACGAGTAG-3´ (*E. coli* 16S rRNA position 1291-1307 (Salman *et al*., *in press*).

The amplification products were analyzed by agarose gel electrophoresis, bands of ca. 1750 bp length (and ca. 940 bp length with the other primer pair) were excised and DNA was purified with a gel extraction kit (Qiagen, Germany). PCR products were sequenced directly using the Big Dye Cycle Sequencing Kit and analyzed on the ABI Prism Genetic Analyzer 3130x (Applied Biosystems, USA). The seven 16S RNA *Thiomargarita*-affiliated sequences have been submitted to the DDBJ/EMBL/GenBank databases. The accession numbers are FN811658 – FN811664. The ITS region (spacer region between 16S rRNA and 23S rRNA as part of the PCR product when using primer ITSReub) was removed before further analysis of sequences.

Phylogenetic tree reconstruction

The phylogenetic affiliation of the retrieved sequences was inferred with the ARB software package (Ludwig *et al.*, 2004) based on release 95 of the SILVA SSURef database (Pruesse *et al*., 2007). The alignment of all sequences was automatically performed with the integrated aligner and checked manually afterwards, taking into account also the secondary structure of the 16S rRNA molecule. For tree reconstruction, a total of 90 nearly full-length sequences including representatives of the Gamma- and Deltaproteobacteria were used. Trees were calculated including nucleotide positions between 279 and 1463 (according to *E. coli* numbering) and performing neighbor-joining, maximum parsimony and maximum likelihood (RAxML) analyses, each with the application of 0%, 30% and 50% positional conservatory filters excluding highly variable regions. Partial sequences (AY632420, AF129012, FN811664) were inserted into each of the reconstructed trees by parsimony criteria without allowing changes in the overall tree topology. A consensus tree was manually produced based on the maximum likelihood tree calculated with a 30% filter using an approach described in (Peplies *et al.*, 2008). The branching patterns of all trees were compared and multifurcations were introduced at nodes where the exact tree topology could not be clearly resolved (Peplies *et al.*, 2008). The tree presented in Figure 5 shows a subset of the sequences originally used for calculation. Neighbor-joining and maximum likelihood trees are included in the supplement for comparison.

Results and Discussion

Occurrence of *Thiomargarita*-like bacteria

Free-living *Thiomargarita*-like cells were observed in sulfidic sediments at several mud mounds and associated methane seeps (Fig. 1A). As in sediments from off Namibia (Schulz *et al.*, 1999), and in the Gulf of Mexico (Kalanetra *et al.*, 2005), free-living sulfur bacteria off Costa Rica resembling *Thiomargarita* were most abundant in white and gray mat associated sediments in the uppermost 3 cm beneath the sediment/water interface. At some localities, cell populations consisted primarily of single *Thiomargarita*-like cells mixed with large *Beggiatoa*-like filaments (Fig. 1A), while at other locations, symmetrical cell clusters (not shown) predominated in gray flocculant material at the seabed. These cell clusters contained 2, 4, 8, and 16 cells, similar to those produced by reductive division in *Thiomargarita* sp. from the Gulf of Mexico

(Kalanetra *et al.*, 2005). In the vicinity of these sediments rich in *Thiomargarita*-like cells, thin films of microbial biomass were found attached to methane-derived authigenic carbonates (Fig. 1B), byssal threads of the mussel *Bathymodiolus* spp. (Fig. 1C), fragments of organic detritus, and most commonly, the integuments of seep-endemic invertebrates (Fig. 2A-C).

Microscopic examination of this benthic biofilm material revealed densely-packed attached vacuolated cells that closely resemble *Thiomargarita*, but with most cells elongated into a stalk-like form. Individual cells occasionally reached nearly 700 µm in length, but average elongate cells were 338 ± 169 µm long (Fig. S3). These elongate cells were observed to bud spherical *Thiomargarita*-like cells from their unattached pole (Fig. 3, 4). The attached poles of the elongate cells ranged from rounded (Fig. 3, 4) to a heavily tapered, and sometimes extruded shape, forming a short appendage (Fig. 4C). Both the attached elongate cells and the buds consisted of a single thin layer of sulfur globule-containing cytoplasm surrounding a single large vacuole (Fig. 3D and TEM observations not shown) – a basic cell structure that they share with all previously described *Thiomargarita* ecotypes (Kalanetra *et al*., 2005; Schulz, 2006; Schulz *et al*., 1999). Elongate morphotypes were not observed in the unattached free-living state. We hypothesize that the attached, budding types, and at least some of the free-living cells, represent not different species, but different stages of a reproductive developmental cycle. We evaluated this hypothesis by directly observing the budding forms after collection.

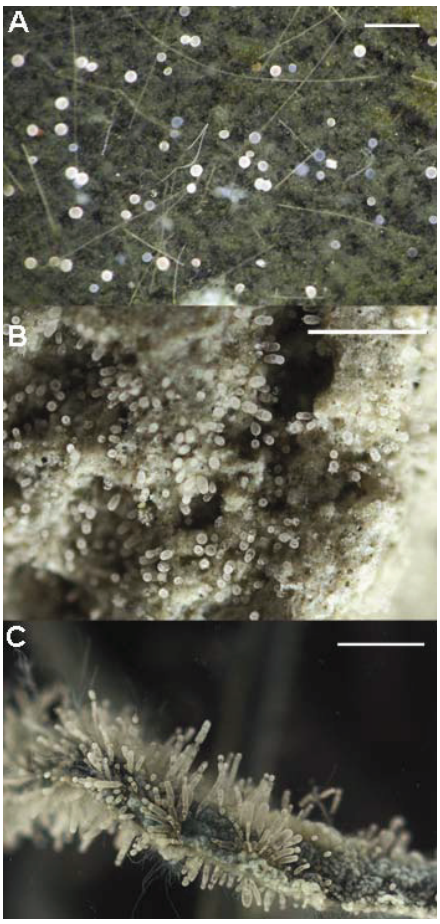


Figure 1. Vacuolate sulfide-oxidizing bacteria are abundant on a variety of substrates along the Costa Rica Margin including: (A) free-living *Thiomargarita*-like cells and large *Beggiatoa*-like filaments inhabiting sulfidic sediments; (B) slightly-elongate *Thiomargarita*-like morphotypes attached to authigenic carbonates; and (C) elongate *Thiomargarita*-like cells attached to *Bathymodiolus* mussel byssal threads. Scale bars represent 500 µm.

Furthermore, we analyzed the 16S ribosomal RNA gene sequences of the new, attached morphotypes, as well as of the free-living canonical morphotypes. Evidence for a developmental relationship between the attached cells and free-living spherical cells was provided by direct observation of buds formed at the time of collection completely detaching from budding parent cells at a rate of approximately one detachment in three hundred budding cells over a 24-hour period (12 in approximately 3,600 individuals observed), demonstrating that the budded daughter cells are indeed being released into the surrounding environment (Fig. 3B, C, F).

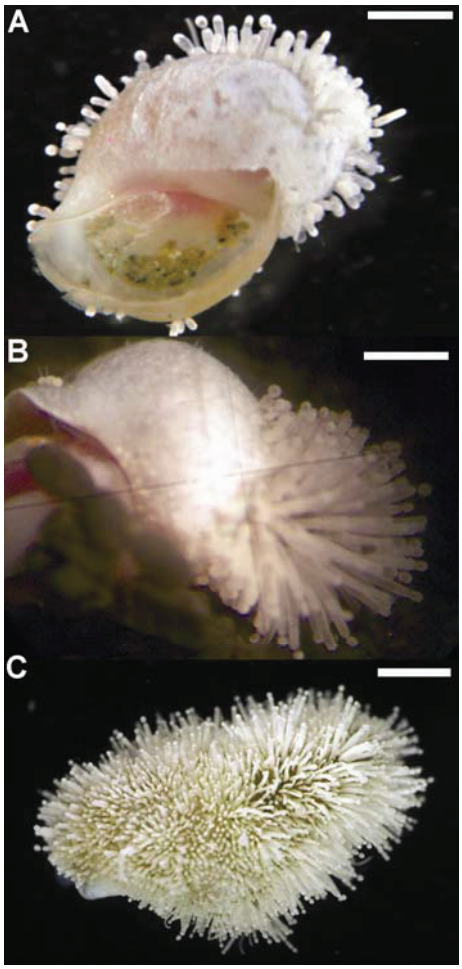


Figure 2. The shells of living gastropods (*Provanna laevis*) from the Costa Rica margin and the Cascadia margin (Hydrate Ridge) are colonized by attached, elongate *Thiomargarita*-like cells. Scale bars represent approximately 1 mm.

Fully-detached, budded daughter cells were morphologically identical to free-living, spherical sulfur bacteria in surrounding sediments and both have similar cell diameters; 71 ± 22 µm, n = 40 vs. 88 ± 14 µm, n = 37. The morphological similarity of the released buds to the free-living *Thiomargarita*-like cells in the surrounding sediments suggests the possibility that at least some of the latter population results from the released spherical buds. However, morphology can be misleading, particularly with respect to the phylogenetic identification of vacuolated sulfur bacteria. For example, marine filaments with a *Beggiatoa*-like morphology, that have similar filament diameters and were collected from the same locality, can be quite divergent with respect to their phylogenetic positions (Mußmann *et al.*, 2007; Teske *et al.*, 1999). The observation that the attached elongate cells are found on hard substrates that are physically separated by sediments that contain spherical, free-living cells, and that elongate cells are never observed in the free-living populations, strongly suggests that at least some of the free-living population includes the spherical offspring of the attached morphotypes. Additionally, the assertion that the life cycle is initiated by the attachment of spherical cells is supported by the observation of attached spherical cells that are just

beginning elongation (Figure S4), along with the spectrum of elongation morphologies from sphere, to prolate spheroid, to increasingly longer cylindrical cells occurring in close proximity (Fig. 4E). However, the percentage of free-living cells that represent the progeny of attached budding forms remains unclear.

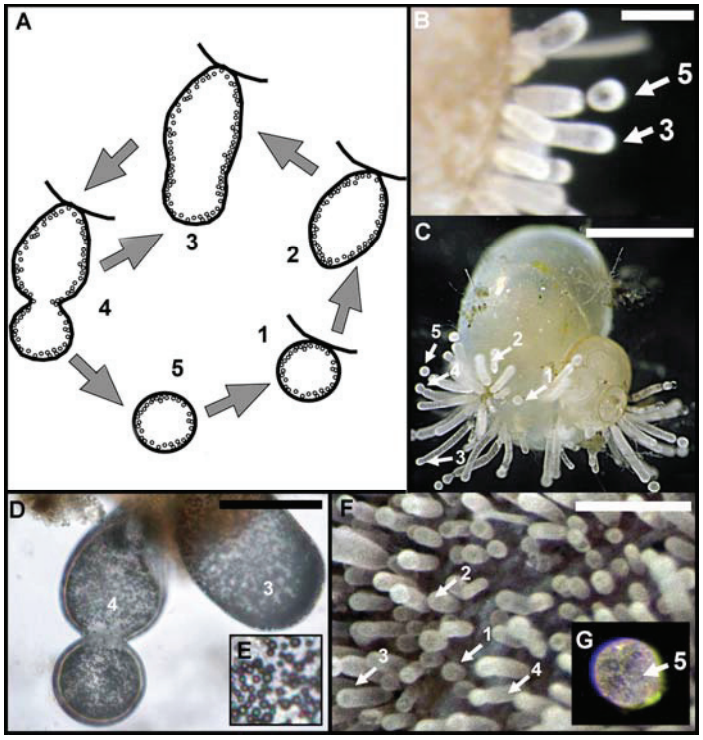


Figure 3. An apparent reproductive developmental cycle in attached sulfur bacteria (A) appears to begin with the attachment of a spherical cell to a solid substrate (arrows 1), followed by elongation to a stalked form (arrows 2), initial bud formation (arrows 3) and finally budding (arrows 4) to produce a spherical daughter cell (arrows 5). (B, C) Attached sulfur bacteria on a gastropod shell. (D, E) Detail of attached cell and intracellular sulfur globules. (F) Bacteria attached to mussel byssal threads. (G) Budded *Thiomargarita*-l ike cell was observed to undergo complete detachment from the mother cell between 12 and 20 hours after collection. After bud detachment, the elongate mother cell (A4 → 3) can continue to produce additional buds. Scale bar in B = 350 µm; C = 1 mm; D = 200 µm; E = 300 µm.

Sequence comparisons of individual cells indicated that Costa Rica margin seeps harbor not just one, but at least three distinct phylogenetic clusters. The relatedness of the organisms in this study to *Thiomargarita namibiensis* and *Thiomargarita* sp. from the Gulf of Mexico is apparent from the phylogenetic tree produced by sequencing individual representatives from Costa Rica and additional specimens from Namibia (Fig. 5). The sequences published prior to this study constitute the closest relatives to at least five sequences obtained in this study. While some sequences of free-living and attached morphotypes are closely related to previously published *Thiomargarita* spp. (Kalanetra *et al*., 2005, Schulz *et al*., 1999), two sequences (COS001 and COS016) each formed a separate branch that did not cluster within the *Thiomargarita* clade. The paraphyletic relationship of the retrieved 16S rRNA gene sequences surely complicates

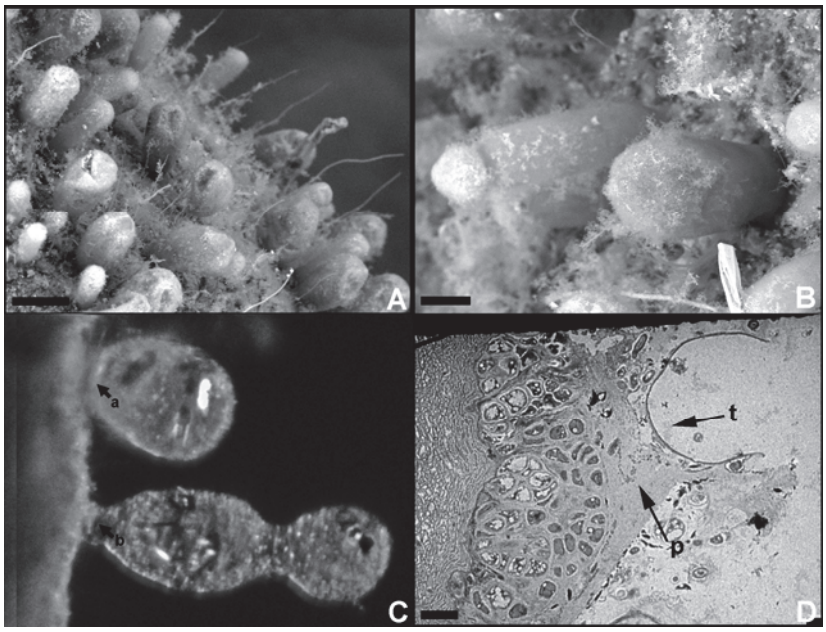


Figure 4. Attached sulfur bacteria. (A, B) Densely-packed *Thiomargarita*-like cells on gastropod shell co-occur with attached filamentous bacteria. Scale bar in A = 30 μm; B = 20 μm. (C) Particularly during early stages of attachment and elongation they have rounded bases where they are attached to benthic substrates. As the cell elongates, the base of the cells becomes tapered. In some specimens, a small appendage formed by extrusion of the cell ultrastructure is present. (D) Transmission electron micrograph of vacuolated *Thiomargarita*-like cells (t) attached to plaque (p) and mucous occupied by smaller filamentous bacteria. This material surrounds the *Bathymodiolus* mussel byssus. Scale bar = 2 μm.

identification of cells based on morphology. Nomenclature for *Thiomargarita*-like cells and other vacuolate sulfur bacteria can be confusing, as some canonical genera include paraphyletic taxa. However, the taxonomy of the colorless sulfur bacteria, including the novel morphotypes identified in this study, was recently revisited by Salman *et al*. (*in press*). Sequences in the three clusters represent either attached, free-living, or paired free-living/attached morphotypes (Fig. 5). Some sequences showed >99% similarity between attached and free-living morphotypes (Table 1) that may represent different developmental stages of the same organism, as the <1% difference is within the range of experimental error. Alternatively, these cells could represent closely related phylotypes that superficially resemble parts of the putative dimorphic life cycle, but do not represent the direct progeny of the attached cells. The disparity could be caused by natural variability within the seep population coupled with the relatively low number of sequenced specimens, and the inherent limitations of sampling a deepwater site via submersible, thereby possibly missing directly corresponding mother and daughter cells. It is also unclear if the reductively-dividing free-living sulfur bacteria present in nearby sediments exhibit attachment or dimorphism as part of their life cycle(s).

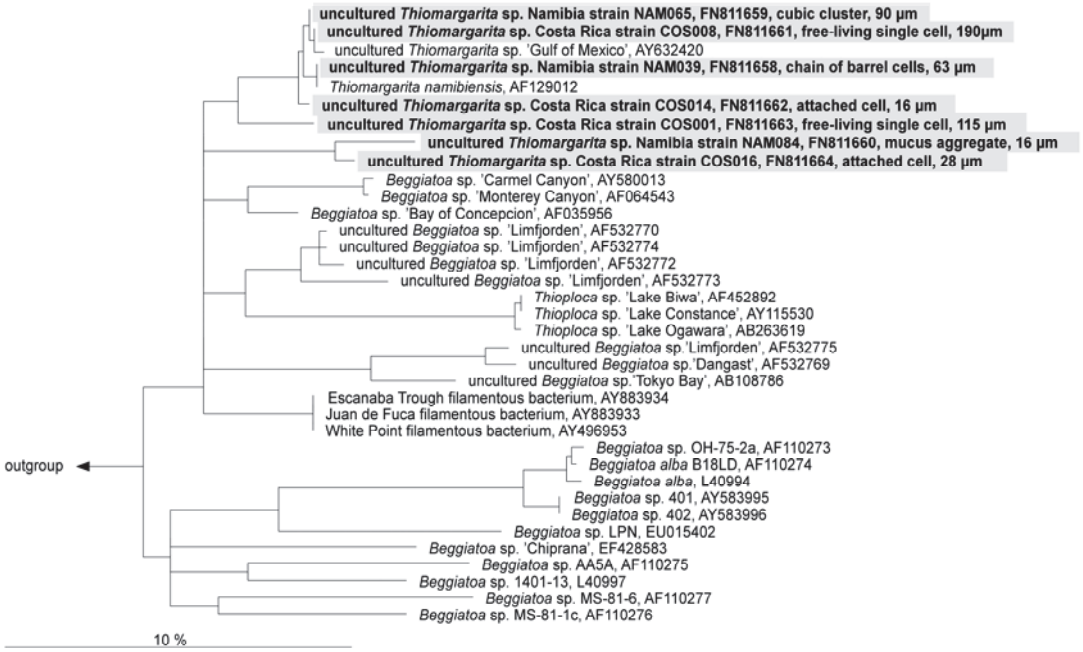


Figure 5. Phylogenetic tree of gammaproteobacterial 16S rRNA gene sequences showing sequences of different morphotypes from Costa Rica (COS) and Namibia (NAM) in gray and their affiliation to previously published *Thiomargarita* spp. These sequences are shown within the context of the broader clade of conspicuous sulfide-oxidizing Gammaproteobacteria. Of each newly sequenced cell the morphology and diameter is indicated. In attached forms, the widest diameter of the attached basal body is reported. The scale bar represents approximately 10% sequence divergence.

Table 1. Levels of 16S rDNA sequence identities of nearly full-length sequences from *Thiomargarita*-like cells from the Costa Rica margin (COS) and Namibia (NAM) along with partial sequences of the formerly described *Thiomargarita namibiensis* (*T. nam*.) and *Thiomargarita* sp. Gulf of Mexico (*T*. GOM) as reference species.

	NAM065	COS008	*T*. GOM	NAM039	*T. nam.*	COS014	COS001	NAM084
NAM065								
COS008	99.8							
T. GOM	99.7	99.8						
NAM039	99.5	99.2	98.8					
T. nam.	99.1	98.7	99.1	100				
COS014	99.4	99.4	99.1	99.3	98.9			
COS001	96.6	96.5	95.7	96.4	94.9	96.8		
NAM084	95.3	95.3	92.4	95.3	92.4	95.2	94.4	
COS016	95.1	95.1	95.1	95	95	94.8	93.8	99.1

Ecological Significance of Dimorphism

Reproductive dimorphism is known in the Alphaproteobacteria, such as *Caulobacter crescentus*, whose parent cells are stalked, and the daughter cells are vibroid – but in those organisms the daughter cells are actively motile (Brun and Janakiraman, 2000; Poindexter, 2006). In the sulfur bacteria detected here, the daughter cells are not yet known to be actively motile – though an inconspicuous sporadic slow rolling motility has been observed in different, spherical sulfur bacteria from Namibia (Salman *et al*., *in press*). Additionally, free-living daughter cells can be readily transported by bottom currents. Broken, transported and re-deposited clasts of methane-derived authigenic carbonate edifices observed on the collection cruises, as well as evidence for local sediment winnowing, record the presence of such currents at the sites where the attached *Thiomargarita*-like morphotypes occur. Upon detachment, buds may also attach to substrates directly adjacent to the mother cell. This process is known to occur in some prosthecate Alphaproteobacteria and would explain the observation that mats of attached *Thiomargarita*-like sulfur bacteria are often densely packed and frequently exhibit more than one developmental stage on the same substrate (Fig. 3B). Alternatively, certain cells in the mat may simply have slower elongation and budding rates than others.

The factors responsible for initiating attachment, elongation and budding in these bacteria are not fully understood because *Thiomargarita* and its vacuolated relatives have yet to be cultivated in the laboratory. However, comparisons with other organisms that exhibit similar behaviors may provide clues. Attachment to solid substrates is also known in other Gammaproteobacteria, including the closely related marine, vacuolated *Thiothrix*-like bacteria (Kalanetra *et al.*, 2004, Kalanetra and Nelson 2010). Indeed, attached filamentous *Thiothrix*-like filaments sometimes co-occur with the attached *Thiomargarita*-like cells described here. Attachment provides stabilization of a parent cell under conditions where it can produce daughter cells for dispersal to other potential habitats via bottom currents. Presumably, stabilization of reproductively-active sulfide-oxidizing cells over zones of sulfide enrichment would be more important in advective seep habitats where sulfide generated by the anaerobic oxidation of methane or vented directly from subsurface fluids is more localized than the sulfide generated by sulfate reduction in upwelling regions (Linke *et al.*, 2005).

Budding is known in *Thiothrix*-like filaments (Kalanetra *et al*., 2004), in certain low-GC gram-positive bacteria, Cyanobacteria, Planctomycetes, and prosthecate Alpha-proteobacteria (refs. in Angert, 2005) among others. In prosthecate bacteria, it is the stalked parent cell that undergoes budding, suggesting that perhaps elongation enhances dispersal of daughter cells. However, elongation is apparently not a prerequisite to budding in attached *Thiomargarita*-like cells, as division of attached but not elongated cells, or slightly elongated cells, was observed on a handful of substrates (Fig. 1B). Access to scarce metabolites may instead drive elongation in attached sulfur bacteria. Cell elongation in *Caulobacter crescentus* is thought to enhance the uptake of phosphorous from the environment by increasing the cell's surface area-to-volume ratio (Poindexter, 1984). *Thiomargarita namibiensis*, which lives in sediments beneath upwelling zones, sequesters phosphorous as intracellular polyphosphate granules (Schulz and Schulz, 2005). In seep settings, phosphorous is less available than in an upwelling environment (Joye *et al.*, 2004; Linke *et al.*, 2005). Thus, cell elongation in seep-dwelling sulfur bacteria may provide for enhanced phosphorous uptake as it does in the Alphaproteobacteria.

Nutrient uptake and ATP synthesis scales with cell surface area, and *Thiomargarita* functions at a relatively low level of metabolic power per gene (Lane and Martin, 2010). An elongate cell has a higher surface area-to-volume ratio than a spherical cell of the same volume (Young, 2006). For example, if a cell was assumed to be a cylinder capped with hemispherical ends, 300 μm in length and 80 μm in diameter, the cell would have a 38.5% greater surface area-to volume ratio (0.054) over a sphere with the same volume (0.039). The larger a *Thiomargarita* cell grows, the larger its potential store of nitrate, but at the cost of decreasing its surface area-to volume ratio. Growth of a cell into an elongate form, as opposed to simply expanding the sphere volume, minimizes the decrease in surface area-to volume ratio that accompanies growth. Reductive cell division, which is observed in free-living *Thiomargarita*-like cells at this site, also effectively increases surface area while maintaining a constant nitrate-storage volume, albeit by producing smaller cells.

The elongation of budding cells may also allow attached cells greater access to the terminal electron acceptors, oxygen and nitrate. It is not known if attached and free-living reductively-dividing *Thiomargarita*-like bacteria represent different stages within the same life cycle. However, one possibility is that elongation increases surface area in

attached cells that have access to nitrate and can readily enlarge, while reductive division apportions stored nitrate into smaller, easily dispersible cells with higher surface-area-volume ratios when nitrate availability is limited. Closely related vacuolate sulfur bacteria resembling *Beggiatoa* and *Thioploca* can actively move along gradients between sulfide and terminal electron acceptors (Huettel *et al.*, 1996; Nelson *et al.*, 1986). In seep settings, alternating conditions of access to electron donor and acceptor can be introduced by brine flows (Girnth *et al*., 2011), changing currents, and heterogenous pore fluid advection, exposing populations of sulfur bacteria to alternating conditions of high sulfide, nitrate and oxygen fluxes. Elongation in sessile bacteria may provide cells with a larger surface area for more efficient uptake of spatially segregated metabolites, under the fluctuating and heterogeneous fluxes of O_2, NO_3^-, and H_2S measured in Costa Rica margin sediments (Linke *et al.*, 2005).

In some freshly collected samples, elongate cells possess a visibly marked increase in sulfur globules at the distal end of the cell. The apical, sulfur-enriched region is the site of bud-formation. These sulfur resources might be incorporated into the budding offspring, providing an initial energy reserve for the free-living state. Alternatively, the irregular distribution of sulfur globules may be driven by the local geochemical environment, where the observed sulfur accumulation at the apical end of the cell is linked to close contact with either oxygenated or sulfidic waters, depending upon the orientation of the cell. However, this explanation seems less likely – as cells attached to opposite sides of byssal threads also showed apical sulfur enrichment in cells oriented in opposite directions.

Metazoan-Bacterial Interactions

The attachment of newly detected sulfur bacteria to benthic substrates also has potentially important implications for metazoan ecology at seeps. Elongate sulfur bacteria were found attached to byssal threads of *Bathymodiolus* spp. (Mytilidae, Bivalvia), the carapace of a Yeti crab specimen (*Kiwa* sp.), the shells of several species of gastropod including *Provanna laevis, Hyalogyrina* sp*., Pyropelta* spp*.,* and the limpet *Lepetodrilus* sp*.* These gastropod shells were commonly covered with thick mats of attached elongate sulfur bacteria (Fig. 2). Interestingly, bacteria-covered *Provanna* collected from Hydrate Ridge in 2010 exhibited an unusual behavior that may benefit the attached sulfide-oxidizing bacteria. When placed in a container with seawater overlying sulfidic sediments, they were observed to either climb to the uppermost

portion of the jar and maintain a position adjacent to the small air headspace (n = 4), or actively dig a burrow into the sulfidic sediments (n = 8) and subsequently inverting themselves so that their shell and attached bacteria were positioned within the freshly dug burrow (Fig. 6). Gastropods removed from the burrows with forceps repeatedly re-established their burrows and inverted positions (n = 6). Epibiont-bearing invertebrates are known to migrate between oxygenated and sulfidic sediment pore waters as a means of supplying symbionts with the metabolic substrates required for sulfide oxidation (e.g., Ott *et al.*, 1991), and a similar behavior may be on display here.

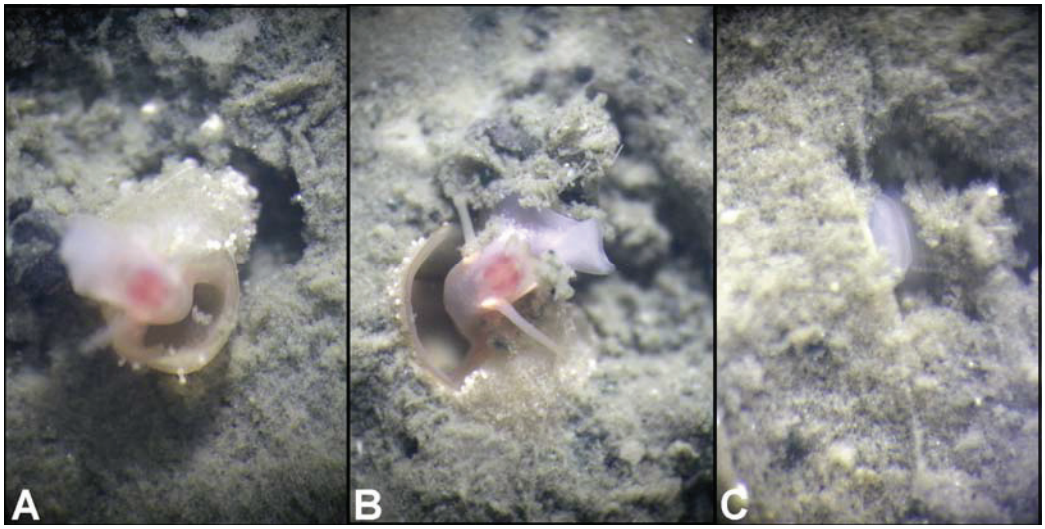


Figure 6. The gastropod *Provanna laevis* burrows into sulfidic mud in a collection container, (A) first digging a burrow, (B) then occupying the burrow with its carapace oriented downward, which exposes the population of attached sulfur bacteria to sulfidic conditions. (C) The head of the gastropod remains partially exposed above the sediment/water interface.

However, this behavior and association with seep invertebrates is apparently not a requirement for growth of the attached *Thiomargarita*-like cells, as these bacteria are sometimes found on stabilized mineral and detrital substrates such as carbonate rocks. The impetus for gastropods to exhibit this behavior that potentially exposes them to predation while lying inverted with their soft parts exposed remains unclear. The bacteria may reduce the animal's exposure to hydrogen sulfide as is thought to occur in other invertebrates (Alain *et al.*, 2002). These species of gastropods are not known to contain sulfide-oxidizing endobionts. Attached bacteria may also serve as food source for grazing gastropods, as occurs in some other animal/epibiont symbioses (e.g., Dattagupta *et al.*, 2009; Gillian and Dubilier, 2004; Zbinden and Cambon-Bonavita, 2003; Ott et al., 1991) although these gastropods are seemingly incapable of consuming the bacteria attached to their own shells.

At Costa Rica seep settings, attached sulfur bacteria are rare on rocks and shells with large surface areas. The substrates most well-colonized by the cells are those with low surface areas such as small shells and mussel byssal threads. When they do occur on larger substrates, they are found in small isolated patches or within small crevices. The apparent absence of attached bacteria on *Bathymodiolus* mussel shells*, Lamellibrachia* tubes, and most carbonate rocks suggests that they may be limited to surfaces that are not subject to intense shearing forces or grazing by gastropods. The shells and carapaces of the grazers themselves and cryptic habitats such as mussel byssal threads may offer a refuge from consumers. Alternatively (or additionally), the availability of sulfide, nitrate, and oxygen may make certain microenvironments more conducive to colonization by attached sulfur bacteria. High oxygen and nitrate fluxes, primarily driven by bacterial oxidation of sulfide (781 and 700 μmol cm^{-2} yr^{-1} respectively) have been reported at the collection sites, but these fluxes vary by up to three orders of magnitude over distances of tens-of centimeters as a result of heterogeneities in seep fluid flow (Linke *et al.*, 2005). Whether these attached bacteria show preferential attachment to specific substrates that guarantee access to sulfide or other metabolites, or if their distribution is influenced by the very abundant microbial grazers in the system remain open questions.

Examination of photographs and specimens of seep fauna collected during 2005 and 2006 cruises to the Eel River Basin and Hydrate Ridge (off the coast of California and Oregon, USA), and observations of gastropods collected from Hydrate Ridge in 2010 both show inhabitation by morphologically-similar elongate *Thiomargarita*-resembling cells (Fig. 2B), suggesting that certain ecotypes of attached sulfur bacteria are geographically widespread in seep settings. The difficulty in collecting these specimens and maintaining them under enrichment conditions that allow for observation of their development renders our understanding of these organisms incomplete. However, the dimorphic life cycle in seep-associated cells closely related to *Thiomargarita* spp. described here reveals a complex reproductive ecology for these giants of the bacterial world that further confounds the generalization that bacteria are small and simple.

Acknowledgements

Funding was provided by the Agouron Institute, the University of Minnesota Department of Geology and Geophysics, the Max Planck Society, and the National Science Foundation (OCE-0826254 and OCE 09-39557). Parts of this work were

carried out in the Characterization Facility, University of Minnesota, a member of the NSF-funded Materials Research Facilities Network (www.mrfn.org) via the MRSEC program with assistance from Chris Frethem and Fang Zhou. Scanning electron microscopy was performed in the University of Minnesota-Twin Cities Limnological Research Center with the support of Amy Myrbo. Furthermore, we thank Tony Rathburn, Shana Goffredi, Anders Waren, Danwei Huang, Anne Dekas, Olivia Mason, Ben Harrison, the crew of the R/V *Atlantis* and pilots of the DSV *Alvin (*15-44 and 15-59) for assistance at sea, and Ben Grupe for providing live gastropods from Hydrate Ridge. We thank the reviewers of this manuscript for valuable comments that helped improve the manuscript.

References

Alain K, Olagnon M, Desbruyères D, Pagé A, Barbier G, Juniper SK et al. (2002). Phylogenetic characterization of the bacterial assemblage associated with mucous secretions of the hydrothermal vent polychaete *Paralvinella palmiformis*. FEMS Microbiol Ecol 42: 463-476.

Angert ER. (2005). Alternatives to binary fission in bacteria. Nat Rev Microbiol 3: 214-224.

Bailey JV, Joye SB, Kalanetra KM, Flood BE, Corsetti FA. (2007). Evidence of giant sulphur bacteria in Neoproterozoic phosphorites. Nature 445: 198-201.

Bohrmann G, Heeschen K, Jung C, Weinrebe W, Baranov B, Cailleau B et al. (2002). Widespread fluid expulsion along the seafloor of the Costa Rica convergent margin. Terra Nova 14: 69-79.

Brun YV, Janakiraman R. (2000). The dimorphic life cycle of *Caulobacter* and stalked bacteria. In: Brun YV and Shimkets LJ (eds). Prokaryotic development. American Society for Microbiology: Washington, D.C. pp 297–317.

Cardinale M, Brusetti L, Quatrini P, Borin S, Puglia AM, Rizzi A et al. (2004). Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. Appl Environ Microbiol 70: 6147-6156.

Dattagupta S, Schaperdoth I, Montanari A, Mariani S, Kita N, Valley JW et al (2009). A novel symbiosis between chemoautotrophic bacteria and a freshwater cave amphipod. The ISME Journal 3: 935-943.

de Beer D, Sauter E, Niemann H, Kaul N, Foucher J-P, Witte U et al (2006). *In situ* fluxes and zonation of microbial activity in surface sediments of the Haakon Mosby Mud Volcano. Limnology and Oceanography 51: 1315-1331.

Gillian DC, Dubilier N (2004). Novel epibiotic *Thiothrix* bacterium on a marine amphipod. Applied and Environmental Microbiology 70: 3773-3775.

Girnth A-C, Grünke S, Lichtschlag A, Felden J, Knittel K, Wenzhöfer F et al. (2011). A novel, mat-forming *Thiomargarita* population associated with a sulfidic fluid flow from a deep-sea mud volcano. Environ Microbiol 13: 495-505.

Gutzke V. (2007). Morphologische und physiologische Untersuchungen an *Thiomargarita* spp. Diploma thesis, University of Hannover, Germany.

Huettel M, Forster S, Klöser S, Fossing H. (1996). Vertical migration in the sediment-dwelling sulfur bacteria *Thioploca* spp. in overcoming diffusion limitations. Applied and Environ Microbiol 62: 1863– 1872.

Joye SB, Boetius A, Orcutt BN, Montoya JP, Schulz HN, Erickson MJ et al. (2004). The anaerobic oxidation of methane and sulfate reduction in sediments from Gulf of Mexico cold seeps. Chem Geol 205: 219-238.

Kalanetra KM, Huston SL, Nelson DC. (2004). Novel, attached, sulfur-oxidizing bacteria at shallow hydrothermal vents possess vacuoles not involved in respiratory nitrate accumulation. Appl Environ Microbiol 70: 7487–7496.

Kalanetra KM, Joye SB, Sunseri NR, Nelson DC. (2005). Novel vacuolate sulfur bacteria from the Gulf of Mexico reproduce by reductive division in three dimensions. Environ Microbiol 7: 1451-1460.

Kalanetra KM, Nelson DC. (2010). Vacuolate-attached filaments: highly productive *Ridgeia piscesae* epibionts at the Juan de Fuca hydrothermal vents. Mar Biol 4: 791-800.

Karnovsky MJ. (1965). A formaldehyde-glutaraldehyde fixative of high osmolarity for use in electron microscopy. J Cell Biol 27: 137.

Lane N, Martin W. (2010). The energetics of genome complexity. Nature 467: 929-934.

Linke P, Wallmann K, Suess E, Hensen C, Rehder G. (2005). *In situ* benthic fluxes from an intermittently active mud volcano at the Costa Rica convergent margin. Earth Planet Sci Lett 235: 79-95.

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

Mußmann M, Hu FZ, Richter M, de Beer D, Preisler A, Jørgensen B et al. (2007). Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. PLoS Biol 5: 1923-1937.

Nelson DC, Revsbech NP, Jørgensen BB. (1986). Microoxic-anoxic niche of Beggiatoa spp.: a microelectrode survey of marine and freshwater strains. Appl Environ Microbiol 52: 161-168.

Ott JA, Novak R, Schiemer F, Hentschel U, Nebelsick M et al. (1991). Tackling the sulfide gradient: a novel strategy involving marine nematodes and chemoautotrophic ectosymbionts. Mar Ecol 12: 261-279.

Peplies J, Kottmann R, Ludwig W, Glöckner FO. (2008). A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. Syst Appl Microbiol 31: 251-257.

Poindexter JS. (1984). The role of calcium in stalk development and in phosphate acquisition in *Caulobacter crescentus*. Arch Microbiol 138: 140-152.

Poindexter JS. (2006). Dimorphic prosthecate bacteria: The genera *Caulobacter*, *Asticcacaulis*, *Hyphomicrobium*, *Pedomicrobium*, *Hyphomonas*, and *Thiodendron*. In: Dworkin M (ed). The Prokaryotes. Springer-Verlag: New York. pp 72-90.

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

Salman V, Amann R, Girnth A-C, Polerecky L, Bailey J, Høgslund S et al. (2011). A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. Syst Appl Microbiol (in press).

Schulz HN. (2006). The genus *Thiomargarita*. In: Dworkin M (ed). The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community. Springer-Verlag: New York.

Schulz HN, Brinkhoff T, Ferdelman T, Marine MH, Teske A, Jørgensen BB. (1999). Dense populations of a giant sulfur bacterium in Namibian shelf sediments. Science 284: 493-495.

Schulz HN, Jørgensen BB. (2001). Big bacteria. Annu Rev Microbiol. 55: 2001.

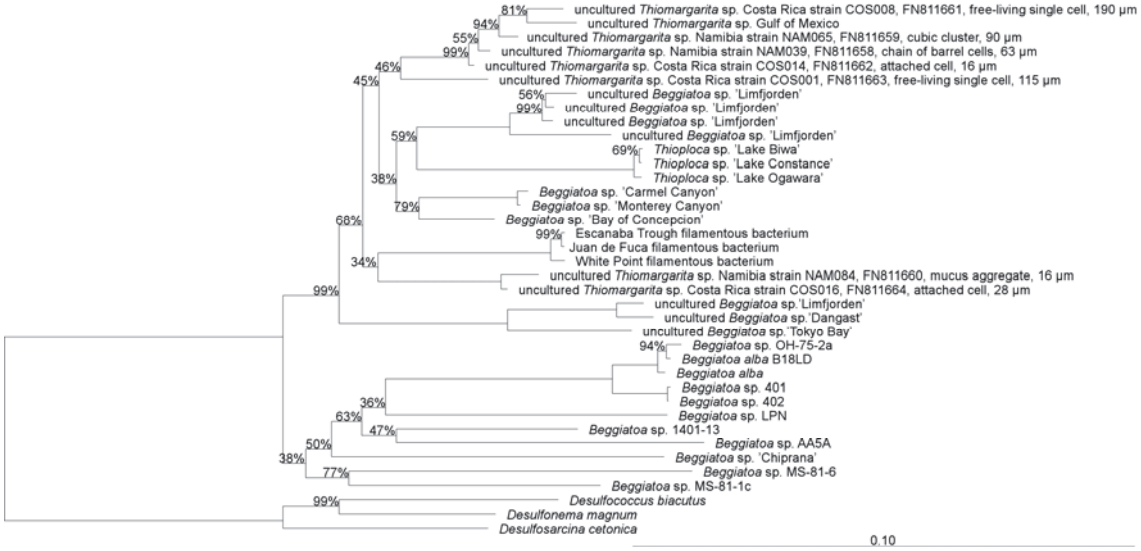
Schulz HN, Schulz HD. (2005). Large sulfur bacteria and the formation of phosphorite. Science 307: 416-418.

Teske A, Sogin ML, Nielsen LP, Jannasch HW. (1999). Phylogenetic position of a large marine *Beggiatoa*. Syst Appl Microbiol 22: 39-44.

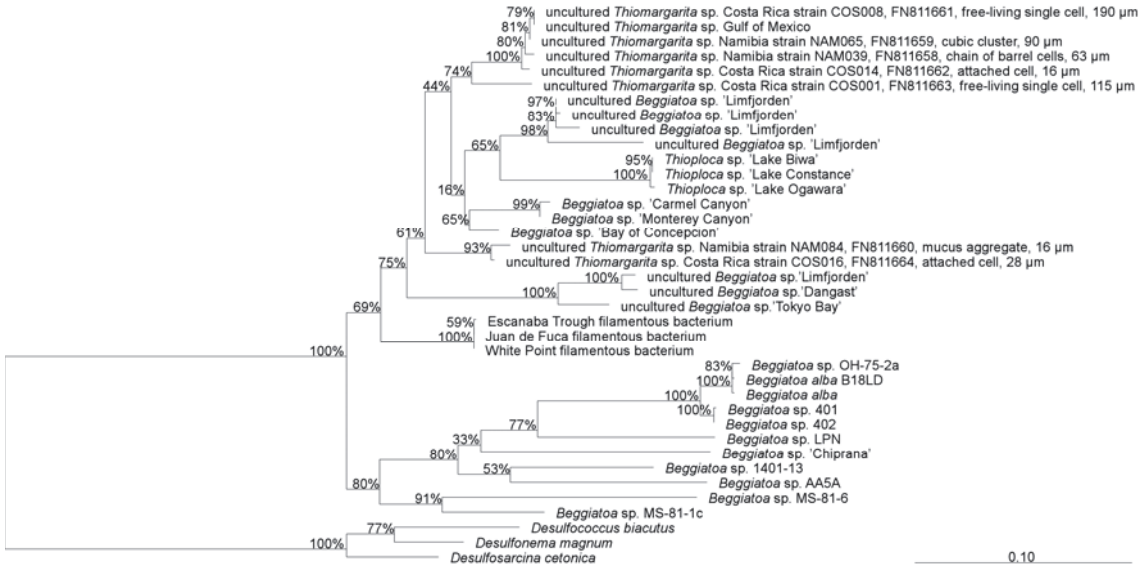
Young KD. (2006). The selective value of bacterial shape. Microbiol Mol Biol Rev. 70: 660-703.

Zbinden M, Cambon-Bonavita MA (2003). Occurrence of *Deferribacterales* and *Entomoplasmatales* in the deep- sea Alvinocarid shrimp *Rimicaris exoculata* gut. FEMS Microbiol Ecol 46: 23-30.

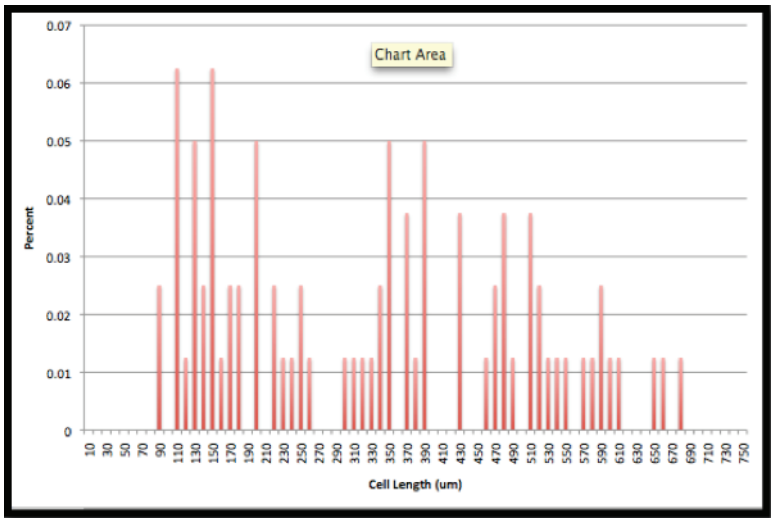
Supplemental Information



Supplemental Figure 1. Neighbor-joining phylogenetic tree of 16S gene sequences showing new *Thiomargarita* sequences from Costa Rica (COS) and Namibia (NAM) within the context of the broader clade of conspicuous sulfur-oxidizing gamma proteobacteria.



Supplemental Figure 2. Maximum likelihood phylogenetic tree of 16S gene sequences showing new *Thiomargarita* sequences from Costa Rica (COS) and Namibia (NAM) within the context of the broader clade of conspicuous sulfur-oxidizing gamma proteobacteria.



Supplemental Figure 3. Long-axis size distribution of budding elongate *Thiomargarita* cells attached to byssal threads.

Supplemental Figure 4. Spherical cells just beginning to elongate provide evidence that the initial attachment stage in a spherical cell. Scale bar = 100 µm.

Chapter 3

A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria

Verena Salman[a], Rudolf Amann[a], Anne-Christin Girnth[a], Lubos Polerecky[a], Jake V. Bailey[b], Signe Høgslund[c], Gerdhard Jessen[a], Silvio Pantoja[d] and Heide N. Schulz-Vogt[a]

[a] Max Planck Institute for Marine Microbiology, Celsiusstraße 1, D-28359 Bremen, Germany

[b] Department of Geology and Geophysics, University of Minnesota – Twin Cities, 310 Pillsbury Drive SE, MN 55455, USA

[c] Department of Biological Sciences, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus, Denmark

[d] Department of Oceanography and Center for Oceanographic Research in the Eastern South Pacific, University of Concepción, Casilla 160-C, Concepción, Chile

Manuscript in press at Systematic and Applied Microbiology;
doi: 10.1016/j.syapm.2011.02.001

Relevant contributions:

The concept of the study was developed together with H. Schulz-Vogt and R. Amann. I performed all morphological analyses, designed novel primers and developed the single-cell 16S rRNA-PCR. I performed the sequencing and phylogenetic tree reconstruction. The manuscript was written mainly in collaboration with H. Schulz-Vogt and R. Amann, but also with assistance of the other co-authors.

Abstract

The colorless, large sulfur bacteria are well known because of their intriguing appearance, size and abundance in sulfidic settings. Since their discovery in 1803 these bacteria have been classified according to their conspicuous morphology. However, in microbiology the use of morphological criteria alone to predict phylogenetic relatedness has frequently proven to be misleading. Recent sequencing of a number of 16S rRNA genes of large sulfur bacteria revealed frequent inconsistencies between the morphologically determined taxonomy of genera and the genetically derived classification. Nevertheless, newly described bacteria were classified based on their morphological properties, leading to polyphyletic taxa. We performed sequencing of 16S rRNA genes and internal transcribed spacer (ITS) regions, together with detailed morphological analysis of hand-picked individuals of novel non-filamentous as well as known filamentous large sulfur bacteria, including the hitherto only partially sequenced species *Thiomargarita namibiensis*, *Thioploca araucae* and *Thioploca chileae*. Based on 128 nearly full-length 16S rRNA-ITS sequences, we propose the retention of the family *Beggiatoaceae* for the genera closely related to *Beggiatoa*, as opposed to the recently suggested fusion of all colorless sulfur bacteria into one family, the *Thiotrichaceae*. Furthermore, we propose the addition of nine *Candidatus* species along with seven new *Candidatus* genera to the family *Beggiatoaceae*. The extended family *Beggiatoaceae* thus remains monophyletic and is phylogenetically clearly separated from other related families.

Introduction

The colorless sulfur bacteria include some of the largest known bacteria, such as the microbial giant *Thiomargarita namibiensis* [63]. The first representative of this extraordinary group, *Beggiatoa alba*, was discovered as early as 200 years ago [81] because of its large size and conspicuous, white appearance. Although a number of other species have been described and investigated, at present only three type strains of colorless, large sulfur bacteria exist in pure cultures (*B. alba* [69], *Thiothrix nivea* [42], *Leucothrix mucor* [21]). Vacuolated representatives have yet to be isolated in pure culture [27,52,56]. Furthermore, only a small number of full-length 16S rRNA gene sequences have been retrieved to date and for some ecotypes genetic data are unavailable or are represented by only partial sequences. Therefore, classification and identification based on morphological properties is continuously applied to newly discovered organisms of the colorless sulfur bacteria [e.g. 2,29,38,49,53,72]. Although it is generally accepted in the discipline of microbiology that this approach may be misleading, the taxonomy of the colorless sulfur bacteria has, until now, never been revised on the basis of phylogenetic data. Accordingly, the traditional nomenclature disagrees with the classification derived from the phylogenetic data, resulting in polyphyletic genera [1,2,53,72].

In the current edition of Bergey´s Manual of Systematic Bacteriology it was suggested that all colorless, large sulfur bacteria should be combined into the family *Thiotrichaceae* [17]. However, according to the International Code of Nomenclature of Bacteria [40], this recently proposed classification is illegitimate because multiple type genera were combined into one family [14], which is furthermore not monophyletic. In fact, all afore published Bergey's Manuals [44,71, citing only the most recent] and the Approved Lists of Bacterial Names [66, 67] use the family name *Beggiatoaceae*, which includes those bacteria closely related to *Beggiatoa*, but excludes more distantly related genera like *Leucothrix* and *Achromatium.*

In this study, we investigated large sulfur bacteria collected from surficial sediments at five distant marine locations between 2004 and 2010. After the initial finding of a large morphological diversity among these bacteria, with many of them resembling described representatives of the genus *Thiomargarita*, extensive analysis of their 16S rRNA genes was initiated to study their phylogenetic relatedness. The large size of these bacteria enabled us to hand-pick single cells and filaments for each amplification reaction. In this

way, sequence data could be directly correlated with morphology and geographical origin of the cell or filament. Now, with a robust genetic basis and the morphology as support, we propose the retention of the originally published family *Beggiatoaceae* and suggest a revised classification of the taxa therein.

Materials and methods

Sample collection and localities

Benguela upwelling zone

Sediments from the South-East Atlantic off the coast of Namibia were obtained during the LEG AHAB 4 cruise (March 2004) and the R/V Meteor cruise M76 (April−June 2008). Samples were taken by a multicorer at 100−200 m water depth across the coordinate block 19°1.01'−25°30.00'S and 12°13.75'−14°23.36'E. The top 3 cm of the sediment were transferred into plastic beakers, covered with about 2 cm of the overlying water and stored at 4°C.

Costa Rica margin

Push cores of sediment and carbonate rocks hosting seep gastropods and *Bathymodiolus* spp. bivalves were sampled along the Pacific Costa Rica margin in January 2010 by the DSV Alvin at Mound 12 (water depth 900−1100 m, 8°55.8'N, 84°18.75'W). Sub-samples of the upper few centimeters of the sediment and hand-picked *Bathymodiolus* byssal threads were covered with overlying water and stored at 4°C.

South-East Mediterranean Sea

Sediment samples of the Nile Deep Sea Fan were taken in October 2006 during the BIONIL cruise M70/2 (R/V Meteor). A sulfide-rich band at the flank of the Amon Mud Volcano (32°22.04'N, 31°42.16'E) was sampled by push coring. Parts of the whitish mat overlying the sulfidic sediment were removed with a pair of tweezers and stored in TE buffer at −20°C.

Chilean margin

Sediment from the Chilean margin was sampled in January 2010 by the L/C Kay Kay II at Station 18 (36º30.80'S, 73º7.75'W, COPAS Time Series). Samples were taken at about 90 m depth with a box corer, filled in plastic bags (ca. 500 mL) and stored at 4°C.

Grund Fjord, Denmark

Samples from the Grund Fjord (56°28.11'N, 10°14.35'E) near Floes, Denmark, were taken in February 2010 through a hole in the ice cover. A push corer was attached to a stick and samples were taken by hand at approximately 1.5 m water depth, kept at room temperature for two days and subsequently stored at 4°C.

Cell preparation and sequencing

Filaments, chains and single cells of large sulfur bacteria were removed from the samples by hand with a Pasteur pipette tip, which was closed and thinned out by melting (except the Amon Mud Volcano-sample; see below). Cells attached to surfaces were scraped off using two needles. Cells from aggregates or filaments from bundles were removed manually with needles. All cells were washed in artificial seawater, removing sediment particles and mucus sheaths as thoroughly as possible. Only single cells, single chains or single filaments were used for polymerase-chain-reaction (PCR). These cells were photographed with a camera attached to a microscope (Sterni 2000-C, Zeiss, Germany; 40× magnification) and added to 25 µl PCR reaction mix containing 0.2 mM of each dNTP, 1 µM each of forward and reverse primer, 1× High Fidelity PCR enzyme mix reaction buffer with $MgCl_2$ (1.5 mM) and 1.25 U High Fidelity enzyme mix (Fermentas, St. Leon-Rot, Germany). Subsequently, the cells were disrupted inside the reaction mix with a sterile needle so that the cytoplasm was released. From the Amon Mud Volcano, only a DNA extract (Mo Bio Ultra Clean™ Soil DNA Isolation Kit, Dianova GmbH, Germany) of the mat sample was available and used for the PCR reaction.

The standard primers used for amplification in this study were VSO233F (5′-CCT ATGCCGGATTAGCTTG-3′, *Escherichia coli* 16S rRNA position 233−251) and ITSReub (5′-GCCAAGGCATCCACC-3′, *E. coli* 23S rRNA position 23−38 [11]). In some cases, single cells removed from an aggregate were used in several parallel PCR reactions containing different primer pairs in order to enhance the chance of retrieving a PCR product from the respective aggregate. A list of all primers utilized, including their specificities, is summarized in the Supplement Table S1.

Irrespective of the primer pair used, the following PCR program was applied: initial denaturation at 94°C for 2 min, followed by 28 cycles of 94°C for 30 s, 56°C for 30 s and 68°C for 4 min. In the last 18 cycles, the elongation step at 68°C was extended each time by an extra 10 s so that the final elongation lasted 7 min. After separation on a

0.8% agarose gel, the PCR products were cut out and extracted with a kit (Qiagen, Hilden, Germany). In most cases, PCR products were re-amplified with the previously used primer pair to increase the amount of PCR product for further processing. DNA fragments, either as PCR products or as cloned inserts, were sequenced using the Big Dye Cycle Sequencing Kit (Applied Biosystems, Darmstadt, Germany) and analyzed on an ABI Genetic Analyzer 3130x (Applied Biosystems, Darmstadt, Germany). The sequences were deposited in the GenBank/EMBL/DDBJ databases under the accession numbers listed in the Supplement Table S2.

Sequence and phylogenetic analyses

The quality of the obtained sequences was manually checked using FinchTV (Geospiza, Inc., USA) and full-length sequences were assembled with SeqMan (Lasergene software package, DNAStar, USA). During assembly, large, unusual regions (260−942 nucleotides) within some 16S rRNA gene sequences were identified. These regions had no resemblance to any part from other 16S rRNA genes from databases and are henceforth referred to as insertions. Prior to phylogenetic analysis, sequence regions of 16S rRNA genes, insertions and internal transcribed spacer (ITS) were determined and analyzed separately. The phylogenetic analysis of the 16S rRNA gene sequences was based on release 102 of the SILVA SSURef database [60] and was carried out using the ARB software package [47]. The sequence alignment, which was produced by the integrated aligner, was manually checked considering also the secondary structure information of the 16S rRNA molecule.

Trees were reconstructed with nearly full-length sequences of 296 selected representatives of the Gammaproteobacteria (*Beggiatoales*) and Deltaproteobacteria (as an outgroup) using neighbour joining, maximum parsimony and maximum likelihood (RAxML) methods. A termini filter considering only sequence information between *E. coli* position 279 and 1463 as well as 0, 30 and 50% positional conservatory filters that exclude highly variable regions were applied. Partial sequences (accession numbers FN811664, FR690910, FR690943, FR690948, FR690960, FR690981) were added to each reconstructed tree by parsimony criteria without allowing changes in the overall tree topology. Previously published partial sequences were not included in the trees. The branching patterns of all resulting trees were compared and a consensus tree based on the maximum likelihood tree calculated without a positional conservatory filter was constructed using the approach outlined in the Standard Operating Procedure for Phylogenetic Inference (SOPPI, [59]). Accordingly, multifur-

cations were introduced at nodes that were not supported by all treeing methods, demonstrating that the actual tree topology at those nodes could not be clearly resolved.

Separate databases of the ITS and insertion regions were constructed by importing the respective sequence parts into ARB. Sequences were aligned using Clustal W criteria [76] and manually checked afterwards. All sequence analyses and identity calculations for the different genetic regions were performed with the ARB software package.

Scanning Electron Microscopy (SEM) analysis

Sampled sulfur bacteria were washed in artificial seawater, fixed in 13% glutaraldehyde for 50 days at 4°C and were subsequently dehydrated in an ethanol series (30, 50, 70, 90, 99%; each step 30 min at room temperature). Cells were then transferred into a drop of 99% ethanol on an adhesive plate (Plano GmbH, Germany) and the ethanol was removed until the cells adhered to the surface but did not run dry. The plate was then transferred to a critical point dryer (Balzers CDP030, Bal-Tec/Leica Microsystems, Lichtenstein, Germany) and the ethanol was replaced by liquid CO_2. After evaporation of the CO_2, the plate was positioned onto a specimen stub and a second adhesive plate was carefully placed on top of the cells. By tearing off the second plate, the cells were disrupted with parts remaining on the first plate and detached parts adhering to the second plate. Both plates were subsequently sputtered with gold and used for SEM-analysis (DSM 940, Zeiss, Germany).

FITC staining

FITC staining is widely used for highlighting cytoplasmic structures. In case of organisms containing vacuoles, this compartment usually remains unstained and thus FITC staining allows indirect detection of vacuoles [18, 33-35]. Single cells and chains were washed in artificial seawater and fixed for 7 days at 4°C in 10% formaldehyde (diluted in 3.5% NaCl). Cells were then washed in 3.5% NaCl for 30 min and subsequently suspended in 1 mL phosphate-buffered saline (PBS). 10 µL FITC (fluorescein-5-isothiocyanate, 10 g L^{-1} in DMSO) was added and the solution was mixed with a pipette tip. The cells were incubated for 24 h at room temperature or at 4°C in the dark and subsequently washed three times in fresh PBS before being analyzed with a confocal laser scanning microscope (DMIRE2, Leica Microsystems, Germany and LSM510, Zeiss, Germany).

Results

16S rRNA gene sequence analysis

In total, we obtained 128 nearly full-length 16S rRNA gene sequences from single filaments, chains and cells. 107 sequences included the ITS region of the rRNA operon and 98 sequences contained insertions within the 16S rRNA gene. The majority of sequences originated from large spherical sulfur bacteria that morphologically resembled *Thiomargarita* cells. The highest morphological diversity within this group was detected in Namibian sediments, from which most sequences originated (n=91). Fewer sequences were obtained from samples taken off Costa Rica (n=17) and Chile (n=6). The sequence obtained from the DNA-extract of the Amon Mud Volcano sample was included in the phylogenetic analysis although it could not be assigned to an individual cell in this study. Its origin from *Thiomargarita*-resembling cells was demonstrated previously by fluorescence *in situ* hybridization [18]. The remaining sequences were obtained from filamentous sulfur bacteria originating from the coast of Chile (n=10) and Denmark (n=2) as well as from the *B. alba* strain B15LD retrieved from the German Type Culture Collection (n=1).

The 16S rRNA gene sequences retrieved in this study were closely related to published *Beggiatoa* and *Thioploca* sequences, together representing a monophyletic group within the consensus tree (Fig. 1). This monophyletic group was well separated from other families within the *Gammaproteobacteria*, including those containing other colorless sulfur bacteria like *Thiothrix* spp. or *Achromatium* spp. (Fig. 1 and Supplement Fig. S1). Therefore, this monophyletic group, which includes the type species *B. alba*, is henceforth referred to as the family *Beggiatoaceae*. The family *Beggiatoaceae* was further sub-divided into 12 clusters (Fig. 1, Clusters I– "XII"). Some clusters included sequences representing the polyphyletic genus *Beggiatoa* (see current nomenclature in the Supplement Fig. S1). Multifurcations at various nodes in Fig. 1 indicated that currently available 16S rRNA data were insufficient to completely resolve the affiliation of the respective groups. In many clusters, the sequenced organisms did not share a uniform morphology or a similar geographic origin, however, in a subset of clusters certain morphological characteristics do seem to be conserved.

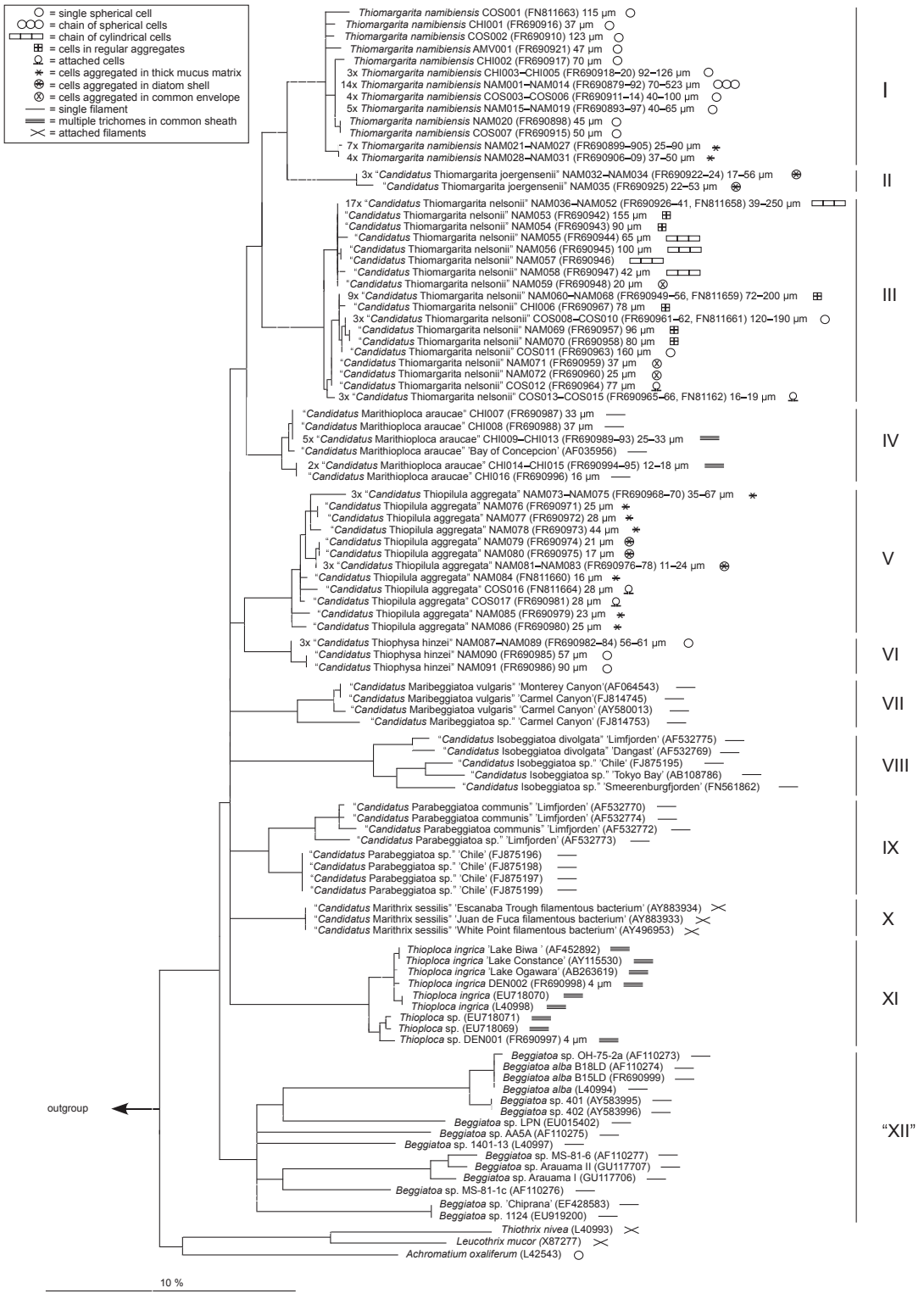


Fig. 1. Consensus tree of nearly full-length 16S rRNA gene sequences of the large, colorless sulfur bacteria, calculated with nucleotides at positions 279–1463 (according to *E. coli* numbering). The nomenclature of the validly published names *Thiomargarita*, *Thioploca* and *Beggiatoa* is sustained in the Clusters I, XI and "XII", respectively. The nomenclature of the remaining clusters is based on the proposal of *Candidatus* taxa made in this study. Roman numbers indicate identified clusters within the proposed family *Beggiatoaceae*. Cluster "XII" is written in quotation marks due to further sub-clustering, which cannot be resolved with the current data. Cell and filament diameters are only given for sequences, for which single cells or filaments were used in this study.

Cluster I in the 16S rRNA tree comprised sequences originating from spherical cells of diverse geographic origins. Within this cluster, 16S rRNA gene sequence identities were between 98.5 and 100% (Table 1). Cluster I also included the typical, spherical, chain-forming morphotype of the validly published type species *T. namibiensis* (Fig. 2a) discovered in 1999 [63]. The original description of *T. namibiensis* included a partial sequence (accession number AF129012, [63]), which was 93.0–94.9% identical with sequences of Cluster I (partial sequence not included in Fig. 1). Interestingly, all sequences from spherical cells in chains (n=14, Fig. 2a and b) retrieved in our study were completely identical. Besides cells occurring in chains (sometimes rather as lumps, Fig. 2a–c) with a thin mucus sheath, cells of Cluster I were single and uncoated or minimally mucus-covered (Fig. 2d−f), or were aggregated in a thick mucus matrix (Fig. 2g–i). The mucus-covered aggregates most frequently contained one or a few cells, but occasionally aggregates with more than 30 cells were found. Generally, diameters of cells in Cluster I ranged from 25 to 532 µm (n=118), but spherical, chain-forming *Thiomargarita* were reported to reach up to 750 µm in diameter [63]. When residing within a chain or an aggregate, diameters of individual cells sometimes varied greatly (see Fig. 2a, 86−155 µm, and Fig. 2g, 25–88 µm). Binary cell division and the presence of a large central vacuole (visualized by FITC staining, Supplement Fig. S2b, d and h) characterized all cells of Cluster I. Intracellular nitrate storage was previously reported in spherical, chain-forming *Thiomargarita namibiensis* cells [63], though it remains to be tested whether all representatives of Cluster I accumulate nitrate. While *Thiomargarita* was initially assumed to be non-motile, slow jerky rolling movement of single, spherical cells has been sporadically observed under the microscope. Single cells belonging to this cluster originated from sediments of the South-East Atlantic (Fig. 1, NAM015−NAM020), the South-East Pacific (Fig. 1, COS001–COS007 and CHI001–CHI005) and the Mediterranean Sea (Fig. 1, AMV001). Chain-forming and aggregate-forming cells of Cluster I have so far only been found in Namibian sediments (Fig. 1, NAM001–NAM014 and NAM021–NAM031).

Cluster II, a sister-group to Cluster I, comprised sequences of spherical cells that resembled single *Thiomargarita*-like cells from Cluster I, but that were found in empty, closed, centric diatom frustules (Fig. 2j–l). Sequence identities within Cluster II ranged between 99.2 and 100% and their nucleotide sequences were ≤97.3% identical to Cluster I (Table 1). Representatives of Cluster II have thus far only been observed in samples from diatomaceous ooze of coastal Namibia (Fig. 1, NAM032−NAM035). All cells in this cluster ranged from 17 to 56 µm (n=56) in diameter. Within individual frustules, up to

27 cells were counted and their diameters also varied (e.g. 22–53 µm). The cells were always arranged at the periphery of the frustule (Fig. 2j and k). Dividing stages as well as a slow rolling movement were frequently observed and vacuoles were detected by FITC staining (Fig. S2g).

Cluster III was composed of sequences from cells of diverse morphologies and different geographic origins with sequence identities ranging from 98.7 to 100%. Cluster III included a chain-forming morphotype similar to chains described for *T. namibiensis* that phylogenetically group with Cluster I, however, with a different cellular morphology. Rather than being spherical, cells in the chains of Cluster III had a cylindrical shape in which the long-axis oriented along the axis of the chain (Fig. 3a–c). Sequence identities between Cluster I and Cluster III were ≤97.1% (Table 1), however, the previously published partial sequence that was assigned to *T. namibiensis* (accession number AF129012, [63]) was 100% identical to several sequences of Cluster III, retrieved from cells with different morphologies (Fig. 1, e.g. NAM036–NAM054 and NAM059). Apart from chain-forming representatives, Cluster III included sequences of aggregated, regularly arranged spherical cells (Fig. 3d, f). This morphotype was discovered in the present study in sediments from the coasts off Namibia and Chile (Fig. 1, NAM053, NAM054, NAM060-NAM068, NAM069, NAM070, CHI006), as well as in previous studies in sediments from the Gulf of Mexico and Costa Rica [5, 33]. These cube-like assemblages of 4–8 or more spherical cells did not exhibit an apparent mucus sheath (Fig. 3d–f). Judged from the arrangement of the cells, other authors previously proposed that the cells divide in three or more planes, possibly constituting a reductive division [4, 5, 33, observed in this study as well]. The previously published partial sequence of aggregate-forming cells collected in the Gulf of Mexico (accession number AY632420, [33]) was 100% identical to sequences of Cluster III resulting from cells with comparable morphology (Fig. 1, NAM069, NAM070, COS010). A second aggregate-type of Cluster III, which featured cells residing randomly inside a rigid envelope (6 to more than 22 cells, Fig. 3g–i), was found in Namibian sediments (Fig. 1, NAM059, NAM071, NAM072). Cells collected at the Pacific coast off Costa Rica were spherical or elongated and attached to surfaces (Fig. 1 and 3j–k, COS012–COS015) or occurred as free-living cells (Fig. 1 and 3l, COS011). In the attached morphotypes, formations of spherical buds at the apical ends of attached elongated cells were previously observed by others [5] and in this study (data not shown). Cells of the remaining three morphotypes of Cluster III divided by regular binary fission. Generally, cell diameters in this cluster ranged from 16 to 250 µm (n=94) and all

morphotypes stained with FITC contained vacuoles (Fig. S2c–f). Nitrate accumulation was detected in aggregate-forming cells from the Gulf of Mexico [33], while a slow rolling movement was sporadically observed in the spherical enveloped cells from Namibia (this study).

Table 1. Identity values of 16S rRNA and ITS sequences among the species of the *Beggiatoaceae* and a comparison with the type species matching their morphology (TS).

Cluster in Fig. 1	name of organisms within the cluster	morphologically associated type species (TS)	16S identity within the cluster	ITS identity within the cluster	16S identity with TS	ITS identity with selected clusters
I	*Thiomargarita namibiensis*	*Thiomargarita namibiensis*	98.5−100%	95.9−100%	98.5−100%	≤82.3% (V)
II	"*Candidatus* Thiomargarita joergensenii"	*Thiomargarita namibiensis*	99.2−100%	99.0−100%	≤97.3%	≤89.4% (I), ≤86.9% (IV), ≤82.7% (V)
III	"*Candidatus* Thiomargarita nelsonii"	*Thiomargarita namibiensis*	98.7−100%	94.1−100%	≤97.1%	≤85.2% (I), ≤84.9% (II)
IV	"*Candidatus* Marithioploca araucae"	*Thioploca schmidlei*	98.3−100%	98.9−100%	≤92.8% (‡)	≤85.4% (I), ≤89.6% (III), ≤84.9% (XI)
V	"*Candidatus* Thiopilula aggregata"	*Thiomargarita namibiensis*	96.5−100%	98.0−100%	≤94.9%	≤83.7% (III), ≤89.1% (IV)
VI	"*Candidatus* Thiophysa hinzei"	*Thiomargarita namibiensis*	99.5−100%	NA	≤95.9%	NA
VII	"*Candidatus* Maribeggiatoa spp."	*Beggiatoa alba* B18LD	97.6−100%	NA	≤88.4%	NA
VIII	"*Candidatus* Isobeggiatoa spp."	*Beggiatoa alba* B18LD	94.9−98.8%	NA	≤86.8%	NA
IX	"*Candidatus* Parabeggiatoa spp."	*Beggiatoa alba* B18LD	95.0−99.9%	NA	≤88.2%	NA
X	"*Candidatus* Marithrix sessilis"	*Thiothrix nivea*	99.5−100%	NA	≤85.9%	NA
XI	*Thioploca* spp.	*Thioploca schmidlei*	96.2−99.8%	81%	NA	≤78.4% (I), ≤83.2% (II), ≤85.6% (III), ≤85.6% (V), ≤80.8% ("XII")
"XII"	*Beggiatoa alba* B15LD	*Beggiatoa alba* B18LD	86.1−100%	only 1 sequence available	99.8%	≤76.2% (I), ≤79.5% (II), ≤76.1% (III), ≤79.9% (IV), ≤75.6%(V)

(‡) Since a 16S rRNA sequence for *T. schmidlei* is not available, this value represents identity towards *Thioploca ingrica*.

Cluster IV, which branched off at the base of Clusters I–III, comprised sequences obtained from filamentous sulfur bacteria collected off the coast of Chile. Both free-living filaments and bundled trichomes that were enclosed by a common sheath were observed in these sediments. The diameter of both types ranged between 12 and 37 µm (n=10). Cluster IV was further divided into two sub-clusters that both contained sequences originating from single filaments and bundled trichomes. The first sub-cluster included seven sequences (Fig. 1, CHI007–CHI013) that were completely identical (100%) with each other as well as with the previously published partial sequence of *Thioploca araucae* (accession number

L41043, [73], not included in Fig. 1). Furthermore, this sub-cluster included the nearly full-length sequence previously assigned to *Beggiatoa* sp. "Bay of Concepción" (accession number AF035956, [72]) that was 99.8% identical to the other seven sequences. Filament diameters of representatives affiliated to this sub-cluster ranged from 25 to 37 µm. The second sub-cluster comprised three identical sequences (Fig. 1, CHI014−CHI016) that shared 98.9% sequence identity with the previously published partial sequence of *Thioploca chileae* (accession number L40999, [73], not included in Fig. 1). Filament diameters in this sub-cluster ranged from 12 to 18 µm. The two sub-clusters were 98.3–99.5% identical (Table 1). For the two previously described species *T. araucae* and *T chileae*, gliding motility of the trichomes, the presence of a central vacuole and storage of nitrate were reported [15,49]. Except for the differences in filament diameters, the bundled morphotype of Cluster IV resembled freshwater *Thioploca* of Cluster XI with the type species *T. schmidlei* (see below), whereas the free-living filaments of Cluster IV resembled *Beggiatoa* of Cluster "XII" with the type species *B. alba* (see below). *T. schmidlei* and *B. alba* feature diameters of 5–9 µm [43] and 1.5–4 µm [77,81], respectively, whereas the bundled Chilean filaments were 12−40 µm in diameter [49, this study] and the free-living ranged from 16 to 40 µm in diameter [72, this study]. Sequence identities between Clusters IV and XI were ≤92.8% (Table 1) and sequences of Cluster IV shared ≤87.8% with the *B. alba* type strain B18LD (accession number AF110274, [2]) in Cluster "XII".

Cluster V consisted of sequences obtained from cell aggregates occurring in Namibian sediments or from cells attached to solid substrates in sediments from the Costa Rica margin. All sequences were 96.5−100% identical and most sequences were retrieved from cells that were aggregated in a thick mucus matrix (Figs. 4a–c and 1, NAM073, NAM078, NAM085, NAM086 and FN811660). This morphotype resembled the respective morphotype of Cluster I (Fig. 2g–l), however, Cluster V was rather distantly related to Cluster I, sharing ≤94.9% sequence identity (Table 1). Aggregated cells affiliated to Cluster V ranged from 25 to 67 µm (n=34) in diameter and were often variable in size within a single aggregate. Cluster V also contained sequences from spherical cells residing in diatom frustules (Figs. 4d–e and 1, NAM079, NAM080) with variable cell diameters within one as well as between different frustules (15–24 µm, n=10). This morphotype was similar to those frustule-inhabiting cells of Cluster II, however, instead of being restricted to the periphery of the frustule (Fig. 2j and k), cells affiliated to Cluster V were evenly distributed within the enclosure (Fig. 4d and e). All spherical cells of Cluster V divided by

binary fission. Furthermore, attached cells from Cluster V were morphologically indistinguishable from attached cells of Cluster III (Figs. 4g and 3j–k, respectively). The putative reproduction mechanism of this morphotype via budding, as well as its segregation in at least two distinct phylogenetic groups, is described in detail by Bailey *et al*. [5]. Diameters of all cells in Cluster V ranged from 15 to 67 µm (n=46). The presence of vacuoles in most morphotypes of this cluster was shown by FITC staining (Fig. S2g and h). A jerky rolling movement was frequently observed in bacteria from aggregates.

Cluster VI was a sister-group of Cluster V. The two clusters shared a maximum sequence identity of 96.9% and the sequence identity of Cluster VI shared with Cluster I was ≤95.9% (Table 1). The sequence data in this cluster were exclusively derived from single spherical cells (Fig. 4 h and i), which originated from Namibian sediments. Cells of this morphotype were vacuolated (Fig. S2d), 56–90 µm (n=5) in diameter and sporadically moved by slow jerky rolling.

Sequences forming Clusters VII– "XII" were not obtained in this study (except for sequences with accession numbers FR690997–99), but are presented here to complete the proposed family *Beggiatoaceae*. Information about the cellular morphology was taken mostly from the available literature. It should also be noted that the respective sequences from the literature were retrieved from bulk samples or from DNA-extractions of numerous cells and filaments (which in some cases were pure cultures) rather than by single-cell or single-filament PCR as applied in this study.

Cluster VII comprised sequences from filamentous bacteria collected at the Carmel and Monterey Canyons (Monterey Bay, CA, USA). Filament diameters generally ranged from 65 to 85 µm, but can be as narrow as 20–30 µm. These bacteria were reported to occur as free-living filaments, to exhibit a gliding motility and to contain intracellular vacuoles [1,35]. Nitrate storage was tested and detected in one case [35]. Sequences of Cluster VII were published as belonging to the genus *Beggiatoa* [1,35], however, they were very distantly related to the genus' type strain *B. alba* B18LD, sharing ≤88.4% sequence identity (Table 1). Moreover, one sequence in Cluster VII (accession number FJ814753) was more distantly related to the others (97.6% sequence identity).

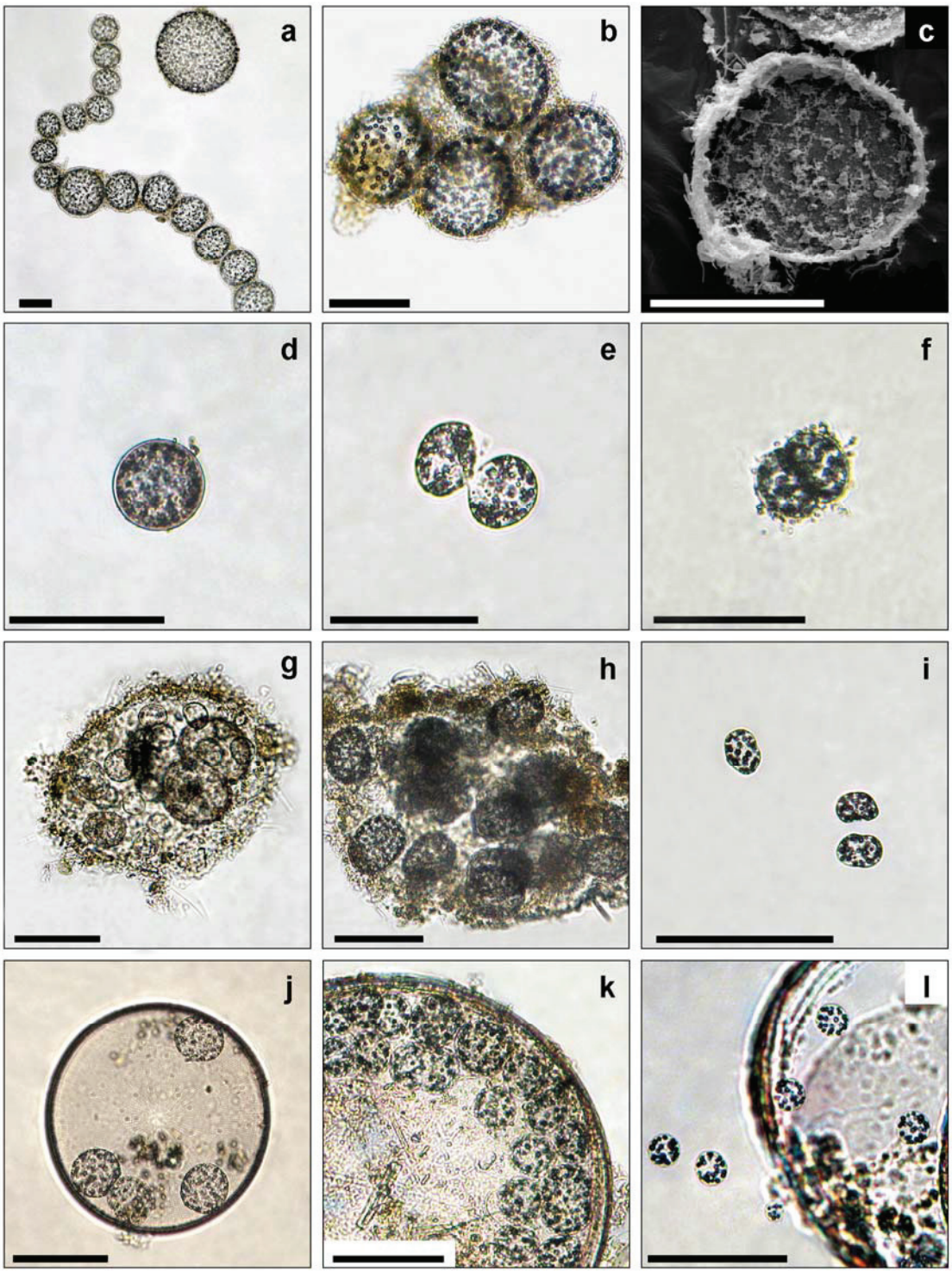
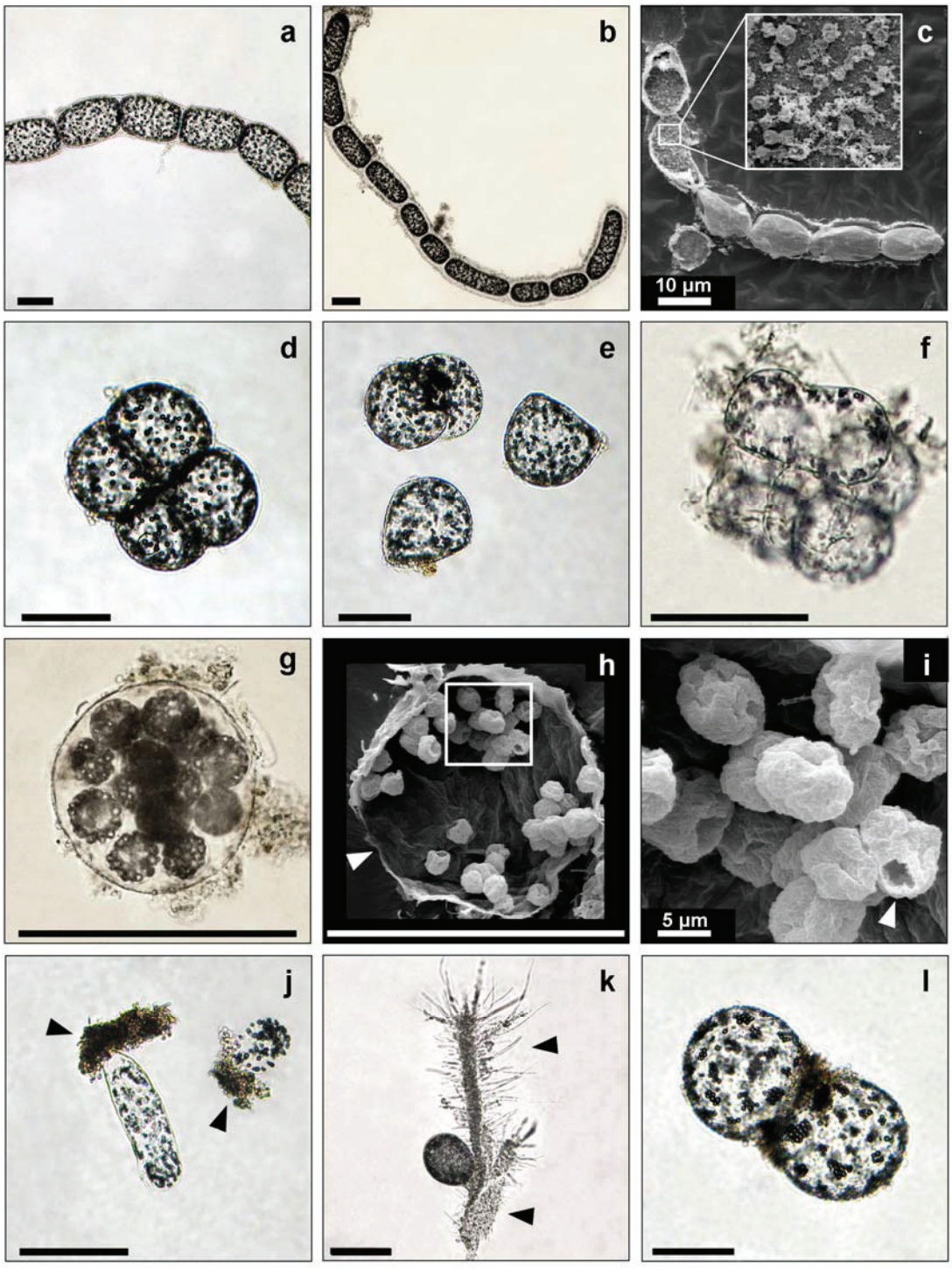


Fig. 2. Images of *T. namibiensis* (a–i) and "*Candidatus* Thiomargarita joergensenii" (j–l). The morphology of *T. namibiensis* is highly diverse, including ensheathed spherical cells that form chains (a) or lumps (b), uncoated or barely covered single cells (d–f) and colony-forming spherical or ovoid cells aggregated in a thick mucus matrix (g and h). The cells can be removed from the mucus when manually manipulated (i). The cytoplasm of *T. namibiensis* consists of a complex network of strings and fibres forming a sponge-like structure surrounding the large central vacuole (c, SEM-image). Cells of "*Candidatus* Thiomargarita joergensenii" have been observed to be exclusively spherical in shape, residing inside empty diatom frustules (j and k) that can be opened manually releasing the interior cells (l). Scale bars are 100 µm.

Sequences of **Cluster VIII** were also referred to as *Beggiatoa* spp. in previous studies that analyzed filamentous free-living bacteria phylogenetically. In this study, Cluster VIII was divided into two sub-clusters, the first of which contained sequences of filaments originating from the brackish Limfjorden, Denmark [53], and the intertidal mud flat of Dangast, Germany [53]. The second sub-cluster comprised sequences of filaments sampled from various marine habitats like sediments off the coast of Chile [3], from Smeerenburgfjorden on Svalbard [30], as well as from estuarine sediments of Tokyo Bay, Japan [38]. Filament diameters were reported to range from 10 to 40 µm in Limfjorden, 9 to 11 µm in Dangast, were 20 µm in Smeerenburgfjorden and 30 µm in Tokyo Bay. All filaments exhibited gliding motility and for some organisms the presence of large central vacuoles and intracellular storage of nitrate were reported [30,53]. The two sub-clusters were 94.9–95.5% identical and shared ≤86.8% identity with *B. alba* B18LD (Table 1).

Similar to Cluster VIII, **Cluster IX** was composed of free-living filaments morphologically resembling *Beggiatoa* spp. Two sub-clusters that appeared to correlate with habitat salinity were found. Sequences of the first sub-cluster originated from samples taken at the brackish Limfjorden, Denmark [53], while sequences of the second sub-cluster originated from marine sediments off Chile [3]. The sequence data of the two sub-clusters were 94.9–97.2% identical and all sequences of Cluster IX were distantly related to the type strain *B. alba* B18LD, sharing ≤88.2% sequence identity (Table 1). More detailed information on the organisms was reported only from the Danish samples, where filament diameters ranged between 5 and 30 µm, the presence of large central vacuoles was confirmed and intracellular nitrate was detected [53].

Fig. 3. Images of "*Candidatus* Thiomargarita nelsonii" exhibiting highly diverse morphologies including: ensheathed chain-forming cylindrical cells (a–c), regularly arranged spherical cells forming tetragonal aggregates that divide in three planes (d–f, in (e) cells were separated manually), colony-forming spherical cells surrounded by a rigid envelope structure (g–i), attached cells (j and k) and single spherical cells (l, binary division stage). Like in *T. namibiensis*, the cytoplasm of "*Candidatus* Thiomargarita nelsonii" cells consists of a network-structure of strings surrounding the central vacuole (c, SEM image, the white box



includes a magnified view onto the inner surface of an opened cell). The envelope (arrow head in h, SEM image) that encloses colony-forming "*Candidatus* T. nelsonii" cells does not exhibit a cytoplasmic-like texture, but rather a smooth surface from the inside (the white box indicates region magnified in picture i). Cells inside the envelope appear hollow showing internal vacuoles (arrow head in i, SEM image). Panels j and k show cells attached via holdfast structures to phytodetritus (arrow heads in j) and a byssal thread (arrow heads in k). Scale bars are 100 µm if not noted differently.

Sequences of **Cluster X** originated from bacteria collected at hydrothermal vents and ridges along the west coast of the USA. These organisms were originally referred to as "vacuolated attached filaments" [34,35]. The filaments were attached to rocks, tubeworms and other solid surfaces surrounding the vents. The formation of rosettes and the production of gonidia were observed occasionally, resembling morphological characteristics of *Thiothrix* spp. [35]. Filaments were non-motile and either narrow (less than 2 μm in diameter) and devoid of a vacuole, or wider (4–112 μm) and vacuolated. Nitrate storage could not be detected, but instead a transient storage of oxygen was suggested [34,35]. Sequences of this cluster shared identity values of ≤87.9% and ≤85.9% with the type strains *B. alba* B18LD and *T. nivea* (accession number for *T. nivea* is L40993, [73]), respectively (Table1).

Cluster XI included sequences obtained from bundled trichomes resembling *Thioploca* spp. encountered in various fresh and brackish water habitats [25,36,37,73, this study]. The trichomes, which measured between 2 and 5.6 μm in diameter, exhibited independent gliding motility [25,36,73, this study]. The *Thioploca* type species *T. schmidlei* [43], with trichome diameters between 5 and 9 μm, was reported from freshwater habitats but neither a culture nor phylogenetic data from this species currently exist. A second freshwater species, *Thioploca ingrica*, has been classified [48, 83] and sequenced (accession numbers L40998 [73] and EU718069–EU18071 [25]). *T. ingrica* occasionally contained vacuoles [50,74,83,84] and internal nitrate was recently detected as well [25,84]. Cluster XI was divided into two sub-clusters, each of them containing one of the sequences obtained in this study and two of the published *T. ingrica* sequences. The first sub-cluster included the oldest available sequence for this species (L40998), three other sequences published as non-vacuolated *Thioploca* spp. [36,37] and the sequence DEN002 (FR690998) retrieved in this study. The second sub-cluster shared ≤97.5% identity with the first and contained two *T. ingrica* sequences together with the here-obtained sequence DEN001 (FR690997). All sequences within Cluster XI were between 96.2 and 99.8% identical (Table 1).

Cluster "XII" contained sequences of free-living filaments published as *Beggiatoa* spp. [2,13,23,73], including the *Beggiatoa* type strain *B. alba* B18LD [70]. Here, the cluster is written in quotation marks because the phylogenetic analysis implied a further sub-clustering that, however, could not be fully resolved based on the current data. The implied sub-clusters (Fig. 1, Cluster "XII" from top to bottom) comprised: (i) a monophyletic group of seven sequences from freshwater strains that included the type strain *B. alba* B18LD, (ii)

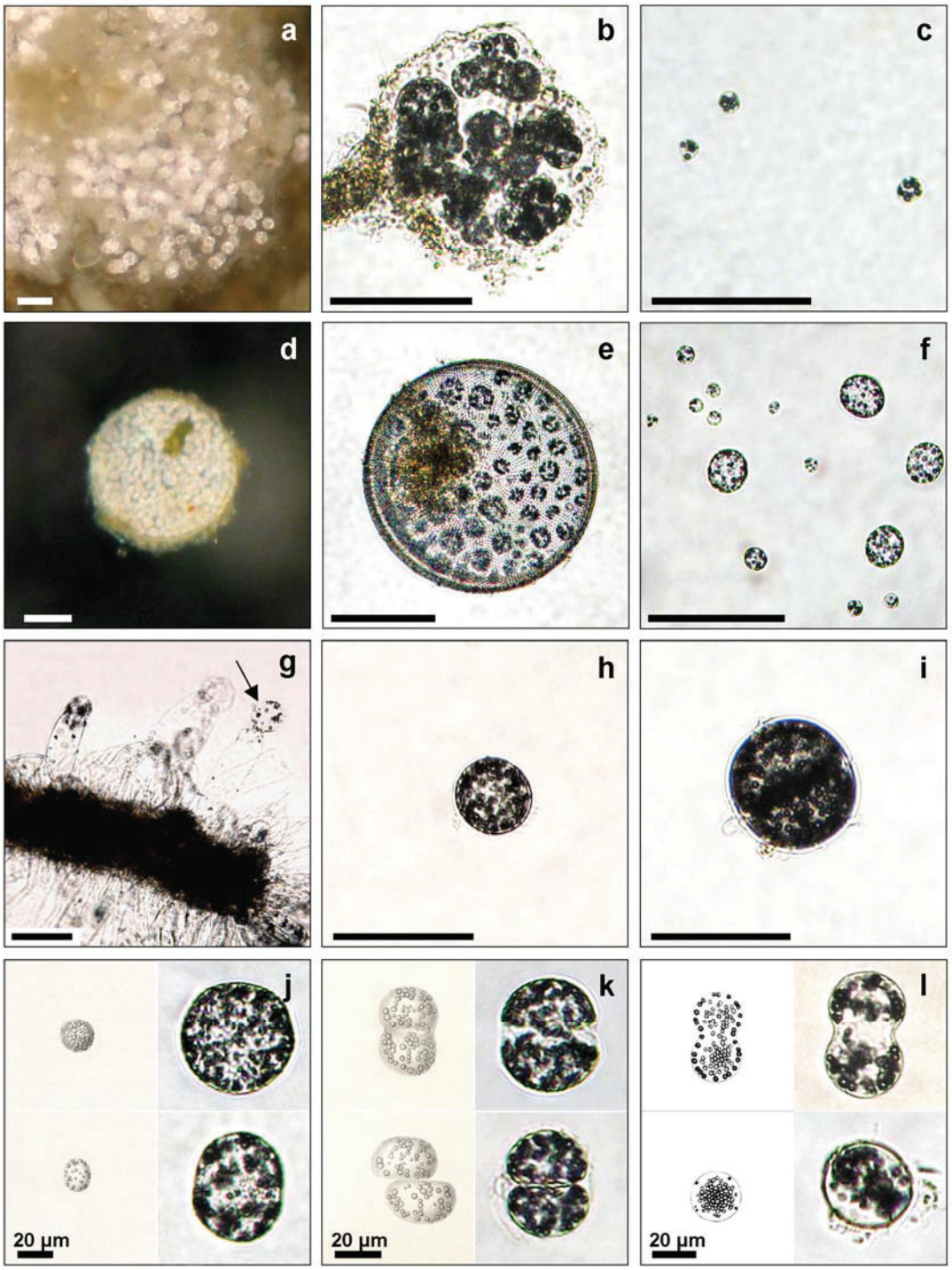


Fig. 4. Images of "*Candidatus* Thiopilula aggregata" (a–g), "*Candidatus* Thiophysa hinzei" (h and i) and comparisons of recent spherical sulfur bacteria with drawings of *Achromatium volutans* by Hinze (1903; j and k) and Nadson (1914; l), originally named *Thiophysa volutans* [24, 80]. "*Candidatus* Thiopilula" cells usually form colonies in thick mucous matrices (a and b) or are residing in empty diatom frustules (d and e). Cells from both structures can be released when manually manipulated (c and f). Also, they occur attached to solid surfaces forming apical gonidia (arrow in g). "*Candidatus* Thiophysa" so far occurred exclusively as single cells (h and i), strongly resembling spherical sulfur bacteria discovered and documented at the beginning of the 20th century (j–l). The different sizes of cells and sulfur globules within panels j, k and l possibly represent natural variability observed in natural populations including those examined in this study. Scale bars are 100 µm if not noted differently.

another monophyletic group of two sequences from non-vacuolated filaments from a hypersaline habitat (GU117707, GU117706) and two sequences from marine strains (AF110277, AF110276) and (iii) two identical sequences of vacuolated *Beggiatoa* spp. from a hypersaline environment (EF428583, EU919200). Two individual sequences remain unclear in their affiliation to any of these groups and therefore emerge from the common baseline (AF110275, L40997). The sub-clusters furthermore exhibited only low identity values, reaching down to merely 86.1% identity among all sequences currently subsumed within Cluster "XII" (Table 1) and likely require future taxonomic revision. *Beggiatoa* filaments vary in their diameters between 2 and 5 μm for freshwater [57,69,70,77] and marine environments [55,73] and range from 2 and 8 μm for hypersaline habitats [13,23]. To date, vacuolation as well as nitrate storage have only been detected in the widest representatives originating from hypersaline environments (EF428583 and EU919200) and have even been ruled out for some narrow, marine pure-culture strains [52,56]. Gliding motility, however, is observed for all of the sequenced organisms. Most sequences in this cluster originated from pure cultures or enrichments.

To clarify the affiliation of the above presented sequences to more distantly related sulfur bacteria, the three type species of related Gammaproteobacteria, *T. nivea* [42,82], *L. mucor* [46,58] and *Achromatium oxaliferum* [22,61], were included in the phylogenetic comparison with the *Beggiatoaceae*. The maximum identity that the 16S rRNA gene sequences of these organisms shared with that of the *Beggiatoaceae* type species *B. alba* was 85.8%. The first two organisms are narrow filaments (0.7–5 μm) and occur either attached to algae, to other benthic surfaces or numerous filaments are connected at one end, forming free-floating rosettes. They are non-motile, but can produce gliding gonidia, either successively at the distal end of the filament (*Thiothrix* spp. [79]) or by a complete disintegration of the filament into numerous gonidia (*Leucothrix* spp. [8]). Internal vacuoles and nitrate-storage have not yet been reported for these taxa. *Thiothrix* spp. have been observed exclusively in freshwater habitats and seven species of the genus are currently recognized. *L. mucor* is presently the only species of its genus and originates from marine sediments. It does not store elemental sulfur as internal globules, in contrast to its sulfur-containing relatives. Both of these type species exist as pure cultures, their 16S rRNA gene sequences were 87.9% identical and shared ≤86.5% identity with the *Beggiatoaceae*. The third type species, *A. oxaliferum* [22,61], comprises spherical to cylindrical bacteria from freshwater habitats, which store conspicuously large calcium carbonate inclusions within the cell [7]. Similar to what was observed in this study for

Thiomargarita-like cells, *A. oxaliferum* exhibits a slow, jerky rolling motility on sediment surfaces and glass slides [24,61]. Internal vacuolation and nitrate-storage for *A. oxaliferum* are not known. No enrichments or cultures are available for this type species. *A. oxaliferum* shared ≤86.8% sequence identity with *T. nivea* and *L. mucor* and ≤87.2% with the *Beggiatoaceae*.

ITS sequences

ITS sequences of the non-filamentous sulfur bacteria were between 395 and 490 nucleotides long. The shortest ITS sequence was detected in filamentous *B. alba* B15LD (337 nt, FR690999) and the longest in trichomes of *Thioploca* sp. DEN001 (543 nt, FR690997). Conserved and variable sequence blocks as well as genes for $tRNA^{ile}$ and $tRNA^{ala}$ were identified by the comparison of all ITS sequences retrieved (Fig. S2a). The phylogenetic analysis of the ITS region reflected the 16S rRNA phylogeny on a fine scale. Generally, identity values of ITS sequences were highest (94.1–100%) within a 16S rRNA-derived phylogenetic clusters and lower (<90%) when separate clusters were compared (Table 1). Within sub-clusters, like those of Cluster IV, ITS sequences were even completely identical and were slightly less identical (≤98.9%) when the two sub-clusters were compared.

Insertions within the 16S rRNA gene

Of all obtained 16S rRNA gene sequences, 77% contained insertions of variable length (260–942 nt). The localization of the insertions at the sites 795, 1078, 1396 and 1495 (based to *E. coli* numbering) was conserved and thus formed the basis for their classification (insertion types 795, 1078, 1396 and 1495; Supplement Tables S2 and S3). The insertions so far occurred in sequences affiliated to Clusters I–VI, and their type and number differed among the clusters (Supplement Table S3). The four different insertion types were not related to one another in their primary sequence (maximum identity 32.7%). Further details on this topic will be given elsewhere (Salman et al. in prep.).

Discussion

Since their discovery, the classification of the large sulfur bacteria has been based primarily on morphological characteristics: (1) single free-living filaments were identified as *Beggiatoa* spp., (2) multiple trichomes surrounded by a common sheath as *Thioploca* spp., (3) single filaments attached via a hold-fast structure and containing sulfur inclusions as *Thiothrix* spp. or when devoid of sulfur inclusions as *Leucothrix* spp., (4) mucus-covered

spherical cells as *Thiomargarita* spp. and (5) spherical cells without an outer mucus layer as *Achromatium* spp. Only a few 16S rRNA gene sequences have been produced, and in many cases, only rather distant relationships between the identified organisms and the type species of the respective assigned genus were found [1,35,53,72,73].

Inconsistencies in the identification of large sulfur bacteria

Within the group of large filamentous sulfur bacteria, the genus *Beggiatoa* with the type species *B. alba* is defined by (1) the existence as single free-living filaments, (2) the formation of sulfur inclusions and (3) a gliding motility. The genus *Thioploca* with the type species *Thioploca schmidlei* differs from the genus *Beggiatoa* by only one property: instead of being free-living, the trichomes of *Thioploca* are bundled in a common sheath. Difficulties in the morphological differentiation between the two genera are caused by the fact that trichomes leaving the common sheath of a *Thioploca* filament are morphologically indistinguishable from single *Beggiatoa* filaments [31,64,72]. In the present study, we retrieved 16S rRNA gene sequences from both single filaments and bundled trichomes from the same sediment sample and found that both were genetically closely related (98.3–100% identical, Fig. 1, Cluster IV). We thus conclude that *Thioploca*-like trichomes can occur and probably live also as free filaments outside of their common sheaths, in agreement with earlier observations [32,49,64]. The hypothesis of free-living *Thioploca*-like trichomes is further supported by the observation of trichomes of different diameters within one sheath, supposedly resulting from free-living filaments entering a sheath populated by trichomes of a different diameter. This has been observed for both marine [49,64] and freshwater species [39]. An example illustrating the difficulty in the discrimination between *Beggiatoa* spp. and *Thioploca* spp. is a study by Teske *et al.* [72], in which phylogenetic analysis of free-living filaments showed their close affiliation with partial sequences of Chilean *Thioploca* spp. (>98% sequence identity). In spite of that, the organisms were classified as *Beggiatoa* sp. based on the lack of a common sheath [72]. We conclude that the sole occurrence as a single filament or the absence of a common mucus sheath should no longer be a determinative property of the genus *Beggiatoa.*

Filamentous sulfur bacteria have not only been identified by morphology on the genus level, but also on the species level, for which the filament diameter was considered. As known from clonal cultures, single *Beggiatoa* strains have a conserved filament diameter [9,55,57,69,70]. Also, in certain habitats it can be observed that filaments exhibit a rather restricted filament diameter [23,30,38,41,56]. However, to identify a certain species or

strain in an environmental sample using filament diameter as known from the culture might be delusive because diameters of strains in natural populations sometimes overlap [29]. It is likely that a frequently changing habitat alternatingly favours various strains of one or even several species, which might have identical filament diameters. In our study, this is supported by the observation of overlapping diameter values even for clusters that are supposed to represent different genera among the large sulfur bacteria: each of the Clusters VII, VIII and IX includes free-living, vacuolated, marine filaments with diameters in the range of 20–40 µm (Table 2). Only free-living filaments of Cluster "XII" can so far be distinguished from the others as they are narrow (≤9 µm) and mostly non-vacuolated. Ensheathed filaments so far seem to be distinguishable based on their habitat and trichome diameters. Wide (12–43 µm, Cluster IV), marine, vacuolated bacteria differ from brackish/freshwater, narrow (≤9 µm, Cluster XI), mostly non-vacuolated representatives (Table 2). However, the finding of narrow (2.5–5 µm), marine, bundled trichomes [31,51,64] raises the question whether these are phylogenetically affiliated to Cluster IV as they are marine or with Cluster XI as they are narrow.

Besides free-living populations, recently also attached marine filaments were found that were covering benthic substrates surrounding hydrothermal vents [20,26,34,35,68] or were attached to each other, forming free-floating rosettes. The filaments produced gonidia at their free ends [20,26,35]. These morphological features resemble narrow (1–4 µm), non-vacuolated freshwater *Thiothrix* spp. [20,26,34,35,68]. However, the diameter of the recently discovered filaments was of a wide range (1–112 µm) and the cells were vacuolated when diameters exceeded 4 µm [35]. Phylogenetic data [34,35] revealed a close relationship of these filaments collected from different hydrothermal vents (Fig. 1, Cluster X), while identifying a profound difference between these organisms and freshwater *Thiothrix* and marine *Leucothrix* (≤85.9% sequence identity). Due to this conflict, the authors refrained from classifying the sequenced organisms. In addition to the attached filamentous sulfur bacteria, attached non-filamentous sulfur bacteria were discovered even more recently [5, this study]. These spherical, ovoid or elongated single cells were 16–100 µm in diameter and up to 700 µm in length. So far, they have only been observed at methane seeps. These attached cells were always vacuolated and often formed gonidia at the free end of an attached cell. Bailey *et al.* [5] suggested a dimorphic life-style for these bacteria, describing gonidia production at the apical end of an attached egg-shaped or stalk-like elongated cell that results in spherical cells being released into the surrounding sediment. Our findings support this hypothesis (Fig. 4g). In contrast to *Thiothrix*-like

filaments, the attached non-filamentous cells have not yet been shown to form free-floating rosettes. In our study, 16S rRNA gene sequences of attached cells and filaments affiliated to at least three clusters (Fig. 1, Clusters III, V and X) that should be treated as three distinct genera (see below). Apparently, the production of hold-fast structures and the formation of free-living offspring from attached ancestral cells is a more frequently appearing characteristic of the large sulfur bacteria than previously recognized.

Non-filamentous, free-living sulfur bacteria have so far been assigned to the genera *Achromatium* or *Thiomargarita*. *Achromatium* is morphologically defined as a genus of single, spherical to ovoid cells that are 10–35 µm wide, contain numerous calcium carbonate inclusions, lack an outer sheath and frequently move by slow jerky rolling. The genus *Thiomargarita* was considered to comprise non-motile, spherical cells of 40–750 µm diameter, that contain sulfur inclusions and have an outer mucus sheath, sometimes forming chains. In several studies from marine environments, single, spherical cells were found [12,18,24,41,54,75, this study] and the organisms were usually assigned to either *Achromatium* or *Thiomargarita*, sometimes without reference to phylogenetic data. In our study, non-filamentous sulfur bacteria varied both in cell diameter (15 to more than 200 µm) and cell sheathing (ranging from uncovered or barely coated cells to several micrometer thick mucus-covered cells), were sporadically motile by slow jerky rolling, but never contained calcium carbonate inclusions. These cells affiliated phylogenetically to several clusters of the *Beggiatoaceae* (Table 2; Fig. 1, Clusters I, II, III, V and VI), but did not show any cluster-specific grouping according to their different morphological features. These examples illustrate that a seemingly homogeneous population of non-filamentous sulfur bacteria may be, similar to the filamentous sulfur bacteria, a mixture of different phylogenetic taxa. In addition, our observation of a sporadic movement in some of the spherical cells implies that, among the non-filamentous sulfur bacteria, not only *Achromatium* spp. have the ability to move. Consequently, morphology is not a reliable criterion for a differentiation between *Achromatium* spp. and *Thiomargarita* spp. However, two unique attributes have so far been identified for each of the two genera. First, chain-formation was found only within the genus *Thiomargarita* and thus discriminates these morphotypes from *Achromatium* spp. that always appeared singly. However, chain-formation is not a discriminative property of the genus *Thiomargarita* because numerous other morphotypes, including single cells, are as well observed in this genus (Fig. 1, Table 2, Clusters I–III). Second, the formation of calcium carbonate inclusions so far

constitutes a discriminative property that subsumes members of the genus *Achromatium*. However, calcium carbonate inclusions were not found in the species *A. volutans* [80].

In conclusion, classification and identification of large sulfur bacteria based on morphological features is ambiguous. When considering possible polymorphisms in individual cells (dimorphic life-style), some morphologies may even be inducible. Therefore, we suggest that classification and identification of large sulfur bacteria should be primarily based on phylogenetic data. Supported by our molecular data, we agree to the revision of the taxonomy of the sulfur bacteria recommended by others [1,53,72] and therefore propose several new *Candidatus* genera and species to take a first step into the direction of the aspired taxonomic revision.

Proposed taxonomic revisions

The *Beggiatoaceae* as an independent family, excluding other genera such as *Achromatium* or *Leucothrix*, was validly published in the current (8th) edition of Bergey's Manual of Determinative Bacteriology [44], in the thereafter published Approved Lists [66,67] as well as in the subsequent (1st) edition of Bergey's Manual of Systematic Bacteriology [71]. The recent (2nd) edition of Bergey's Manual of Systematic Bacteriology [17], however, introduced a combined family name, the *Thiotrichaceae*, which included the following genera: *Beggiatoa*, *Thioploca*, *Thiomargarita*, *Thiothrix*, *Leucothrix*, *Achromatium*, *Thiobacterium* and *Thiospira* [17]. Certainly, this classification is illegitimate according to rule 51b of the International Code of Nomenclature of Bacteria, because the proposed family *Thiotrichaceae* would contain three genera that are type genera (*Beggiatoa*, *Leucothrix* and *Achromatium*) of already validly published families (see also [14]). 16S rRNA data reveal a sequence identity of only ≤85.8% among the type species of these genera and furthermore show that the individual families (*Beggiatoaceae*, *Leucotrichaceae* and *Achromatiaceae*) are monophyletic while the combined family *Thiotrichaceae* is not (Fig. S1). Accordingly, we support the retention of the original family name *Beggiatoaceae*, including the validly published genera *Beggiatoa*, *Thioploca* and *Thiomargarita*, and propose to add nine *Candidatus* species along with seven new *Candidatus* genera, of which the 16S rRNA gene sequences are monophyletic within the family. We furthermore suggest excluding all other genera that are not monophyletic with the family or that were in the past included merely according to morphological similarities.

Table 2. List of proposed genera of the family *Beggiatoaceae* together with a summary of their currently available characteristics. 'Yes' denotes that characteristics have been found in at least one member of the genus.

Cluster in Fig. 1	Former genus	New genus	Type species	Habitat [a]	Oxygen preference	Shape and arrangement	Diameter	Motility	Sulfur inclusions	Internal vacuolation	Internal nitrate tested / detected	Production of gonidia	References
I, II, III	*Thiomargarita*	*Thiomargarita*	*T. namibiensis*	m	Presumably aerobic or microaerophilic	Spherical or cylindrical cells in chains, spherical or elongated cells, single or in aggregates or attached	16–750 µm	None or slow rolling	Yes	Yes	Yes / Yes	Yes	[5,18,33,63,5]
IV	*Thioploca*	"*Candidatus* Marithioploca"	"*Candidatus* M. araucae"	m	Microaerophilic	Single filaments or multiple trichomes in a sheath	12–43 µm	Gliding	Yes	Yes	Yes / Yes	No	[15,16,49,64]
V	–	"*Candidatus* Thiopilula"	"*Candidatus* T. aggregata"	m	Presumably aerobic or microaerophilic	Spherical or elongated cells, single or in aggregates or attached	15–65 µm	None or slow rolling	Yes	Yes	No / No	Yes	this study
VI	–	"*Candidatus* Thiophysa"	"*Candidatus* T. hinzeii"	m	Presumably aerobic or microaerophilic	Single spherical cells	56–90 µm	None or slow rolling	Yes	Yes	No / No	No	this study
VII	*Beggiatoa*	"*Candidatus* Maribeggiatoa"	"*Candidatus* M. vulgaris"	m	Presumably microaerophilic	Single filaments	20–85 µm	Gliding	Yes	Yes	Yes / Yes	No	[1,35]
VIII	*Beggiatoa*	"*Candidatus* Isobeggiatoa"	"*Candidatus* I. divolgata"	b, m	Microaerophilic	Single filaments	10–40 µm	Gliding	Yes	Yes	Yes / Yes	No	[30,38,53]
IX	*Beggiatoa*	"*Candidatus* Parabeggiatoa"	"*Candidatus* P. communis"	b, m	Microaerophilic	Single filaments	10–40 µm	Gliding	Yes	Yes	Yes / Yes	No	[53]
X	*"vacuolated attached filaments"*	"*Candidatus* Marithrix"	"*Candidatus* M. sessilis"	m	Presumably aerobic or microaerophilic	Filaments, attached or rosette-forming	10–96 µm	None	Yes	Yes	Yes / No	Yes	[34, 5]
XI	*Thioploca*	*Thioploca*	*T. schmidlei*	f, b	Microaerophilic	Multiple trichomes in a sheath	1–9 µm	Gliding	Yes	Yes	Yes / Yes	No	[25, 6,43,50,73, 74,84]
"XII"	*Beggiatoa*	*Beggiatoa*	*B. alba*	f, m, h	Aerobic or microaerophilic	Single filaments	1–9 µm	Gliding	Yes	Yes	Yes / Yes	No	[2,13,19,23,55, 69,70,77,81]

[a] f=freshwater, b=brackish, m=marine, h=hypersaline

In the original characterization of the latest, validly published genus of the *Beggiatoaceae*, *Thiomargarita* [63], cells were described as spheres arranged in chains. This morphology was exclusively found in **Cluster I** in this study, whereas the originally retrieved partial *T. namibiensis* sequence was assigned to Cluster III. This disagreement might have been caused by using a mixed sample of chain-forming round and chain-forming cylindrical cells for the amplification by Schulz *et al*. [63]. At that time, it was not assumed that the two chain-forming morphotypes were distinct phylotypes and therefore the obtained (partial) sequence was assigned to chain-forming *Thiomargarita* in general. In this study, however, we performed PCRs from both chain-forming morphotypes and indeed they are separate phylotypes, supposedly two different species. We propose to assign the sequences in Cluster I to the species *T. namibiensis* because this cluster contains the originally described spherical and chain-forming morphology of the validly published species *T. namibiensis.* **Clusters II and III** are monophyletic sister-groups of *T. namibiensis*. Maximum sequence identities between the three clusters range from 97.1 to 97.3%, while sequences within each single cluster are identical by 98.5–100% (Cluster I), 99.2–100% (Cluster II) and 98.7–100% (Cluster III). These clusters are thereby identified as separate species of the common genus *Thiomargarita.* We recommend naming the remaining two clusters after two extraordinary scientists devoting decades of research to the versatile group of large sulfur bacteria encompassing many milestones along the path: Bo Barker Jørgensen and Douglas C. Nelson. Therefore, the two novel species are suggested to be named "*Candidatus* Thiomargarita joergensenii" (Cluster II) and "*Candidatus* Thiomargarita nelsonii" (Cluster III).

Cluster IV of the 16S rRNA tree contains sequences that group with partial sequences of *T. araucae, T. chileae* and the full-length sequence of *Beggiatoa* sp. 'Bay of Concepcion' [49,72]. The morphology of the sequenced filaments of this cluster strongly resembles freshwater *Thioploca* spp. when trichomes occur in a bundle but is similar to *Beggiatoa* spp. when filaments occur as individuals. Yet, the 16S rRNA gene sequence data delineate this cluster from the validly published genera *Thioploca* and *Beggiatoa* at the genus level (≤93.9% sequence identity). According to morphological characteristics and recognizing their occurrence in marine environments, we propose the new genus name "*Candidatus* Marithioploca" for all organisms in Cluster IV. The separation of two species within this cluster was suggested elsewhere based on morphology [49]. Here, we confirm the finding of at least two groups within genus "*Candidatus* Marithioploca" based on different diameter ranges (12–18 μm and 25–37 μm, Fig. 1), however, the separation of two distinct

species is weakly supported due to very high identities in both their 16S rRNA (98.3–99.5%) and ITS (98.9%) sequences. We conclude that all organisms in Cluster IV should be subsumed into one species of the genus "*Candidatus* Marithioploca". As a consequence, we suggest transferring organisms formerly described as *T. araucae* and *T. chileae* into the new species "*Candidatus* Marithioploca araucae" (sic), according to the specific epithet proposed by Maier and Gallardo [49] but noting that the specific epiphet is grammatically incorrect [78].

Clusters V and VI are sister-groups comprising two novel genera of large sulfur bacteria. These clusters share 95.5–96.9% sequence identity with each other and ≤95.9% with the *T. namibiensis* cluster. **Cluster V** contains non-filamentous, spherical cells that were either found in colony-like aggregates or attached to benthic surfaces, often populating them densely [5]. According to the formation of colonial aggregates and high population densities on solid surfaces, we propose the genus and species name "*Candidatus* Thiopilula aggregata" for these organisms, meaning "little, aggregated sulfur balls".

Cluster VI contains sequences of exclusively spherical cells that delineated from the other clusters at the genus level. The first description of spherical, vacuolated cells, that are inconspicuously motile by a slow jerky rolling movement and contain sulfur inclusions, was given by Hinze in 1903 [24], who named them *Thiophysa volutans*. Later, van Niel reclassified this species as *A. volutans* [80]. The genus *Achromatium* belongs the family *Achromatiaceae* and therefore we propose to revive the genus name *Thiophysa* Hinze 1903 [24] for organisms grouping within Cluster VI of the *Beggiatoaceae*, according to Rule 33c of the International Code of Nomenclature of Bacteria [40]. Because of the lack of cultures, we are unable to retrieve the exact same sample of cells as Hinze. Therefore, we cannot be certain to include the same species that Hinze defined as *T. volutans*, nor can we confirm the genetic homogeneity of the sample that Hinze investigated. Apart from that, the epithet *volutans* suggests a putative motility of the cells, which can be confirmed as a frequent characteristic for all spherical sulfur bacteria so far investigated in this study, irrespective of being free-living, aggregated in diatom frustules or surrounded by a common envelope. Therefore, a rolling motility cannot be assigned to a certain taxon of non-filamentous, large sulfur bacteria. Accordingly, we recommend combining the revived genus name *Thiophysa* with a new species name, honouring the discoverer of these marvellous bacteria, G. Hinze, and propose the name "*Candidatus* Thiophysa hinzei" for organisms of this cluster.

The bacteria of **Clusters VII, VIII and IX** occur as free-living filaments in the sediment. Their 16S rRNA genes were sequenced in earlier studies and they have been classified as *Beggiatoa* spp. even though sequences of these clusters are different from the *Beggiatoa* type species and from each other at the genus level (identity values between the clusters are minimally between 85.3 and 93% and maximally between 88.1 and 94.1%). They bear a paraphyletic affiliation towards the *Beggiatoa* type species, *B. alba*, and therefore the conclusion is justified that bacteria of the Clusters VII, VIII and XI are genetically diverging genera exhibiting *Beggiatoa*-like properties. Therefore, they should no longer be considered as *Beggiatoa.* Instead, we propose the new genus names "*Candidatus* Maribeggiatoa", "*Candidatus* Isobeggiatoa" and "*Candidatus* Parabeggiatoa" and suggest transferring the species of the Clusters VII, VIII and IX into the new respective species "*Candidatus* Maribeggiatoa vulgaris" (meaning "usual or plentiful *Beggiatoa*-like organisms", Cluster VII), "*Candidatus* Isobeggiatoa divolgata" (meaning "widespread *Beggiatoa*-like organisms", Cluster VIII) and "*Candidatus* Parabeggiatoa communis" (meaning "common or general *Beggiatoa*-like organisms", Cluster IX). Sequences represented by a genus name, but appended by "sp." (Fig. 1) instead of the respective *Candidatus* species names, exhibited identities between 95.4 and 97.6% with the named species, which implies that the *Candidatus* genus possibly contains more than one species.

Filaments of **Cluster X** have not been classified subsequent to their discovery but a close morphological resemblance to *Thiothrix*-filaments was noted [35]. The 16S rRNA sequence data of this group were only 85.9% identical with the validly published *Thiothrix* spp., requiring the introduction of a new genus name. Recognizing both their marine habitat and their morphological similarity to *Thiothrix* spp., we propose "*Candidatus* Marithrix" as the new genus name. With respect to their observed life-mode, we propose "*Candidatus* Marithrix sessilis" as the name for the new species, meaning "sessile marine hair". Furthermore, this *Candidatus* genus is included in the monophylum of the family *Beggiatoaceae* (Fig. 1), sharing 87.1–94.5% sequence identity with other genera of this family, while it delineates from the genus *Thiothrix* at the family level (≤85.9% sequence identity).

Cluster XI comprises freshwater organisms morphologically resembling *T. schmidlei*, which was first discovered by Lauterborn in 1907 [43] in Lake Constance, Germany, and was defined as the type species of the genus *Thioploca*. The 16S rRNA gene of *T. schmidlei* was never sequenced and therefore cannot be used as a type sequence of the

genus. However, we suggest keeping the genus name *Thioploca* for sequences that are subsumed in Cluster XI and that originate from freshwater trichomes bundled in a common sheath. Teske *et al.* [73] were the first to sequence an organism of the genus *Thioploca* (accession number L40998), which they assigned to the species *T. ingrica* on a morphological basis [48, 83]. We suggest classifying all sequences affiliated to sequence L40998 as *T. ingrica.* Genus names appended by "sp." in this cluster exhibited 16S rRNA gene sequence identities ≤97.5% and an ITS sequence identity of 81% with *T. ingrica*, which implies that the genus possibly contains more than one species.

Table 2 summarizes the characteristics of the currently recognized genera and species of the *Beggiatoaceae* as well as the newly identified organisms added as *Candidatus* taxa to the family.

ITS confirms clusters defined by 16S rRNA data

The internal transcribed spacer (ITS) region can be used as a fast molecular chronometer with a good resolution of short-term phylogenetic evolution [28]. This is possible because of the lower evolutionary pressure on the non-coding parts of this spacer compared to the surrounding functional regions [6]. In this study, ITS identity values supported the clustering pattern derived from the 16S rRNA gene analysis. Indeed, clusters with individuals of putatively identical species showed high sequence identity values in the range of 94–100% at the ITS level (e.g. 95.9–100% in Cluster I or 94.1–100% in Cluster III), thereby confirming the close affiliation of the individuals sequenced. On the other hand, ITS identity values dropped to 85–89% between two putative species of the same genus, overlapping with ITS identity values of 75–89% between different genera. This finding reveals that the ITS region is not useful for higher-level classification. Instead, the ITS region is reliable for delineating fine-scale phylogeny on the intraspecies level, as reported earlier by others [45]. In most cases, where 16S rRNA genes were 100% identical, the ITS sequences were highly identical as well (99.1−100%) and in the majority of these cases the sequenced cells originated from the same geographical sampling area. Yet, also cells from different geographic origins shared identical 16S rRNA and ITS sequences, including two cells from Namibia and Costa Rica (NAM071 and COS012). Among cells from the South-Eastern Atlantic and the Mid-Pacific, we even found individuals 100% identical on the entire sequence obtained from the rRNA operon (identical 16S rRNA, ITS and insertion sequences). This finding implies that dispersal is faster than evolution within this species and that geographical barriers do not promote diversification.

Furthermore, it was an important observation that cells with different morphologies may feature 100% identical ITS sequences. This was detected for Cluster III with cells forming regularly arranged clusters and cells of cylindrical shape forming chains (both from Namibia), with attached cells and enveloped cells (from Costa Rica and Namibia), or for Cluster I with single cells and spherical, chain-forming cells (from Costa Rica and Namibia). This finding leads us to speculate about polymorphisms in some species or sub-species of sulfur bacteria. Yet, this hypothesis remains to be tested by temporal observations of single cells or better by culturing.

Towards an improved definition of the family *Beggiatoaceae*

Members of the family *Beggiatoaceae* are not only filamentous, as suggested earlier [44,71], but of highly variable morphology with complexities ranging from simple free-living cells, over chains and aggregates to highly structured filaments that even temporarily reside in a bundle. They vary greatly in size, starting with diameters of about 1 µm but reaching up to several hundreds of microns, making their largest representatives the largest known bacteria. Among the different sub-groups, morphologies may be conserved (e.g. *Beggiatoa*-like filaments), but occasionally differ greatly within a single species (e.g. *T. namibiensis*). Motility was noticed in all genera, although of different modes and velocities. Unique characteristics frequently occurring in the *Beggiatoaceae* are the ability to grow to gigantic dimensions (several 100 µm), the possibility to compartmentalize the cytoplasm via internal vacuolation (vacuoles can comprise up to 98% of the cellular biovolume) and the capacity to store nitrate (reaching close to 1 M). However, these features are not compulsory as filaments of Cluster X contain vacuoles without storing nitrate [34,35] and some narrow representatives of Cluster XI may even store nitrate without obvious vacuolation [25, 84]. Moreover, filaments from Cluster "XII" are usually narrow in diameter without vacuoles and nitrate storage. Only two sequences of Cluster "XII" can be assigned to wider, vacuolated, nitrate storing filaments [23]. Notably, the phylogeny of the organisms now subsumed in Cluster "XII" requires clarification in the future as the current data do not allow assessing whether the organisms might need to be segregated into different phylogenetic taxa with more restrictive characteristics.

Vacuolation, as mentioned above, seems to be a frequent property in the *Beggiatoaceae* and supposedly correlates with the diameter of the cell or filament. Most filamentous freshwater organisms have a small diameter (approximately 1–4 µm) and do not possess vacuoles, whereas marine species or those found in hypersaline environments can grow to

wider dimensions and are compartmentalized when their diameter exceeds roughly 5 µm [2,55,69]. The smallest marine non-filamentous sulfur bacterium observed in this study had a diameter of 9 µm and still contained a single central vacuole (data not shown). Few freshwater species like *T. schmidlei* and *T. ingrica* can grow to diameters >4 µm and in that case were found to contain several small or a large central vacuole [25,48,50,84]. As discussed earlier by Schulz and Jørgensen [62], vacuolation allows for large cell diameters while maintaining an efficient diffusion of solutes into the cytoplasm. In most cases, vacuolation of the cytoplasm was concomitant with nitrate storage [15,23,25,35,53,63]. In some species, vacuoles were reported to contain other compounds like polyphosphate [10] or possibly oxygen [34,35]. However, varieties and possible functions of vacuoles in members of the *Beggiatoaceae* or their association with specific habitat characteristics require further investigation.

Different genera of the *Beggiatoaceae* show correlations between their level of oxygen tolerance and motility. In the filamentous free-living or bundle-forming sulfur bacteria, low tolerance to oxygen is accompanied by the ability to withdraw from zones with sub-optimal conditions by gliding [31]. On the other hand, higher oxygen tolerance correlates with sporadic motility, as observed in the genera *Thiomargarita*, "*Candidatus* Thiopilula" and "*Candidatus* Thiophysa", or even a sessile life-style, as described for filamentous "*Candidatus* Marithrix sessilis" [35] and dimorphic forms of "*Candidatus* Thiomargarita nelsonii" and "*Candidatus* Thiopilula aggregata" [5].

Sessile ecotypes of the *Beggiatoaceae* share some properties with members of the *Leucotrichaceae*, including oxygen tolerance, production of gonidia and attachment to substrates or formation of rosettes. In all these organisms, a stationary life-mode seems to be complemented by the formation of a smaller and free-living offspring. These cells probably serve as a dispersal stage to populate more distant habitats. This life cycle has been known since the studies of Winogradsky [82] for the narrow *Leucotrichaceae*, but was only recently found in members of the wide *Beggiatoaceae*, more specifically in members of the species "*Candidatus* Marithrix sessilis" [35], "*Candidatus* Thiopilula aggregata" and "*Candidatus* Thiomargarita nelsonii" [5]. Whereas in the *Leucotrichaceae*, attached filaments produce gonidia that, after detachment, move by gliding [79] or show a jerky gliding motility [8], the definite motility of gonidia in the *Beggiatoaceae* still remains to be shown. In the well-studied *Leucotrichaceae*, gonidia production is always concomitant with rosette formation [79], whereas in the *Beggiatoaceae*, this correlation does not hold in

general: while gonidia production has been observed in rosette-forming filaments of "*Candidatus* Marithrix sessilis" [35], the sessile non-filamentous "*Candidatus* Thiopilula aggregata" and "*Candidatus* Thiomargarita nelsonii" produce gonidia from single cells attached to solid surfaces [5].

Finally, the members of the *Beggiatoaceae* are seemingly lacking a confined biogeographical distribution. Extremely similar or even identical sequences were detected in at least five different marine sites around the globe. This includes also the partial sequence of spherical sulfur bacteria from the Gulf of Mexico [33], which is 100% identical to some full-length sequences of this study from Namibia and Costa Rica. Already in previous studies, *Thioploca* filaments sampled from distant freshwater lakes in Germany and Japan revealed almost identical 16S rRNA sequences (>99.6% [36]). In conclusion, we propose that the habitats of the *Beggiatoaceae* are not isolated on the scale of evolution and that spatially separated upwelling areas or other habitats such as lakes are populated by similar cosmopolitan sulfur bacteria.

Etymology of the *Candidatus* taxa

"*Candidatus* Thiomargarita joergensenii"

Thi.o.mar.ga.ri'ta. Gr. neut. n. *theion* (Latin transliteration *thium*), sulfur; L. n. *margarita* pearl; N.L. fem. n. *Thiomargarita* sulfur pearl; joer.gen.se'ni.i. N.L. gen. n., *joergensenii* of Jørgensen, named in honor of Bo Barker Jørgensen, a Danish microbiologist.

Single, spherical cells; so far occurring in empty diatom frustules; sporadic slow jerky rolling movement; vacuolated; sulfur inclusions; marine; accession numbers FR690922– FR690925.

"*Candidatus* Thiomargarita nelsonii"

Thi.o.mar.ga.ri'ta. Gr. neut. n. *theion* (Latin transliteration *thium*), sulfur; L. n. *margarita* pearl; N.L. fem. n. *Thiomargarita* sulfur pearl; nel.so'ni.i. N.L. gen. n. *nelsonii* of Nelson, named in honor of Douglas C. Nelson, an American microbiologist.

Cells of highly diverse morphology and life-modes; ability to divide in multiple planes, to attach or to form gonidia; sporadic slow jerky rolling movement; vacuolated; sulfur inclusions; marine; accession number FR690926– FR690966, FN811658– FN811659, FN811661– FN811662.

"*Candidatus* Marithioploca araucae"

Ma.ri.thi.o.plo'ca. L. gen. n. *maris* of the sea; N.L. fem. n. *Thioploca* a genus name; N.L. fem. n. *Marithioploca* the *Thioploca* of the sea, the truly marine *Thioploca*; *T. araucae* [49], Approved Lists 1980; a.rau'ca.e. N.L. fem. adj. araucae of Arauco in Central Chile (sic).
Multicellular filaments either free-living filaments or bundled by a common mucus sheath; gliding motility; vacuolated; ability to store nitrate; sulfur inclusions; marine; accession numbers FR690987– FR690996, AF035956.

"*Candidatus* Thiopilula aggregata"

Thi.o.pi'lu.la. Gr. neut. n. *theion* (Latin transliteration *thium*), sulfur; L. fem. n. *pilula* little ball, little globule; N.L. fem. n. *Thiopilula* little sulfur ball; ag.gre.ga'ta. L. fem. adj. *aggregata* joined together.
Spherical cells aggregated in variable arrangements; ability to attach and form gonidia; sporadic slow jerky rolling movement; vacuolated; sulfur inclusions; marine; accession numbers FR690968– FR690980, FN811660, FN811664.

"*Candidatus* Thiophysa hinzei"

Thi.o.phy'sa. Gr. neut. n. *theion* (Latin transliteration *thium*), sulfur; Gr. fem. n. *physa* bubble, breath; N.L. fem n. *Thiophysa* sulfur bubble; hin'zei. N.L. gen. n. *hinzei* of Hinze; named in remembrance of G. Hinze, a German microbiologist, who first described marine large, spherical sulfur bacteria. The genus name *Thiophysa* was proposed by G. Hinze [24] for these marine sulfur bacteria. Yet, the name has never been validly published and the organisms were named "*Achromatium volutans*" [80], which was also never validly published in the Approved Lists. The name *Thiophysa* is revived according to Rule 33c of the International Code of Nomenclature of Bacteria.
Single, spherical cells; vacuolated; sporadic slow jerky rolling movement; sulfur inclusions; marine; accession numbers FR690982– FR690986.

"*Candidatus* Maribeggiatoa vulgaris"

Ma.ri.beg.gi.a.to'a. L. gen. n. *maris* of the sea; N.L. fem. n. *Beggiatoa* a genus name; N.L. fem. n. *Maribeggiatoa* the *Beggiatoa* of the sea, the truly marine *Beggiatoa*; vul.ga'ris. L. fem. adj. *vulgaris* usual, common.
Disc-shaped cells forming multicellular filaments; gliding motility; vacuolated; ability to store nitrate; sulfur inclusions; marine; at hydrothermal vents; accession numbers for "*Candidatus* Maribeggiatoa spp." F064543, FJ814745, FJ814753, AY580013, FJ14753.

"*Candidatus* Isobeggiatoa divolgata"

I.so.beg.gi.a.to'a. Gr. adj. *isos* equal, similar; N.L. fem. n. *Beggiatoa* a genus name; N.L. fem. n. *Isobeggiatoa* the similar *Beggiatoa*, the bacterium similar to *Beggiatoa*; di.vol.ga'ta. L. fem. adj. *divolgata* widespread, common.

Disc-shaped cells forming multicellular filaments; gliding motility; vacuolated; ability to store nitrate; sulfur inclusions; brackish or marine, also arctic latitudes; accession numbers for "*Candidatus* Isobeggiatoa spp." AF532769, AF532775, FJ875195, AB108786, FN561862.

"*Candidatus* Parabeggiatoa communis"

Pa.ra.beg.gi.a.to'a. Gr. prep. *para* beside, like; N.L. fem. n. *Beggiatoa* a genus name; N.L. fem. n. *Parabeggiatoa* resembling the genus *Beggiatoa*; com.mu'nis. L. fem. adj. *communis* common, widespread.

Disc-shaped cells forming multicellular filaments; gliding motility; vacuolated; ability to store nitrate; sulfur inclusions; brackish or marine; accession numbers for "*Candidatus* Parabeggiatoa spp." AF532770, AF532772–AF532774, FJ875196–FJ875199.

"*Candidatus* Marithrix sessilis"

Ma'ri.thrix. L. gen. n. *maris* of the sea; Gr. n. *thrix* hair; N.L. fem. n. *Marithrix* hair of the sea; ses'si.lis. L. adj. *sessilis* sitting, adhering to a surface.

Attached, multicellular filaments, sometimes forming rosettes; ability to produce gonidia; non-motile; sulfur inclusions; vacuolated or non-vacuolated; marine; at hydrothermal vents; accession numbers AY883933, AY883934, AY496953.

Acknowledgements

We would like to thank Jean Euzéby, Bernhard Fuchs, Bo Barker Jørgensen, Ramon Roselló-Mora, Douglas Nelson and Lars Peter Nielsen for valuable discussions and comments on the manuscript. We are also grateful to Wolfgang Ludwig for his help with the ARB software and sequence alignments, Anja Kamp, Gaute Lavik and Tim Ferdelman for providing samples from Namibia during LEG AHAB 4 cruise and R/V Meteor cruise M76, Lars Peter Nielsen for taking samples from the Grund Fjord, Stefanie Grünke for providing samples from the Amon Mud Volcano, Victoria Orphan, Lisa Levin, Greg Rouse, Tony Rathburn and the participants and crew of cruises AT 15-44 and AT 15-59 (supported by the National Science Foundation OCE-0826254 and OCE-0939557) for assisting with the collection of samples from Costa Rica. This study was funded by the Max Planck Society.

References

[1] Ahmad, A., Barry, J.P., Nelson, D.C. (1999) Phylogenetic affinity of a wide, vacuolate, nitrate-accumulating *Beggiatoa* sp. from Monterey Canyon, California, with *Thioploca* spp. Appl. Environ. Microbiol. 65, 270-277.

[2] Ahmad, A., Kalanetra, K.M., Nelson, D.C. (2006) Cultivated *Beggiatoa* spp. define the phylogenetic root of morphologically diverse, noncultured, vacuolate sulfur bacteria. Can. J. Microbiol. 52, 591-598.

[3] Aranda, C., Paredez, J., Valenzuela, C., Lam, P., Guillou, L.G. Coexistence of *Beggiatoa* spp., Delta-, Epsilonproteobacteria and *Fusibacter* in organically-enriched salmon farm sediment in southern Chile. EMBL/GenBank/DDBJ accession numbers FJ814745, FJ875195-99.

[4] Bailey, J.V., Joye, S.B., Kalanetra, K.M., Flood, B.E., Corsetti, F.A. (2007) Evidence of giant sulphur bacteria in Neoproterozoic phosphorites. Nature 445, 198-201.

[5] Bailey, J.V., Salman, V., Rouse, G., Schulz-Vogt, H.N., Levin, L., Orphan, V. Dimorphism in methane seep-dwelling ecotypes of the largest known bacteria, in preparation.

[6] Barry, T., Colleran, G., Glennon, M., Dunican, L.K., Gannon, F. (1991) The 16S/23S ribosomal spacer region as a target for DNA probes to identify eubacteria. PCR Methods Appl. 1, 51-56.

[7] Bersa, E. (1920) Über das Vorkommen von kohlensaurem Kalk in einer Gruppe von Schwefelbakterien. Sitzungsbericht Akademie der Wissenschaften, mathematisch-naturwissenschaftliche Klasse, I Abteilung, Wien

[8] Bland, J.A., Brock, T.D. (2005) Genus IV. *Leucothrix* Oersted. In: Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey´s Manual of Systematic Bacteriology, Springer, New York, Edition 2, Vol. 2, pp. 162-168.

[9] Brock, J., Schulz-Vogt, H.N. (2011) Sulfide induces phosphate release from polyphosphate in cultures of a marine *Beggiatoa* strain. ISME J. 5, 497-506.

[10] Brock, J. (2011) Impact of sulfide-oxidizing bacteria on the phosphorus cycle in marine sediments. Thesis, University of Bremen, Bremen, Germany.

[11] Cardinale, M., Brusetti, L., Quatrini, P., Borin, S., Puglia, A.M., Rizzi, A., Zanardini, E., Sorlini, C., Corselli, C., Daffonchio, D. (2004) Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. Appl. Environ. Microbiol. 70, 6147-6156.

[12] Dando, P.R., Hughes, J.A., Thiermann, F. (1995) Preliminary observations on biological communities at shallow hydrothermal vents in the Aegean Sea. In: Parson,

L.M., Walker, C.L., Dixon, D.L. (Eds.), Hydrothermal Vents and Processes, London, pp. 303-317.

[13] de Albuquerque, J.P., Keim, C.N., Lins, U. (2010) Comparative analysis of *Beggiatoa* from hypersaline and marine environments. Micron 41, 507-517.

[14] Euzéby, J.P. List of prokaryotic names with standing in nomenclature. http://www.bacterio.cict.fr/ Last updated 12th January 2011.

[15] Fossing, H., Gallardo, V.A., Jørgensen, B.B., Huttel, M., Nielsen, L.P., Schulz, H., Canfield, D.E., Forster, S., Glud, R.N., Gundersen, J.K., Kuver, J., Ramsing, N.B., Teske, A., Thamdrup, B., Ulloa, O. (1995) Concentration and transport of nitrate by the mat-forming sulfur bacterium *Thioploca*. Nature 374, 713-715.

[16] Gallardo, V.A. (1977) Large benthic microbial communities in sulfide biota under Peru-Chile subsurface countercurrent. Nature 268, 331-332.

[17] Garrity, G.M., Bell, J.A., Lilburn, T. (2005) Family I. *Thiotrichaceae* fam. nov. In: Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey´s Manual of Systematic Bacteriology, Springer, New York, Edition 2, Vol. 2, pp. 131.

[18] Girnth, A.-C., Grünke, S., Lichtschlag, A., Felden, J., Knittel, K., Wenzhöfer, F., de Beer, D., Boetius, A. (2011) A novel, mat-forming *Thiomargarita* population associated with a sulfidic fluid flow from a deep-sea mud volcano. Environ. Microbiol. 13, 495-505.

[19] Grabovich, M.Y., Patritskaya, V.Y., Muntyan, M.S., Dubinina, G.A. (2001) Lithoautotrophic growth of the freshwater strain *Beggiatoa* D-402 and energy conservation in a homogeneous culture under microoxic conditions. FEMS Microbiol. Lett. 204, 341-345.

[20] Guezennec, J., Ortega-Morales, O., Raguenes, G., Geesey, G. (1998) Bacterial colonization of artificial substrate in the vicinity of deep-sea hydrothermal vents. FEMS Microbiol. Ecol. 26, 89-99.

[21] Harold, R., Stanier, R.Y. (1955) The genera *Leucothrix* and *Thiothrix*. Bacteriol. Rev. 19, 49-64.

[22] Head, I.M., Gray, N.D., Clarke, K.J., Pickup, R.W., Jones, J.G. (1996) The phylogenetic position and ultrastructure of the uncultured bacterium *Achromatium oxaliferum*. Microbiol. 142, 2341-2354.

[23] Hinck, S., Neu, T.R., Lavik, G., Mussmann, M., De Beer, D., Jonkers, H.M. (2007) Physiological adaptation of a nitrate-storing *Beggiatoa* sp. to diel cycling in a phototrophic hypersaline mat. Appl. Environ. Microbiol. 73, 7013-7022.

[24] Hinze, G. (1903) *Thiophysa volutans*, ein neues Schwefelbakterium. Berichte der deutschen botanischen Gesellschaft, Heft 6,

[25] Høgslund, S., Nielsen, J.L., Nielsen, L.P. (2010) Distribution, ecology and molecular identification of *Thioploca* from Danish brackish water sediments. FEMS Microbiol. Ecol. 73, 110-120.

[26] Jacq, E., Prieur, D., Nichols, P., White, D.C., Porter, T., Geesey, G.G. (1989) Microscopic examination and fatty-acid characterization of filamentous bacteria colonizing substrata around subtidal hydrothermal vents. Arch. Microbiol. 152, 64-71.

[27] Jannasch, H.W., Nelson, D.C., Wirsen, C.O. (1989) Massive natural occurrence of unusually large bacteria (*Beggiatoa* sp.) at a hydrothermal deep-sea vent site. Nature 342, 834-836.

[28] Jensen, M.A., Webster, J.A., Straus, N. (1993) Rapid identification of bacteria on the basis of polymerase chain reaction - amplified ribosomal DNA spacer polymorphisms. Appl. Environ. Microbiol. 59, 945-952.

[29] Jørgensen, B.B. (1977) Distribution of colorless sulfur bacteria (*Beggiatoa* spp.) in a coastal marine sediment. Mar. Biol. 41, 19-28.

[30] Jørgensen, B.B., Dunker, R., Grünke, S., Roy, H. (2010) Filamentous sulfur bacteria, *Beggiatoa* spp., in arctic marine sediments (Svalbard, 79 degrees N). FEMS Microbiol. Ecol. 73, 500-513.

[31] Jørgensen, B.B., Gallardo, V.A. (1999) *Thioploca* spp.: filamentous sulfur bacteria with nitrate vacuoles. FEMS Microbiol. Ecol. 28, 301-313.

[32] Jørgensen, B.B., Teske, A., Ahmad, A. (2005) Genus VII. *Thioploca* Lauterborn. In: Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey´s Manual of Systematic Bacteriology, Springer, New York, Edition 2, Vol. 2, pp. 171-178.

[33] Kalanetra, K.M., Joye, S.B., Sunseri, N.R., Nelson, D.C. (2005) Novel vacuolate sulfur bacteria from the Gulf of Mexico reproduce by reductive division in three dimensions. Environ. Microbiol. 7, 1451-1460.

[34] Kalanetra, K.M., Nelson, D.C. (2010) Vacuolate-attached filaments: highly productive *Ridgeia piscesae* epibionts at the Juan de Fuca hydrothermal vents. Mar. Biol. 157, 791-800.

[35] Kalanetra, K.M., Huston, S.L., Nelson, D.C. (2004) Novel, attached, sulfur-oxidizing bacteria at shallow hydrothermal vents possess vacuoles not involved in respiratory nitrate accumulation. Appl. Environ. Microbiol. 70, 7487-7496.

[36] Kojima, H., Teske, A., Fukui, M. (2003) Morphological and phylogenetic characterizations of freshwater *Thioploca* species from Lake Biwa, Japan, and Lake Constance, Germany. Appl. Environ. Microbiol. 69, 390-398.

[37] Kojima, H., Koizumi, Y., Fukui, M. (2006) Community structure of bacteria associated with sheaths of freshwater and brackish *Thioploca* species. Microb. Ecol. 52, 765-773.

[38] Kojima, H., Fukui, M. (2003) Phylogenetic analysis of *Beggiatoa* spp. from organic rich sediment of Tokyo Bay, Japan. Water Res. 37, 3216-3223.

[39] Koppe, F. (1924) Die Schlammflora der ostholsteinischen Seen und des Bodensees. Arch. Hydrobiol. Band XIV, 619-675.

[40] Lapage, S.P., Sneath, P.H.A., Lessel, E.F., Skerman, V.B.D., Seeliger, H.P.R., Clark, W.A. (1992) International Code of Nomenclature of Bacteria: Bacteriological Code. ASM Press, Washington (DC)

[41] Larkin, J.M., Henk, M.C. (1996) Filamentous sulfide-oxidizing bacteria at hydrocarbon seeps of the Gulf of Mexico. Microsc. Res. Tech. 33, 23-31.

[42] Larkin, J.M., Shinabarger, D.L. (1983) Characterization of *Thiothrix nivea*. Int. J. Syst. Bacteriol. 33, 841-846.

[43] Lauterborn, R. (1907) Eine neue Gattung der Schwefelbakterien (*Thioploca schmidlei* nov. gen. nov. spec.). Berichte der deutschen botanischen Gesellschaft 25, 238-242.

[44] Leadbetter, E.R. (1974) Family II. *Beggiatoaceae*. In: Buchanan, R.E., Gibbons, N.E. (Eds.), Bergey´s Manual of Determinative Bacteriology, The Williams & Wilkins Company, Baltimore, Edition 8, pp. 112-116.

[45] Leblond-Bourget, N., Philippe, H., Mangin, I., Decaris, B. (1996) 16S rRNA and 16S to 23S internal transcribed spacer sequence analyses reveal inter- and intraspecific *Bifidobacterium* phylogeny. Int. J. Syst. Bacteriol. 46, 102-111.

[46] Ludwig, W., Rossello-Mora, R., Aznar, R., Klugbauer, S., Spring, S., Reetz, K., Beimfohr, C., Brockmann, E., Kirchhof, G., Dorn, S., Bachleitner, M., Klugbauer, N., Springer, N., Lane, D., Nietupsky, R., Weizenegger, M., Schleifer, K.H. (1995) Comparative sequence analysis of 23S ribosomal RNA from Proteobacteria. Syst. Appl. Microbiol. 18, 164-188.

[47] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A.,

Lenke, M., Ludwig, T., Bode, A., Schleifer, K.-H. (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363-1371.

[48] Maier, S. (1984) Description of *Thioploca ingrica* sp. nov., nom. rev. Int. J. Syst. Bacteriol. 34, 344-345.

[49] Maier, S., Gallardo, V.A. (1984) *Thioploca araucae* sp. nov. and *Thioploca chileae* sp. nov. Int. J. Syst. Bacteriol. 34, 414-418.

[50] Maier, S., Murray, R.G.E. (1965) Fine structure of *Thioploca ingrica* and a comparison with *Beggiatoa*. Can. J. Microbiol. 11, 645-655.

[51] Maier, S. (1989) Genus III. *Thioploca* Lauterborn 1907. In: Staley, J.T., Bryant, M.P., Pfennig, N., Holt, J.G. (Eds.), Bergey´s Manual of Systematic Bacteriology, The Williams and Wilkins Company, Baltimore, Edition 1, Vol. 3, pp. 2101-2106.

[52] McHatton, S.C., Barry, J.P., Jannasch, H.W., Nelson, D.C. (1996) High nitrate concentrations in vacuolate, autotrophic marine *Beggiatoa* spp. Appl. Environ. Microbiol. 62, 954-958.

[53] Mussmann, M., Schulz, H.N., Strotmann, B., Kjaer, T., Nielsen, L.P., Rosselló-Mora, R.A., Amann, R.I., Jørgensen, B.B. (2003) Phylogeny and distribution of nitrate-storing *Beggiatoa* spp. in coastal marine sediments. Environ. Microbiol. 5, 523-533.

[54] Nadson, G.A. (1913) Ueber Schwefelmikroorganismen des Hapsaler Meerbusens. Bulletin du Jardin Impériale Botanique de St.-Pétersbourg

[55] Nelson, D.C., Waterbury, J.B., Jannasch, H.W. (1982) Nitrogen-fixation and nitrate utilization by marine and freshwater *Beggiatoa*. Arch. Microbiol. 133, 172-177.

[56] Nelson, D.C., Wirsen, C.O., Jannasch, H.W. (1989) Characterization of large, autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas Basin. Appl. Environ. Microbiol. 55, 2909-2917.

[57] Nelson, D.C., Castenholz, R.W. (1981) Organic nutrition of *Beggiatoa* sp. J. Bacteriol. 147, 236-247.

[58] Oersted, A.S. (1844) De Regionibus marinis. Elementa Topographiae Historiconaturalis Freti Oeresund. Thesis, University of Copenhagen, Copenhagen.

[59] Peplies, J., Kottmann, R., Ludwig, W., Glöckner, F.O. (2008) A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. Syst. Appl. Microbiol. 31, 251-257.

[60] Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J., Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188-7196.

[61] Schewiakoff, W. (1892) Ueber einen neuen bacterienähnlichen Organismus des Süsswassers. Thesis, University Heidelberg, Heidelberg.

[62] Schulz, H.N., Jørgensen, B.B. (2001) Big bacteria. Annu. Rev. Microbiol. 55, 105-137.

[63] Schulz, H.N., Brinkhoff, T., Ferdelman, T.G., Marine, M.H., Teske, A., Jørgensen, B.B. (1999) Dense populations of a giant sulfur bacterium in Namibian shelf sediments. Science 284, 493-495.

[64] Schulz, H.N., Jørgensen, B.B., Fossing, H.A., Ramsing, N.B. (1996) Community structure of filamentous, sheath-building sulfur bacteria, *Thioploca* spp., off the coast of Chile. Appl. Environ. Microbiol. 62, 1855-1862.

[65] Schulz, H.N. (2006) The genus *Thiomargarita*. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes, Springer, Edition 3, Vol. 6, pp. 1156-1163.

[66] Skerman, V.B.D., McGowan, V., Sneath, P.H.A. (1980) Approved lists of bacterial names. Int. J. Syst. Bacteriol. 30, 225-420.

[67] Skerman, V.B.D., McGowan, V., Sneath, P.H.A. (1989) Approved lists of bacterial names (amended edition). In: American Society for Microbiology, Washington, DC, pp. 1-298.

[68] Stein, J.L. (1984) Subtidal gastropods consume sulfur-oxidizing bacteria - evidence from coastal hydrothermal vents. Science 223, 696-698.

[69] Strohl, W.R., Larkin, J.M. (1978) Enumeration, isolation, and characterization of *Beggiatoa* from freshwater sediments. Appl. Environ. Microbiol. 36, 755-770.

[70] Strohl, W.R., Cannon, G.C., Shively, J.M., Gude, H., Hook, L.A., Lane, C.M., Larkin, J.M. (1981) Heterotrophic carbon metabolism by *Beggiatoa alba*. J. Bacteriol. 148, 572-583.

[71] Strohl, W.R. (1989) Family I. *Beggiatoaceae*. In: Staley, J.T., Bryant, M.P., Pfennig, N., Holt, J.G. (Eds.), Bergey´s Manual of Systematic Bacteriology, Williams & Wilkins, Baltimore, Edition 1, Vol. 3, pp. 2089-2106.

[72] Teske, A., Sogin, M.L., Nielsen, L.P., Jannasch, H.W. (1999) Phylogenetic relationships of a large marine *Beggiatoa*. Syst. Appl. Microbiol. 22, 39-44.

[73] Teske, A., Ramsing, N.B., Küver, J., Fossing, H. (1995) Phylogeny of *Thioploca* and related filamentous sulfide-oxidizing bacteria. Syst. Appl. Microbiol. 18, 517-526.

[74] Teske, A., Nelson, D.C. (2006) The genera *Beggiatoa* and *Thioploca*. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes, Springer, Edition 3, Vol. 6, pp. 784-810.

[75] Thar, R., Fenchel, T. (2005) Survey of motile microaerophilic bacterial morphotypes in the oxygen gradient above a marine sulfidic sediment. Appl. Environ. Microbiol. 71, 3682-3691.

[76] Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

[77] Trevisan, V. (1842) Prospetto della Flora Euganea. Coi Tipi del Seminario, Padua

[78] Trüper, H.G., Euzeby, J.P. (2009) International code of nomenclature of prokaryotes. Appendix 9: Orthography. Int. J. Syst. Evol. Microbiol. 59, 2107-2113.

[79] Unz, R.F., Head, I.M. (2005) Genus I. *Thiothrix* Winogradsky. In: Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey´s Manual of Systematic Bacteriology, Springer, Baltimore, Edition 2, Vol. 2, pp. 131-141.

[80] Van Niel, C.B. (1948) Family A. *Achromatiaceae* Massart. In: Breed, R.S., Murray, E.G.D., Hitchens, A.P. (Eds.), Bergey´s Manual of Determinative Bacteriology, The Williams and Wilkins Company, Baltimore, Edition 6, pp. 997-999.

[81] Vaucher, J.P. (1803) Histoire des conferves d´eau douce, contenant leurs différents modes de reproduction, et la description de leurs principales espèces. Paschoud, Geneva

[82] Winogradsky, S. (1888) Beiträge zur Morphologie und Physiologie der Bacterien, Heft 1. Zur Morphologie und Physiologie der Schwefelbacterien, Leipzig

[83] Wislouch, S.M. (1912) *Thioploca ingrica* nov. sp. Berichte der deutschen botanischen Gesellschaft 30, 470-474.

[84] Zemskaya, T.I., Namsaraev, B.B., Dul'tseva, N.M., Khanaeva, T.A., Golobokova, L.P., Dubinina, G.A., Dulov, L.E., Wada, E. (2001) Ecophysiological characteristics of the mat-forming bacterium *Thioploca* in bottom sediments of the Frolikha Bay, northern Baikal. Microbiol. 70, 335-341.

Supplemental Information

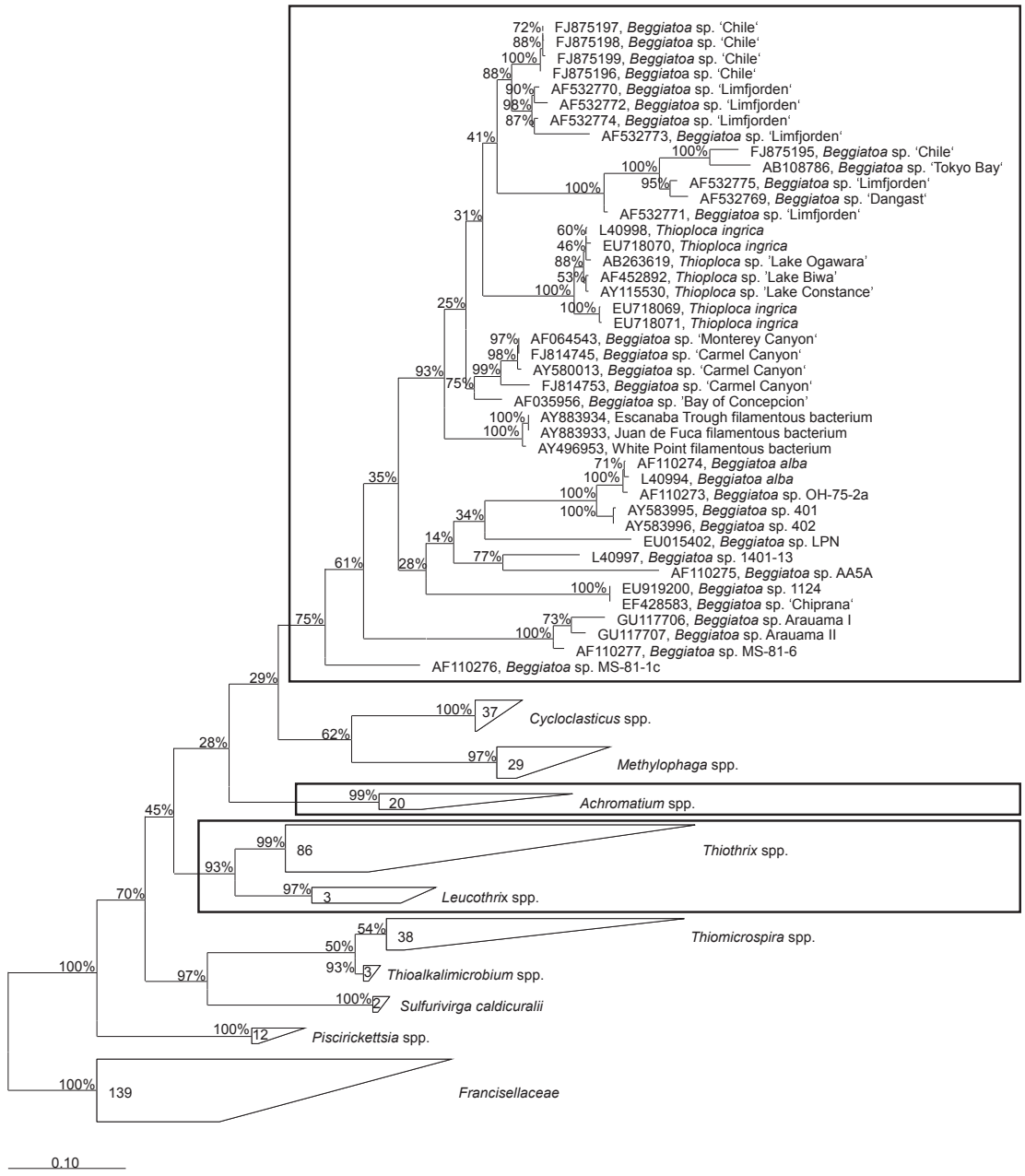


Figure S1. Maximum likelihood tree of all available nearly full-length 16S rRNA sequences of the large, colorless sulfur bacteria and their closest relatives. Percentage numbers on the branches indicate bootstrap values of 1000 runs. The classification is based on the current nomenclature. Boxes mark the currently available sequences of genera, which were recently combined into one family, the *Thiotrichaceae*, by Garrity et al. [3]. The branching pattern shown here, however, suggests that the groups should be classified according to the previous classification [5,7-9] that knows three distinct families, the *Beggiatoaceae*, *Achromatiaceae* and *Leucotrichaceae* (boxes from top to bottom). Each of them is a clearly separated monophyletic group supported by a high bootstrap value (75%, 99% and 93%, respectively). Therefore, it is recommended to retain the original classification of three separate families.

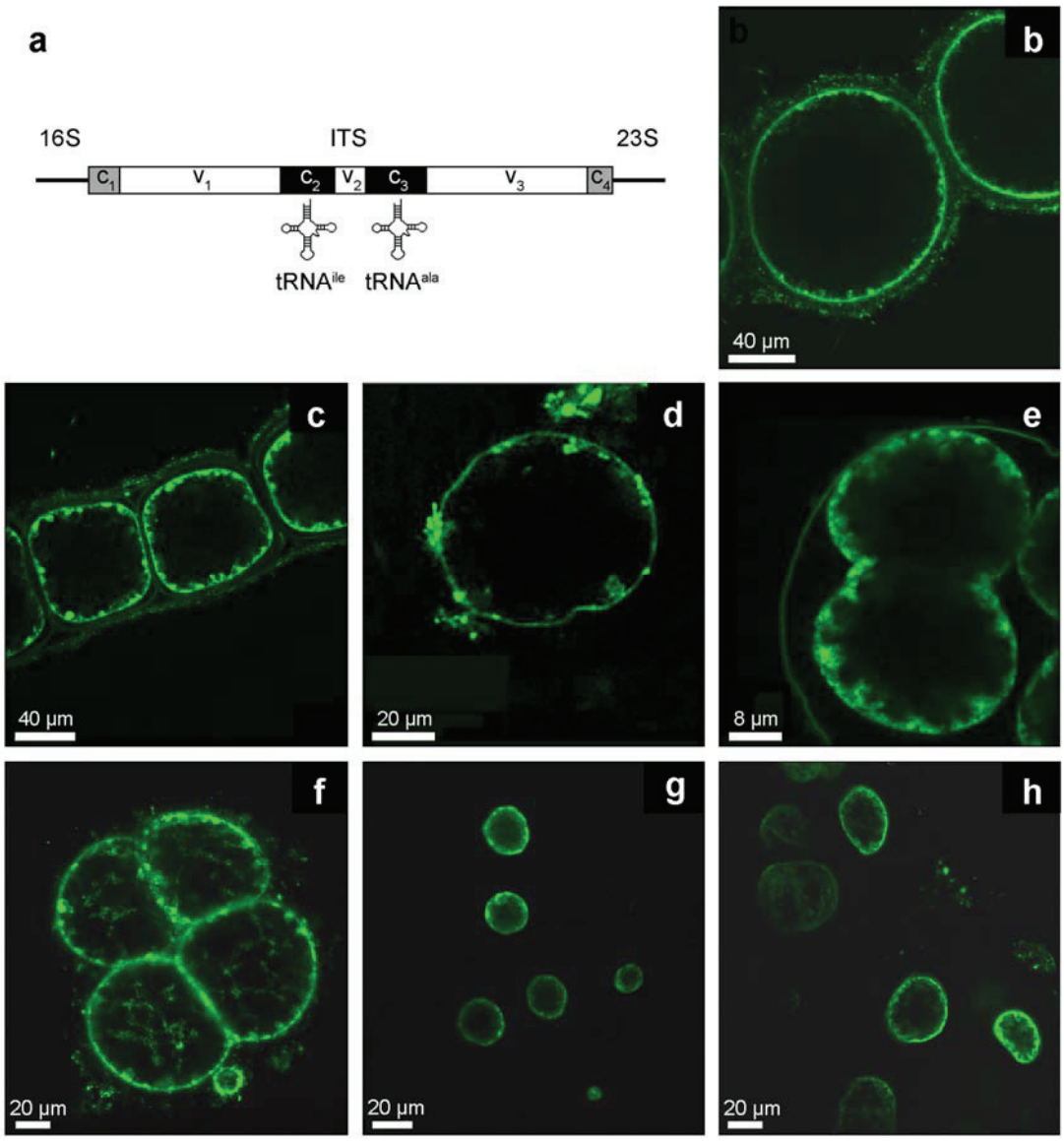


Figure S2. (a) Schematic diagram of the ITS region of the *Beggiatoaceae* divided into conserved (grey, c1 and c4) and variable regions (white, v1, v2 and v3), with special emphasis on conserved regions coding for tRNA genes (black, c2 and c3). (b–h) FITC staining of the cytoplasmic rim surrounding the central vacuole in the large sulfur bacteria morphotypes: chain-forming sulfur bacteria with round (b) and cylindrical-shaped (c) cells, a single free-living bacterium (d), spherical cells aggregated in an envelope (e), cells aggregated regularly forming a cube-like structure (f), cells released from a diatom frustule (g) and bacteria aggregated randomly within a thick mucus matrix (h).

Table S1. Primers used for the phylogenetic analysis of the large sulfur bacteria. Position numbering is according to the 16S/23S rRNA gene numbering of *E. coli*.

primer	position	sequence	target gene	reference	specificity
VSO233F	233−251	5´-CCTATGCCGGATTAGCTTG-3´	16S rDNA	this study	*Thiomargarita* spp., "*Candidatus* Marithioploca", "*Candidatus* Thiophysa", "*Candidatus* Maribeggiatoa", "*Candidatus* Marithrix", "*Candidatus* Parabeggiatoa"
AMV442F	442−461	5´-GTTGGGAAGAAAAGCTTTAG-3´	16S rDNA	this study	*Thiomargarita namibiensis*, "*Candidatus* Marithioploca"
VSO86F	86−105	5´-GGGGTGATGACGAGTGG-3´	16S rDNA	this study	*Thiomargarita* spp., "*Candidatus* Thioploca", "*Candidatus* Thiopilula", "*Candidatus* Marithrix", "*Candidatus* Parabeggiatoa", "*Candidatus* Maribeggiatoa"
VSO1300R	1291−1307	5´-ATCCGGACTACGAGTAG-3´	16S rDNA	this study	*Beggiatoaceae* excluding *Thioploca* spp. and *Beggiatoa* spp.
FWTpl131F	131−148	5´-TCTGCCTTGTAGTGGGGA-3´	16S rDNA	this study	*Thioploca* spp., "*Candidatus* Thiomargarita nelsonii"
VSO673F	656−673	5´-GTACAGTAGAGGGAAGCG-3´	16S rDNA	[4]	"*Candidatus* Thiomargarita nelsonii", "*Candidatus* Parabeggiatoa", "*Candidatus* Marithioploca", "*Candidatus* Maribeggiatoa", "*Candidatus* Thiopilula", "*Candidatus* Thiophysa", *Thioploca* spp.
VSO360F	361−379	5´- GGAATATTGGACAATGGGC-3´	16S rDNA	[1]	*Proteobacteria*, *Bacteroidetes*, *Actinobacteria*
GM3F	8−23	5´-AGAGTTTGATCMTGGC-3´	16S rDNA	[6]	*Eubacteria*
GM4R	1492−1507	5´-TACCTTGTTACGACTT-3´	16S rDNA	[6]	*Eubacteria*
ITSReub	23−38	5´-GCCAAGGCATCCACC-3´	23S rDNA	[2]	*Eubacteria*

Table S2. Summary of all individual cells and filaments sequenced in this study.

name	morphology	cell / filament diameter [µm]	sample origin	primers utilized	PCR product length (rounded)	position of insertion(s)	accession number
Thiomargarita namibiensis NAM001	chain of round cells	141	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690879
Thiomargarita namibiensis NAM002	chain of round cells	172	22° 59.87´S 14° 3.12´E	233F / ITSReub	2900 bp	1078, 1396	FR690880
Thiomargarita namibiensis NAM003	chain of round cells	164	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690881
Thiomargarita namibiensis NAM004	chain of round cells	136	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690882
Thiomargarita namibiensis NAM005	chain of round cells	122	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690883
Thiomargarita namibiensis NAM006	chain of round cells	160	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690884
Thiomargarita namibiensis NAM007	chain of round cells	148	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690885
Thiomargarita namibiensis NAM008	chain of round cells	145	25° 00.04´S 14° 23.36´E	233F / ITSReub	2900 bp	1078, 1396	FR690886
Thiomargarita namibiensis NAM009	round double-cell	131	19° 1.01´S 12° 13.75´E	233F / ITSReub	2900 bp	1078, 1396	FR690887
Thiomargarita namibiensis NAM010	chain of round cells	121	19° 48.83´S 12° 46.37´E	233F / ITSReub	2900 bp	1078, 1396	FR690888
Thiomargarita namibiensis NAM011	single, round cell	146	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690889
Thiomargarita namibiensis NAM012	chain of round cells	122	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690890
Thiomargarita namibiensis NAM013	round cells, lump-forming	70	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690891
Thiomargarita namibiensis NAM014	round cells, lump-forming	532	24° 03.19´S 14° 15.69´E	233F / ITSReub	2900 bp	1078, 1396	FR690892
Thiomargarita namibiensis NAM015	single, round cell	47	24° 03.19´S 14° 15.69´E	233F / ITSReub	2300 bp	1396	FR690893
Thiomargarita namibiensis NAM016	single, round cell	40	24° 03.19´S 14° 15.69´E	233F / ITSReub	2300 bp	1396	FR690894
Thiomargarita namibiensis NAM017	single, round cell	64	24° 03.19´S 14° 15.69´E	233F / ITSReub	2300 bp	1396	FR690895
Thiomargarita namibiensis NAM018	single, round cell	52	24° 03.19´S 14° 15.69´E	233F / ITSReub	2300 bp	1396	FR690896
Thiomargarita namibiensis NAM019	single, round cell	65	24° 03.19´S 14° 15.69´E	233F / ITSReub	2300 bp	1396	FR690897
Thiomargarita namibiensis NAM020	single, round cell	45	24° 17.15´S 14° 16.10´E	233F / ITSReub	2300 bp	1396	FR690898
Thiomargarita namibiensis NAM021	round cells in mucus aggregate	40	24° 03.19´S 14° 15.69´E	233F / ITSReub	3500 bp	795, 1495	FR690899
Thiomargarita namibiensis NAM022	round cells in mucus aggregate	52	24° 03.19´S 14° 15.69´E	233F / ITSReub	3500 bp	795, 1495	FR690900
Thiomargarita namibiensis NAM023	round cells in mucus aggregate	50	24° 03.19´S 14° 15.69´E	233F / ITSReub	3500 bp	795, 1495	FR690901
Thiomargarita namibiensis NAM024	round cells in mucus aggregate	40	24° 03.19´S 14° 15.69´E	233F / ITSReub	3500 bp	795, 1495	FR690902
Thiomargarita namibiensis NAM025	round cells in mucus aggregate	47	24° 17.15´S 14° 16.10´E	233F / ITSReub	3500 bp	795, 1495	FR690903
Thiomargarita namibiensis NAM026	round cells in mucus aggregate	25-88	25° 00.04´S 14° 23.36´E	233F / ITSReub	3500 bp	795, 1495	FR690904
Thiomargarita namibiensis NAM027	round cells in mucus aggregate	44	24° 03.19´S 14° 15.69´E	233F / ITSReub	3500 bp	795, 1495	FR690905
Thiomargarita namibiensis NAM028	round cells in mucus aggregate	48	24° 03.19´S 14° 15.69´E	233F / ITSReub	3900 bp	795, 1396, 1495	FR690906
Thiomargarita namibiensis NAM029	round cells in mucus aggregate	50	24° 03.19´S 14° 15.69´E	233F / ITSReub	3900 bp	795, 1396, 1495	FR690907
Thiomargarita namibiensis NAM030	round cells in mucus aggregate	37	24° 03.19´S 14° 15.69´E	233F / ITSReub	3900 bp	795, 1396, 1495	FR690908
Thiomargarita namibiensis NAM031	round cells in mucus aggregate	40	24° 03.19´S 14° 15.69´E	233F / ITSReub	3900 bp	795, 1396, 1495	FR690909
Thiomargarita namibiensis COS001	single round cell	115	8° 55.8´N 84° 18.75´W	233F/ITSReub	1800 bp	-	FN811663
Thiomargarita namibiensis COS002	single round cell	123	8° 55.8´N 84° 18.75´W	233F/ITSReub	3400 bp	795, 1396	FR690910

Table S2 (continued)

Thiomargarita namibiensis COS003	single round cell	40	8° 55.8´N 84° 18.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690911
Thiomargarita namibiensis COS004	single round cell	44	8° 55.8´N 84° 18.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690912
Thiomargarita namibiensis COS005	single round cell	100	8° 55.8´N 84° 18.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690913
Thiomargarita namibiensis COS006	single round cell	82	8° 55.8´N 84° 18.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690914
Thiomargarita namibiensis COS007	round double cell	50	8° 55.8´N 84° 18.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690915
Thiomargarita namibiensis C□I001	single round cell	37	36□30.80´S 73□7.75´W	233F/ITSReub	2600 bp	1396	FR690916
Thiomargarita namibiensis C□I002	single round cell	70	36□30.80´S 73□7.75´W	233F/ITSReub	3500 bp	795, 1396	FR690917
Thiomargarita namibiensis C□I003	single round cell	126	36□30.80´S 73□7.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690918
Thiomargarita namibiensis C□I004	single round cell	118	36□30.80´S 73□7.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690919
Thiomargarita namibiensis C□I005	single round cell	92	36□30.80´S 73□7.75´W	233F/ITSReub	2900 bp	1078, 1396	FR690920
Thiomargarita namibiensis AM□001	single round cell	ca. 47	32° 22.04´N 31° 42.16´E	233F/1300R and AM□442F/ITSReub	3500 bp	795, 1078, 1495	FR690921
"Candidatus Thiomargarita □oergensenii□ NAM032	round cells in diatom frustule	56	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1078	FR690922
"Candidatus Thiomargarita □oergensenii□NAM033	round cells in diatom frustule	NA	19° 1.01´S 12° 13.75´E	233F / ITSReub	2600 bp	1078	FR690923
"Candidatus Thiomargarita □oergensenii□NAM034	round cells in diatom frustule	17-38	24° 17.15´S 14° 16.10´E	233F / ITSReub	2600 bp	1078	FR690924
"Candidatus Thiomargarita □oergensenii□NAM035	round cells in diatom frustule	22-53	24° 03.19´S 14° 15.69´E	233F / ITSReub	3800 bp	795, 1078, 1495	FR690925
"Candidatus Thiomargarita nelsonii□NAM036	chain of barrel-shaped cells	117	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690926
"Candidatus Thiomargarita nelsonii□NAM037	chain of barrel-shaped cells	39	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690927
"Candidatus Thiomargarita nelsonii□NAM038	chain of barrel-shaped cells	96	25° 00.04´S 14° 23.36´E	233F / ITSReub	2600 bp	1396	FR690928
"Candidatus Thiomargarita nelsonii□NAM039	chain of barrel-shaped cells	63	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FN811658
"Candidatus Thiomargarita nelsonii□NAM040	chain of barrel-shaped cells	45	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690929
"Candidatus Thiomargarita nelsonii" NAM041	chain of barrel-shaped cells	58	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690930
"Candidatus Thiomargarita nelsonii" NAM042	chain of barrel-shaped cells	80	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690931
"Candidatus Thiomargarita nelsonii" NAM043	chain of barrel-shaped cells	125	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690932
"Candidatus Thiomargarita nelsonii" NAM044	chain of barrel-shaped cells	90	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690933
"Candidatus Thiomargarita nelsonii" NAM045	chain of barrel-shaped cells	100	19° 48.83´S 12° 46.37´E	233F / ITSReub	2600 bp	1396	FR690934
"Candidatus Thiomargarita nelsonii" NAM046	chain of barrel-shaped cells	120	24° 17.15´S 14° 16.10´E	233F / ITSReub	2600 bp	1396	FR690935
"Candidatus Thiomargarita nelsonii" NAM047	chain of barrel-shaped cells	158	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690936
"Candidatus Thiomargarita nelsonii" NAM048	chain of barrel-shaped cells	85	19° 1.01´S 12° 13.75´E	233F / ITSReub	2600 bp	1396	FR690937
"Candidatus Thiomargarita nelsonii" NAM049	chain of barrel-shaped cells	182	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690938
"Candidatus Thiomargarita nelsonii" NAM050	chain of barrel-shaped cells	61	19° 1.01´S 12° 13.75´E	233F / ITSReub	2600 bp	1396	FR690939
"Candidatus Thiomargarita nelsonii" NAM051	chain of barrel-shaped cells	250	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690940
"Candidatus Thiomargarita nelsonii" NAM052	chain of barrel-shaped cells	105	24° 17.15´S 14° 16.10´E	233F / ITSReub	2600 bp	1396	FR690941
"Candidatus Thiomargarita nelsonii" NAM053	regularly arranged round cells	155	19° 1.01´S 12° 13.75´E	233F / ITSReub	2600 bp	1396	FR690942
"Candidatus Thiomargarita nelsonii" NAM054	regularly arranged round cells	100	25° 00.04´S 14° 23.36´E	233F / ITSReub	1760 bp	1396	FR690943
"Candidatus Thiomargarita nelsonii" NAM055	chain of barrel-shaped cells	65	24° 03.19´S 14° 15.69´E	233F / ITSReub	2600 bp	1396	FR690944

Table S2 (continued)

"Candidatus Thiomargarita nelsonii" NAM056	chain of barrel-shaped cells	100	24° 03.19´S 14° 15.69´E	□M3 / □M4	2800 bp	1396	FR690945
"Candidatus Thiomargarita nelsonii" NAM057	chain of barrel-shaped cells	NA	19° 1.01´S 12° 13.75´E	□M3 / □M4	2800 bp	1396	FR690946
"Candidatus Thiomargarita nelsonii" NAM058	chain of barrel-shaped cells	42	24° 17.15´S 14° 16.10´E	233F / ITSReub	2600 bp	1396	FR690947
"Candidatus Thiomargarita nelsonii" NAM059	round cells in en□elope	20	NA	233F / 1300R	1100 bp	-	FR690948
"Candidatus Thiomargarita nelsonii" NAM060	regularly arranged round cells	114	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690949
"Candidatus Thiomargarita nelsonii" NAM061	regularly arranged round cells	200	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690950
"Candidatus Thiomargarita nelsonii" NAM062	regularly arranged round cells	NA	19° 1.01´S 12° 13.75´E	233F / ITSReub	2600 bp	1396	FR690951
"Candidatus Thiomargarita nelsonii" NAM063	regularly arranged round cells	118	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690952
"Candidatus Thiomargarita nelsonii" NAM064	regularly arranged round cells	92	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690953
"Candidatus Thiomargarita nelsonii" NAM065	regularly arranged round cells	90	25° 00.04´S 14° 23.36´E	233F / ITSReub	1760 bp	-	FN811659
"Candidatus Thiomargarita nelsonii" NAM066	regularly arranged round cells	72	25° 00.04´S 14° 23.36´E	233F / ITSReub	1760 bp	-	FR690954
"Candidatus Thiomargarita nelsonii" NAM067	regularly arranged round cells	NA	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690955
"Candidatus Thiomargarita nelsonii" NAM068	regularly arranged round cells	95	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690956
"Candidatus Thiomargarita nelsonii" NAM069	regularly arranged round cells	96	24° 17.15´S 14° 16.10´E	233F / ITSReub	1760 bp	-	FR690957
"Candidatus Thiomargarita nelsonii" NAM070	regularly arranged round cells	80	24° 03.19´S 14° 15.69´E	233F / ITSReub	1760 bp	-	FR690958
"Candidatus Thiomargarita nelsonii" NAM071	round cells in en□elope	37	NA	233F / ITSReub	1750 bp	-	FR690959
"Candidatus Thiomargarita nelsonii" NAM072	round cells in en□elope	25	NA	673F / 1300R	618 bp	-	FR690960
"Candidatus Thiomargarita nelsonii" COS008	single, round cell	190	8° 55.8´N 84° 18.75´W	233F/ITSReub	1800 bp	-	FN811661
"Candidatus Thiomargarita nelsonii" COS009	single, round cell	190	8° 55.8´N 84° 18.75´W	233F/ITSReub	2600 bp	1396	FR690961
"Candidatus Thiomargarita nelsonii" COS010	round double-cell	120	8° 55.8´N 84° 18.75´W	233F/ITSReub	1800 bp	-	FR690962
"Candidatus Thiomargarita nelsonii" COS011	round double-cell	160	8° 55.8´N 84° 18.75´W	233F/ITSReub	2600 bp	1396	FR690963
"Candidatus Thiomargarita nelsonii" COS012	attached, round cell	77	8° 55.8´N 84° 18.75´W	233F/ITSReub	1800 bp	-	FR690964
"Candidatus Thiomargarita nelsonii" COS013	attached, elongated cell	18	8° 55.8´N 84° 18.75´W	233F/ITSReub	1800 bp	-	FR690965
"Candidatus Thiomargarita nelsonii" COS014	attached, elongated cell	16	8° 55.8´N 84° 18.75´W	233F/ITSReub	1800 bp	-	FN811662
"Candidatus Thiomargarita nelsonii" COS015	attached, elongated cell	19	8° 55.8´N 84° 18.75´W	86/ITSReub	2000 bp	-	FR690966
"Candidatus Thiomargarita nelsonii" C□I006	regularly arranged round cells	78	36□30.80´S 73□7.75´W	233F/ITSReub	2600 bp	1396	FR690967
"Candidatus Thiopilula aggregata" NAM073	round cells in mucus aggregate	35	24° 03.19´S 14° 15.69´E	□M3 / □M4	2300 bp	1078	FR690968
"Candidatus Thiopilula aggregata" NAM074	round cells in mucus aggregate	35	24° 03.19´S 14° 15.69´E	□M3 / □M4	2300 bp	1078	FR690969
"Candidatus Thiopilula aggregata" NAM075	round cells in mucus aggregate	30-67	24° 03.19´S 14° 15.69´E	□M3 / □M4	2300 bp	1078	FR690970
"Candidatus Thiopilula aggregata" NAM076	round cells in mucus aggregate	25	24° 17.15´S 14° 16.10´E	□M3 / □M4	2800 bp	1396	FR690971
"Candidatus Thiopilula aggregata" NAM077	round cells in mucus aggregate	28	24° 17.15´S 14° 16.10´E	□M3 / □M4	2800 bp	1396	FR690972
"Candidatus Thiopilula aggregata" NAM078	round cells in mucus aggregate	30-44	24° 17.15´S 14° 16.10´E	□M3 / 673R and 233F / ITSReub	670 bp and 2600 bp	1396	FR690973
"Candidatus Thiopilula aggregata" NAM079	round cells in diatom frustule	21	24° 03.19´S 14° 15.69´E	□M3 / □M4	2800 bp	1396	FR690974
"Candidatus Thiopilula aggregata" NAM080	round cells in diatom frustule	12-17	19° 1.01´S 12° 13.75´E	□M3 / □M4	2800 bp	1396	FR690975
"Candidatus Thiopilula aggregata" NAM081	round cells in diatom frustule	NA	24° 03.19´S 14° 15.69´E	□M3 / 673R and 233F / ITSReub	670 bp and 2600 bp	1396	FR690976

Table S2 (continued)

"Candidatus Thiopilula aggregata*"* NAM082	round cells in diatom frustule	NA	24° 03.19´S 14° 15.69´E	GM3 / GM4	2800 bp	1396	FR690977
"Candidatus Thiopilula aggregata*"* NAM083	round cells in diatom frustule	11-24	24° 17.15´S 14° 16.10´E	GM3 / GM4	2800 bp	1396	FR690978
"Candidatus Thiopilula aggregata*"* NAM084	round cells in mucus aggregate	16	24° 03.19´S 14° 15.69´E	GM3 / GM4	1500 bp	-	FN811660
"Candidatus Thiopilula aggregata*" NA* M085	round cells in mucus aggregate	23	24° 17.15´S 14° 16.10´E	GM3 / GM4	1500 bp	-	FR690979
"Candidatus Thiopilula aggregata*"* NAM086	round cells in mucus aggregate	25	24° 17.15´S 14° 16.10´E	GM3 / GM4	2800 bp	1396	FR690980
"Candidatus Thiopilula aggregata*"* COS016	attached, elongated cell	28	8° 55.8´N 84° 18.75´W	360F/1300R	950 bp	-	FN811664
"Candidatus Thiopilula aggregata*"* COS017	attached, elongated cell	28	8° 55.8´N 84° 18.75´W	673F/1300R	600 bp	-	FR690981
"Candidatus Thiophysa hinzei*"* NAM087	single, round cell	60	24° 03.19´S 14° 15.69´E	233F/ITSReub	3600 bp	1396, 1495	FR690982
"Candidatus Thiophysa hinzei*"* NAM088	single, round cell	56	24° 17.15´S 14° 16.10´E	233F/ITSReub	3600 bp	1396, 1495	FR690983
"Candidatus Thiophysa hinzei*"* NAM089	single, round cell	61	24° 03.19´S 14° 15.69´E	233F/ITSReub	3600 bp	1396, 1495	FR690984
"Candidatus Thiophysa hinzei*"* NAM090	single, round cell	57	24° 03.19´S 14° 15.69´E	233F/ITSReub	3600 bp	1396, 1495	FR690985
"Candidatus Thiophysa hinzei*"* NAM091	single, round cell	90	24° 03.19´S 14° 15.69´E	233F/ITSReub	3600 bp	1396, 1495	FR690986
"Candidatus Marithioploca araucae*"* CHI007	single filament	33	36º 30.80´S 73º 7.75´W	233F/ITSReub	2600 bp	1396	FR690987
"Candidatus Marithioploca araucae*"* CHI008	single filament	37	36º 30.80´S 73º 7.75´W	233F/ITSReub	2600 bp	1396	FR690988
"Candidatus Marithioploca araucae*"* CHI009	single filament	32	36º 30.80´S 73º 7.75´W	GM3/673R and 233F/ITSR	670 bp and 2600 bp	1396	FR690989
"Candidatus Marithioploca araucae*"* CHI010	single filament	32	36º 30.80´S 73º 7.75´W	233F/ITSReub	2600 bp	1396	FR690990
"Candidatus Marithioploca araucae*"* CHI011	single filament	33	36º 30.80´S 73º 7.75´W	233F/ITSReub	2600 bp	1396	FR690991
"Candidatus Marithioploca araucae*"* CHI012	single filament	30	36º 30.80´S 73º 7.75´W	233F/ITSReub	2600 bp	1396	FR690992
"Candidatus Marithioploca araucae*"* CHI013	single filament	25	36º 30.80´S 73º 7.75´W	GM3/673R and 233F/ITSR	670 bp and 2600 bp	1396	FR690993
"Candidatus Marithioploca araucae*"* CHI014	single filament	18	36º 30.80´S 73º 7.75´W	GM3/673R and 233F/ITSR	670 bp and 2600 bp	1396	FR690994
"Candidatus Marithioploca araucae*"* CHI015	single filament	12	36º 30.80´S 73º 7.75´W	GM3/673R and 233F/ITSR	670 bp and 2600 bp	1396	FR690995
"Candidatus Marithioploca araucae*"* CHI016	single filament	16	36º 30.80´S 73º 7.75´W	233F/ITSReub	2600 bp	1396	FR690996
Thioploca ingrica DEN001	filaments in bundle	4	56° 28.11´N 10° 14.35´E	FWTpl131F/ITSReub	1900 bp	-	FR690997
Thioploca ingrica DEN002	filaments in bundle	4	56° 28.11´N 10° 14.35´E	FWTpl131F/ITSReub	1900 bp	-	FR690998
Beggiatoa alba B15LD	single filament	3	DSM1416	GM3/ITSReub	1900 bp	-	FR690999

Table S3. Overview of the insertion types in the 16S rRNA genes of large, colorless sulfur bacteria identified in the study. The presence of the insertion in at least one sequence of the given species is indicated by '+', whereas '–' indicates that the insertion was absent in all sequences of the species.

Cluster	species name	insertion type			
		795	1078	1396	1495
I	*Thiomargarita namibiensis*	+	+	+	+
II	"*Candidatus* Thiomargarita joergensenii"	+	+	-	+
III	"*Candidatus* Thiomargarita nelsonii"	-	-	+	-
VI	"*Candidatus* Marithioploca araucae"	-	-	+	-
V	"*Candidatus* Thiopilula aggregata"	-	+	+	-
VI	"*Candidatus* Thiophysa hinzeii"	-	-	+	+

References

[1] Bailey, J.V., Salman, V., Rouse, G., Schulz-Vogt, H.N., Levin, L., Orphan, V. Dimorphic development in methane seep dwelling ecotypes of the largest known bacteria. in prep.

[2] Cardinale, M., Brusetti, L., Quatrini, P., Borin, S., Puglia, A.M., Rizzi, A., Zanardini, E., Sorlini, C., Corselli, C., Daffonchio, D. (2004) Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. Appl. Environ. Microbiol. 70, 6147-6156.

[3] Garrity, G.M., Bell, J.A., Lilburn, T. (2005) Family I. *Thiotrichaceae* fam. nov. In: Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey´s Manual of Systematic Bacteriology, Springer, New York, Edition 2, Vol. 2, pp. 131.

[4] Kalanetra, K.M., Joye, S.B., Sunseri, N.R., Nelson, D.C. (2005) Novel vacuolate sulfur bacteria from the Gulf of Mexico reproduce by reductive division in three dimensions. Environ. Microbiol. 7, 1451-1460.

[5] Leadbetter, E.R. (1974) Family II. *Beggiatoaceae*. In: Buchanan, R.E., Gibbons, N.E. (Eds.), Bergey´s Manual of Determinative Bacteriology, The Williams & Wilkins Company, Baltimore, Edition 8, pp. 112-116.

[6] Muyzer, G., Teske, A., Wirsen, C.O., Jannasch, H.W. (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel-electrophoresis of 16S rDNA fragments. Arch. Microbiol. 164, 165-172.

[7] Skerman, V.B.D., McGowan, V., Sneath, P.H.A. (1980) Approved lists of bacterial names. Int. J. Syst. Bacteriol. 30, 225-420.

[8] Skerman, V.B.D., McGowan, V., Sneath, P.H.A. (1989) Approved lists of bacterial names (amended edition). In: American Society for Microbiology, Washington, DC, pp. 1-298.

[9] Strohl, W.R. (1989) Family I. *Beggiatoaceae*. In: Staley, J.T., Bryant, M.P., Pfennig, N., Holt, J.G. (Eds.), Bergey´s Manual of Systematic Bacteriology, Williams & Wilkins, Baltimore, Edition 1, Vol. 3, pp. 2089-2106.

Chapter 4

Large and frequent self-splicing introns in the 16S rRNA genes of large sulfur bacteria

Verena Salman[a], Rudolf Amann[a], and Heide N. Schulz-Vogt[a]

[a] Max Planck Institute for Marine Microbiology, Celsiusstraße 1, D-28359 Bremen, Germany

Manuscript in preparation

Relevant contributions:

The concept of the study was developed together with H. Schulz-Vogt and R. Amann. I designed probes and helpers and developed the protocol for CARD-FISH application on large sulfur bacteria, including the precursor hybridization technique. Development of primers and sequencing of the 23S rRNA were done by me and I performed *in vitro* splicing experiments. I also performed *in silico* analysis of the introns.

The manuscript was written in collaboration with H. Schulz-Vogt and R. Amann.

Abstract

The 16S rRNA of the small ribosomal subunit serves as a prominent tool for the phylogenetic analysis and classification of *Bacteria* and *Archaea* due to its conserved size of ~1.6 kb and its fundamental function in living organisms. Here we show that the 16S rRNA genes of not yet cultivated large sulfur bacteria, among them the largest known bacterium, *Thiomargarita namibiensis*, regularly contain numerous, long introns and can be enlarged to up to 3.5 kb. CARD-FISH on both ribosomes and rRNA precursor molecules, as well as *in vitro* splicing experiments show self-splicing of the intron and ligation of the exons, which explains, why the introns are not inhibiting the formation of functional ribosomes. The most frequent intron exhibits comparable phylogenetic relations as 16S rRNA genes within the different species, but diverging relationships between the species and genera of the family. This suggests that the introns are passed on to daughter cells, but may also be transferred horizontally to closely related bacteria. Remarkably, introns have never been identified in a bacterial 16S rRNA gene before, which is probably caused by a bias during PCR discriminating against longer homologous, as we can show experimentally. In conclusion, we suggest that the possibility of extremely long 16S rRNA genes has to be considered when generating clone libraries for diversity studies.

Introduction

Large sulfur bacteria are subject to fast and easy identification in environmental samples because of their conspicuous macroscopic appearance. In spite of this, they are usually underrepresented or even entirely missing from universal clone libraries in the course of diversity studies (Angert et al., 1998; Gillan et al., 1998; Edgcomb et al., 2002; Lopez-Garcia et al., 2003; Sekar et al., 2006; Stevens and Ulloa, 2008). Studies on microbial diversity in natural environments or in enrichment cultures usually involve the construction of clone libraries, for which the technique of polymerase-chain-reaction (PCR) is used. PCR, however, is commonly known to be prone to bias. Biases are usually systematic and can be already introduced during sample preparation before applying PCR (v. Wintzingerode et al., 1997) and are certainly caused by the fact that primers are not universal (Frank et al., 2008; Hong et al., 2009). Furthermore, the heterogeneity in concentration of source organisms (Suzuki and Giovannoni, 1996) or gene copy numbers (Cilia et al., 1996; Dennis et al., 1998; Coenye and Vandamme, 2003; Nannya et al., 2005) cause bias in the relative abundance of organisms in clone libraries, which are therefore not considered to be quantitative (Chandler et al., 1997). Eventually, there exists another bias in PCR, which has not yet been considered when constructing 16S rRNA-based clone libraries: length heterogeneity of the target gene. This bias is caused by kinetic reasons, during which a 16S rRNA gene of significantly extended length in a mixed sample will be less likely amplified than 16S rRNA gene of the standard length of ~1.6 kb. In turn, the microbial diversity as assessed by a clone library is thus specifically distorted towards shorter 16S rRNA genes.

In a previous study, we investigated the phylogenetic diversity of large sulfur bacteria by developing new protocols to retrieve nearly full-length 16S rRNA gene sequences, and applied this on more than one hundred single cells and filaments (Salman et al., 2011). In the course of this study, it became apparent that large sulfur bacteria commonly contain one or multiple introns in their 16S rRNA genes. As a result, their 16S rRNA gene sequences can be more than 3 kb long. Here we show that the introns are removed efficiently from host rRNA precursor molecules, which is why they possibly have no negative impact on the host. However, the overall presence of introns in this group of bacteria leads to the hypothesis that there could be many other intron-containing groups of bacteria, which are probably less morphologically conspicuous, but likewise systematically excluded from detection in universal clone libraries.

Material and Methods

16S rRNA sequences

A detailed description of the amplification of 16S rRNA gene sequences including introns has recently been published (Salman et al., 2011).

23S rRNA sequences

The 23S rRNA gene of a single "*Candidatus* Thiomargarita nelsonii" chain was amplified according to (Salman et al., 2011) using the primers ITS350F (5´-aattaggaagctgatgtaaa-3´) binding in the ITS1 region of *Thiomargarita* spp. and 97ar (Van Camp et al., 1993) binding at the 3´-end of bacterial 23S rRNA genes. The elongation time during PCR was extended to 5 minutes. The amplicon of ca. 4500 nt was purified using a spin column containing Sephadex^{TM} G-50 Superfine (Amersham Biosciences, Uppsala, Sweden), directly sequenced with the Big Dye Cycle Sequencing Kit and further analyzed on an ABI Genetic Analyzer 3130x (Applied Biosystems, USA). The localization of the three introns was done by manually aligning the obtained sequence with the 23S rRNA gene sequence of the close relative *Leucothrix mucor* (accession number X87285) in BioEdit (Hall, 1999). The complete sequence was deposited in the GenBank/EMBL/DDBJ databases under accession number FR774200.

Catalyzed reporter deposition-fluorescence *in situ* hybridization

Probe 1284 was designed matching 16S rRNA of all non-filamentous large sulfur bacteria using the ARB ProbeDesign tool (Ludwig et al., 2004). Prior to performing catalyzed reporter deposition-fluorescence *in situ* hybridization, CARD-FISH (Pernthaler et al., 2002), its discrimination against all database entries besides non-filamentous members of the *Beggiatoaceae* was tested with the online Probe Match tool provided by the Ribosomal Database Project (Cole et al., 2005). Specificity was tested using filamentous "*Candidatus* Marithioploca araucae", containing one central mismatch to the probes´ sequence (data not shown). All other probes and helpers applied in this study (for introns and ITS regions) were designed manually. Horseradish peroxidase (HRP)-labelled probes and non-labelled helpers were purchased from Biomers (Ulm, Germany).

Chains of *Thiomargarita* spp. were removed from Namibian sediment (Salman et al., 2011), transferred into 5% formaldehyde in artificial seawater and fixed for 1 h at room temperature. Subsequently, they were transferred into 50% ethanol (in 3.5% NaCl) and

incubated for 1 h at room temperature before being washed in 3.5% NaCl. The individual chains were fixed on polylysine-covered glass slides (Menzel, Braunschweig, Germany), sprayed with 0.4% low melting agarose (Nu Sieve® GTG® Agarose, Cambrex Bioscience Rockland, Inc.; Rockland, ME, USA), dried at 37°C and stored at -20°C. Cells were dehydrated in ethanol (50%, 80%, 99%, each for 1 minute) and air-dried at room temperature. Permeabilization with 10 mg/mL lysozyme was done for 1 h at 37°C in a moisture chamber, followed by washing in MilliQ twice. Internal peroxidises were inactivated by incubation in 0.15% H_2O_2 in methanol for 30 minutes at room temperature. Slides were washed twice in MilliQ and once in 99% ethanol before air-drying. Hybridization was performed in a pre-warmed humid chamber. Twenty microliters of hybridization buffer (20 mM Tris-HCl, 0.01% SDS, 900 mM NaCl, 10−35% formamide) containing 50 ng of (each) HRP-labelled probe(s), and if necessary the respective helpers, were spotted onto the cells and incubated for 3 h at the appropriate hybridization temperature (see Table 2). Slides were washed in washing buffer (20 mM Tris-HCl, 5 mM EDTA, 0.01% SDS, 80−450 mM NaCl according to the formamide concentration in hybridization buffer) for 15 minutes at 48°C or 37°C and transferred twice to fresh 1× PBS (pH 7.6). For signal amplification, 70 µL of amplification buffer containing 0.015% H_2O_2 and either FITC-labelled or Alexa488-labelled tyramide were applied and incubated for 45 minutes at 46°C or 35°C in the dark. Slides washed once in 1× PBS (pH 7.6, 5 minutes), twice in MilliQ (5 minutes each), once in 99% ethanol (2 minutes), air-dried in the dark and subsequently mounted with 4:1 Citifluor AF1 (Citifluor, Leicester, UK) and Vectashield (Vector laboratories, Burlinghame, CA, USA) before freezing at -20°C. In case of double hybridization, a second round of hybridization was applied before mounting the slides. Peroxidases of the first probe(s) were inactivated as previously described. For second the hybridization, the time was extended to 15 h and Alexa594-labelled tyramide was used for signal amplification. Examination of the fluorescent signals was performed with a Zeiss epifluorescence microscope.

In vitro splicing

PCR amplicons of 16S rRNA genes containing an intron were cloned using the TOPO TA cloning® kit (Invitrogen, Darmstadt, Germany). Plasmids were purified (Qiagen, Hilden, Germany), linearized (*SpeI* or *PstI*), purified (Qiagen, Hilden, Germany) and separated on a 1% agarose gel to determine the concentration. For *in vitro* transcription, 20 µL reaction mix containing 1 µg DNA template, 0.5 mM of each NTP, 1× T7 Transcription Buffer and 20 U T7 RNA Polymerase-Plus™ Enzyme Mix (Ambion, Darmstadt, Germany) were

incubated for 2 h at 37°C. After the addition of 2 U DNase I (Fermentas, St. Leon-Rot, Germany) and another incubation at 37°C for 15 minutes, 1 µL EDTA (50 mM) was added and the reaction mix was heated to 65°C for 10 minutes. RNA was mixed with 2× RNA loading dye (Fermentas, St. Leon-Rot, Germany), heated to 70°C for 10 minutes and subject to agarose gel electrophoresis on a 1% gel in 1× TBE buffer. To confirm splicing and religation of the exons, the RNA was subject to RT-PCR. 20 µL reaction mix were prepared containing 0.5 µg RNA, 0.5 µM of the product-specific reverse primer, 1 mM of each dNTP, 1× RT Reaction Buffer, 20 U RiboLock^{TM} RNase Inhibitor and 40 U M-MuLV Reverse Transcriptase (Fermentas, St. Leon-Rot, Germany). Reverse transcription was performed by incubation at 37°C for 1 h and stopped by heating to 70°C for 10 minutes. Subsequently, 1 µL cDNA was added to a 25 µL PCR reaction mix containing 1 µM each of a product-specific forward and reverse primer, 0.2 mM of each dNTP, 1× High Fidelity PCR Enzyme Mix Reaction Buffer with $MgCl_2$ (1.5 mM) and 1.25 U High Fidelity Enzyme Mix (Fermentas, St. Leon-Rot, Germany). The PCR program included initial heating to 94°C for 2 minutes, followed by 30 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 68°C for 2 minutes, and a final elongation time of 7 minutes at 68°C. PCR products were separated on a 1% agarose gel, the resulting bands were cut out and extracted (Qiagen, Hilden, Germany), sequenced using the Big Dye Cycle Sequencing Kit (Applied Biosystems, Darmstadt, Germany) and analyzed on an ABI Genetic Analyzer 3130x (Applied Biosystems, Darmstadt, Germany).

PCR of homologous genes with different lengths

A 16S rRNA gene lacking introns (1500 nt) and a gene containing either intron S1396 (2350 nt) or intron S1078 (2200 nt) were cloned using the TOPO TA cloning® kit (Invitrogen, Darmstadt, Germany). Vectors were amplified individually or simultaneously in a mix containing equal amounts of vectors with and without introns. The PCR reaction always contained 1 µM each of universal forward (GM3F) and reverse (GM4R) primer (Muyzer et al., 1995), 0.2 mM of each dNTP, 1× High Fidelity PCR Enzyme Mix Reaction Buffer with $MgCl_2$ (1.5 mM) and 1.25 U High Fidelity Enzyme Mix (Fermentas, St. Leon-Rot, Germany). The PCR program started with heating to 94°C for 2 minutes, followed by 25 cycles of 94°C for 30 seconds, 42°C for 30 seconds and 72°C for 2 minutes, and a final elongation time of 10 minutes at 72°C. The resulting PCR amplicons were subject to agarose gel electrophoresis.

Results

Introns in rRNA genes

We identified introns with variable lengths (290−942 nucleotides) at up to four sites (Figure 1A) within the 16S rRNA genes of large sulfur bacteria (mentioned in Salman et al., 2011). Corresponding to their insertion site within these genes, they were classified as S795, S1078, S1396 and S1495 (according to E. coli numbering, adapted from Salman et al., 2011). In the present study, we additionally identified three inserted sequences in the 23S rRNA gene of "*Candidatus* Thiomargarita nelsonii". These novel insertions were classified as L1920, L1932 and L1955, accordingly (Figure 1A). The seven insertion sites in the two genes did not show sequence homology and sequence identities between the seven intron types accounted for merely 40% or less. This implies that the seven introns were not homologues. Intron distribution varies between the different phylogenetic groups with the *Beggiatoaceae* (Figure 1B) and also between each individual member of a given species (for details see Figure S1, supplementary material). Those introns inserted at the same site were conserved in their nucleotide sequence (usually 85−100% identical, except for S1495 with 46.8%, Table 1) and within a given species, introns were 92−100% identical (Table 1). S1396 was the most frequent intron in the group of large sulfur bacteria. Its phylogeny within a given species reflected phylogeny of the 16S rRNA gene, but differs in the relationship between species and genera (for details see Figure S2, supplementary material).

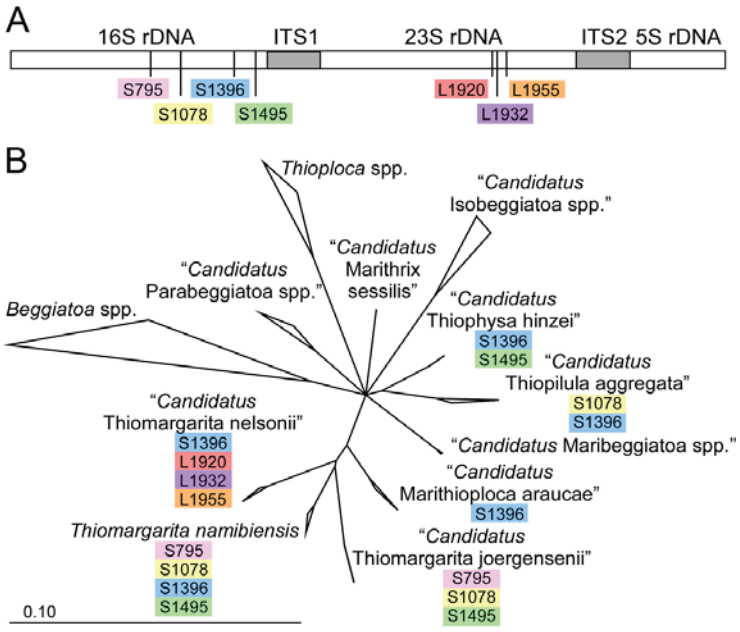


Figure 1. (A) In large sulfur bacteria, four introns can be inserted in the gene for the small ribosomal subunit (16S rRNA) and three introns can be inserted in the gene for the large ribosomal subunit (23S rRNA). The letters 'S' and 'L' indicate the gene, in which the intron is inserted, and the number refers to the insertion site within the gene according to *E. coli* numbering. (B) Multifurcation tree based on nearly full-length 16S rRNA gene sequences of members of the family *Beggiatoaceae*. So far, introns were located in most 16S rRNA genes of the genera *Thiomargarita*, "*Candidatus* Marithioploca", "*Candidatus* Thiopilula" and "*Candidatus* Thiophysa". The analysis of one 23S rRNA gene of "*Candidatus* Thiomargarita nelsonii" revealed the presence of another tree intron types.

Introns S795, S1078, S1495, L1920, L1935 and L1955 were characterized as group I introns (Figure 2A) and S1396 as a group II intron (Figure 2B). *In silico* modelling of RNA folding patterns revealed conserved domains and stem-loop-structures for these groups of introns (reviewed in Ref. Lambowitz and Belfort, 1993). Furthermore, of 148 investigated introns, 112 contained an open reading frame (ORF) coding for an intron encoded protein (IEP), which is a common property of introns (Lambowitz and Belfort, 1993). The primary sequence of all seven intron types investigated here are only distantly related to other known bacterial or organellar introns from public databases, which is also common for introns (Lambowitz and Belfort, 1993; Saldanha et al., 1993). The nested ORF coded for a putative site-specific endonuclease of the LAGLIDADG/HNH superfamily. Maximum sequence identity (67%) was detected between intron L1955 and a corresponding gene for a homing endonuclease inserted into a group I intron in the 23S rRNA gene of the bacterium *Coxiella burnetii* (accession number CP001020). This family of homing endonucleases is usually encoded by group I introns, but occasionally also found in group II introns (Toor and Zimmerly, 2002; Haugen et al., 2005).

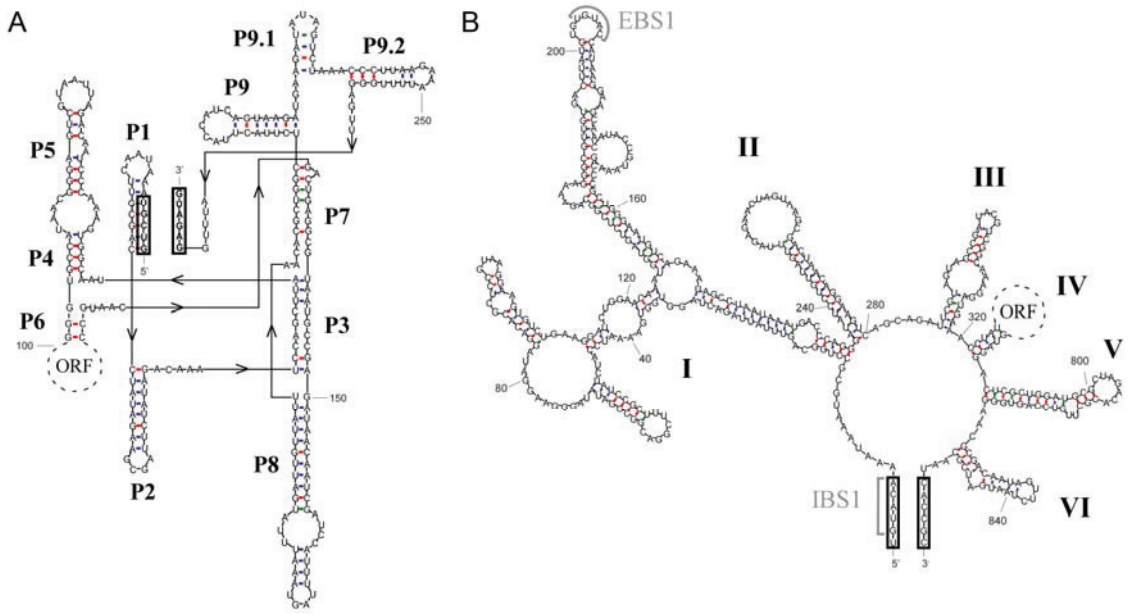


Figure 2. Secondary structure diagram of (A) the group I intron S1078 (in *Thiomargarita namibiensis* NAM001, accession number FR690879) and (B) the group II intron S1396 (in "*Candidatus* Thiomargarita nelsonii" NAM036, accession number FR690926). Conserved domains in the respective diagram are indicated with P1−P9 (A) and DI−DVI (B) at each helix structure. The locations of the ORFs are illustrated by a dashed loop. In boxes are nucleotides of the exon regions adjacent to both sides of the intron.

Ligation of exons during *in vitro* splicing

The capability of self-splicing and ligation was demonstrated for several bacterial introns (e.g., Adamidi et al., 2003; Nesbø and Doolittle, 2003; Birgisdottir and Johansen, 2005; Chien et al., 2009) and in this study three out of seven introns were investigated for their self-splicing potential (Figure 3A−D). S1078, S1396 and S1495 were tested as ORF-less group I and group II introns. S1078 from a different donor was also analyzed as ORF-containing group I intron. All four tested intron types showed self-splicing in transcription buffer during RNA synthesis (Figure 3B). Ligation of the exons was demonstrated by RT-PCR (Figure 3C). Sequencing of the resulting amplicons (Figure 3D) confirmed ligation and revealed that exons were not truncated during splicing.

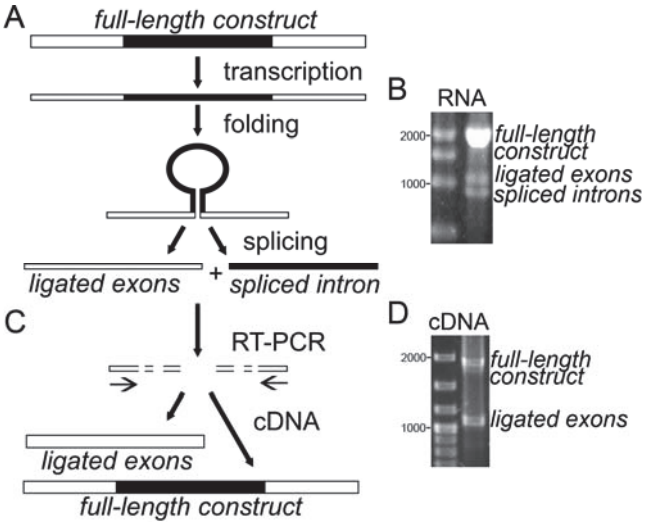


Figure 3. *In vitro* splicing and ligation of the two exons. DNA is shown as wide bars and RNA as thin bars, exon regions are white and the intron region is black. (A) During transcription, the RNA intron folds and splices, thereby ligating the two exons. (B) The RNA gel shows three bands, indicating the presence of the not yet spliced full-length construct (1900 nt), the ligated exons (1200 nt) and the spliced intron (700 nt). (C) Bulk RNA was subject to reverse transcription PCR (RT-PCR) with primers on both exon ends. (D) The resulting cDNA was separated on a DNA gel, showing the presence of the full-length construct and the ligated exons.

Detection of S1078 and S1396 in precursor rRNA

CARD-FISH hybridizations of whole fixed cells of large sulfur bacteria indicated that introns were absent from mature 16S rRNA (Figure 5A, step 5). An oligonucleotide targeting position 1285−1304 of the 16S rRNA (VSO1285, Table S1) hybridized to nearly all cells, yielding strong signals indicative of high cellular concentration of target molecules (Fig. 4A). In contrast, single oligonucleotides targeting intron sequences yielded no signal above background (Fig. 4B). The application of a probe, of which the sequence bridged the intron insertion site, revealed likewise strong fluorescent signals (Figure 5E, exons). In contrast, a probe that targeted partly the exon and partly the adjacent spliced intron sequence showed no detectable hybridization (Figure 5E, exon+S1396).

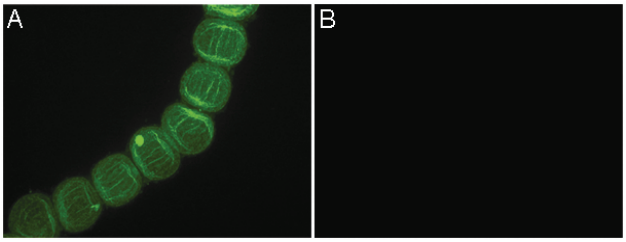


Figure 4. (A) CARD-FISH hybridization of the 16S rRNA using probe VSO1285 revealed strong signals for nearly all cells examined. (B) Hybridization of intron regions using a single probe, e.g. S1396 with probe S1396-179 as shown here, never revealed detectable signals.

To exclude that the intron-containing 16S rRNA genes are pseudogenes we applied a more sensitive CARD-FISH assay and tested whether the 16S rRNA introns were transcribed as part of the rRNA precursor (Figure 5A, step 1). The ITS region in the primary transcript was hybridized as positive control. Each of the excized segments (ITS, S1078 or S1396) was targeted with a mix of three to four probes, hybridized at lower temperatures (35°C instead of 46°C) and by using several helpers to open putative secondary structures (Table S1, Fuchs et al., 2000). Two *Thiomargarita* species, distinguishable by a distinct intron composition, were used for this hybridization technique: *T. namibiensis* containing S1078 and S1396 and "*Candidatus* T. nelsonii" containing only S1396. ITS- and intron-specific signals could be obtained and double-hybridization experiments targeting both precursor segments revealed that all cells having intron-specific signals simultaneously showed ITS-specific signals (Figure 5B and C). It should be noted that only few of the analysed *Thiomargarita* spp. cells showed signals when targeting excised regions of the precursor RNA, because its concentration is low and can vary strongly between cells (Britschgi and Cangelosi, 1995). This finding confirmed that intron-containing 16S rRNA genes were indeed transcribed (Figure 5A, step 1). Frequently, ITS signals were detectable although intron signals were not, and never vice versa. This indicates that introns were excised and/or degraded more efficiently than ITS (Figure 5, steps 2−5). Double hybridization on both intron sequences in *T. namibiensis* furthermore revealed simultaneous presence of S1078 and S1396, however, in some chains, only signals for the longer S1396 were detectable (Figure 5D). This suggests that also the shorter intron S1078 might be removed/degraded faster than the longer intron S1396.

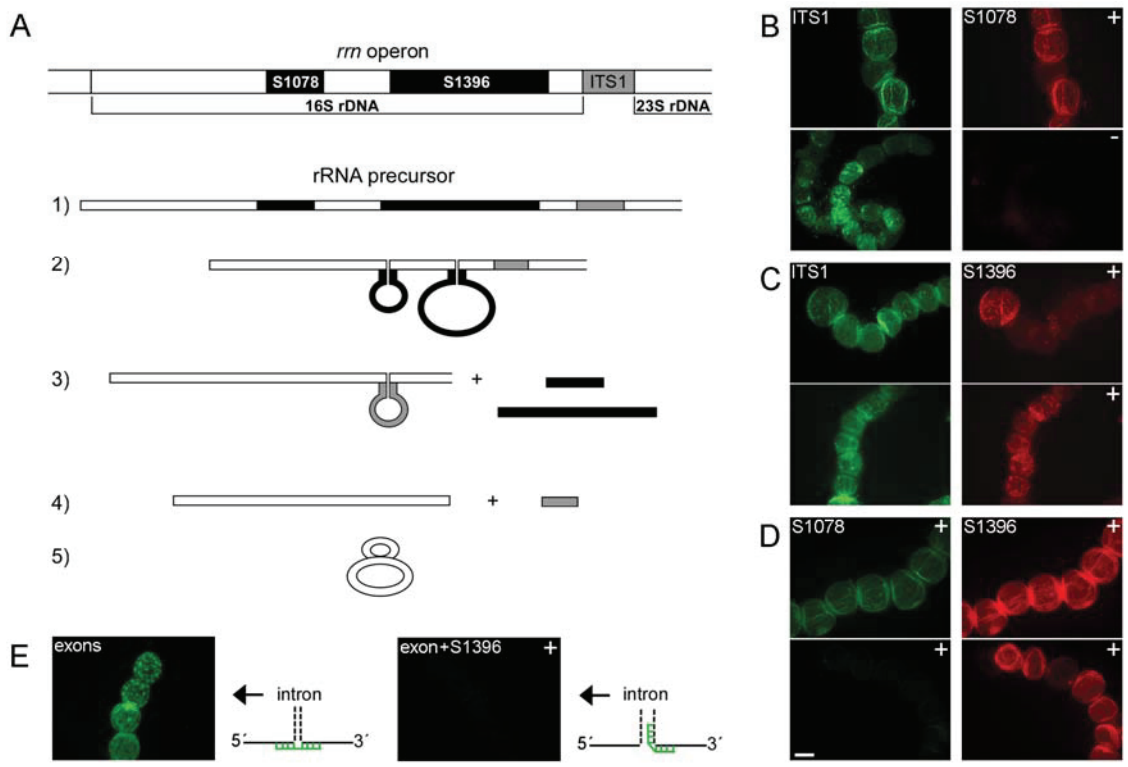


Figure 5. (A) Processing of the rRNA precursor and different maturation steps (1−5) in large sulfur bacteria. DNA is shown as wide bars and RNA as thin bars, exon regions are shown in white, intron regions in black and the ITS region in grey. At first, both introns and the ITS region are present in the rRNA precursor (step 1). The introns fold (step 2) and are spliced out, thereby ligating the two exons and giving way for folding of ITS region (step 3). Cellular enzymes then remove the ITS region (step 4) and the native ribosome is formed (step 5). (B) Simultaneous hybridization of the ITS region and S1078 on the precursor in *Thiomargarita* sp. containing S1078 (+) and another species lacking S1078 (-). (C) Simultaneous hybridization of the ITS region and S1396 present in both *Thiomargarita* spp. (+). (D) Simultaneous hybridization of S1078 and S1396 present in *Thiomargarita* sp. (+). (E) Hybridization of the native ribosome using either one probe that bridges the intron insertion site (exons) or one probe that binds to the 3´-exon and the 3´-intron region (exon+S1396).

Discussion

Introns are a common feature within the rRNA genes of large sulfur bacteria. There are at least four different types in the 16S rRNA gene and three different types in the 23S rRNA gene. All types are analogous because they do not share a common target sequence and not related among each other. Thus every intron type has probably individually inserted into the genome. Phylogenetic analysis of the most frequently occurring intron S1396 revealed that it was probably acquired both vertically and horizontally within the family. This feature is well known among group II introns (Itoh et al., 1998; Toro, 2003; Raghavan et al., 2007; Tourasse and Kolstø, 2008). On the level of individuals within a species, the sequence of intron S1396 generally follows the phylogeny suggested by 16S rRNA sequences (Table 1),

indicating that the intron coevolved with the host after insertion into its genome. On the family level, the relationship of introns between genera is different from the 16S rRNA-derived phylogenetic pattern, suggesting horizontal transfer of introns between different species or genera (for details see Figure S2). This phenomenon is likely caused by the high mobility of these elements, frequently inserting and exciting from genomes (Dai and Zimmerly, 2002; Toro, 2003).

Table 1. Introns identified in the 16S rRNA genes of the different species of large sulfur bacteria and identity values calculated within and among species.

intron	intron-containing species	sequence identities within the species	sequence identities among the different intron containing species
S795 with ORF	*Thiomargarita namibiensis*	94.6−100%	≤89% (“*Candidatus* Thiomargarita joergensenii”)
	“*Candidatus* Thiomargarita joergensenii”	-	
S1078 without ORF	*Thiomargarita namibiensis*	98.1−100%	-
S1078 with ORF	“*Candidatus* Thiomargarita joergensenii”	98.8−100%	85.2% (“*Candidatus* Thiopilula aggregata”)
	“*Candidatus* Thiopilula aggregata”	99.9−100%	
S1396 without ORF	*Thiomargarita namibiensis*	98.0−100%	-
S1396 with ORF	*Thiomargarita namibiensis*	96.6−100%	≤93.5% (“*Candidatus* Thiomargarita nelsonii”)
	“*Candidatus* Thiomargarita nelsonii”	97.9−100%	-
	“*Candidatus* Marithioploca araucae”	92.1−100%	≤96.5% (*Thiomargarita* spp.), 91.2% (“*Candidatus* Thiopilula spp.”)
	“*Candidatus* Thiopilula aggregata”	96.0−100%	≤92.5% (*Thiomargarita* spp.), 91.6% (“*Candidatus* Thiophysa hinzei”)
	“*Candidatus* Thiophysa hinzei”	98.8−100%	93.9% (*Thiomargarita* spp.), 92.2% (“*Candidatus* Marithioploca araucae”)
S1495 without ORF	*Thiomargarita namibiensis*	-	46.8% (“*Candidatus* Thiomargarita joergensenii”)
	“*Candidatus* Thiomargarita joergensenii”	-	
S1495 with ORF	*Thiomargarita namibiensis*	97.9−100%	≤89.1% (“*Candidatus* Thiophysa hinzei”)
	“*Candidatus* Thiophysa hinzei”	99.0−100%	

The mixed presence and absence of open reading frames (ORFs) in the different intron types is in accordance with the Goddard-Burt-model (1999). This model describes a course of ORF-gain, ORF-truncation and ORF-loss in a population of group I intron containing organisms, and was confirmed in several studies (Foley et al., 2000; Bhattacharya et al., 2002; Cannone et al., 2002; Nozaki et al., 2002; Haugen et al., 2005). ORFs contain information for intron encoded proteins (IEP) and are required for intron mobility in group I introns, but are considered unnecessary for intron splicing. Accordingly, once an intron has spread to all possible target sites within a genome, the IEP is redundant and looses its function (Goddard and Burt, 1999; Haugen et al., 2005). Apparently, in some organisms of the large sulfur bacteria, certain introns have saturated their targets and are seemingly in the course of loosing their IEPs. Group II introns without IEPs, on the other hand, are very uncommon (Simon et al., 2008), but are known to recruit the IEPs of closely related group II introns, which are simultaneously present in the host (Nakamura et al., 2002; Dai and Zimmerly, 2003). Thus, it can be hypothesized that some sulfur bacteria, featuring and

ORF-less S1396, might have additional, closely related ORF-containing group II introns in their genomes.

The splicing mechanism from RNA precursors and the dispersal within the host genome are studied very well in bacterial introns (Nielsen and Johansen, 2009; Toor et al., 2009). Also sulfur bacterial introns show self-splicing capabilities *in vitro* (Figure 3) and a removal from precursor rRNA molecules *in vivo* (Figure 4). CARD-FISH on the rRNA precursor was previously performed targeting the ITS region (Oerther et al., 2000; Schmid et al., 2001), and the technique was expanded in this study hybridizing also RNA introns. The successful and specific hybridization of the different introns demonstrates their transcription into RNA, strongly indicating that the16S rRNA genes retrieved by PCR were no pseudogenes. The introns have no clear target sequence for precursor processing enzymes like RNAse III, which indicates that the mechanism of intron splicing is probably uncoupled from precursor maturation. The efficient intron splicing prevents negative impact on ribosome functionality. The rRNA genes are not translated into proteins and the rRNA precursor is not subject to simultaneous transcription and translation like protein-coding RNA precursors. Thus, the time between precursor synthesis and cellular precursor maturation is seemingly sufficient to allow intron splicing.

The detection of large and frequent introns in the rRNA genes of large sulfur bacteria raises questions on a possible biological in the host organisms. Currently, we can only speculate about possible reasons for their presence. The introns might not be a big burdon since the physiologically specialized chemolithotrophic bacteria grow very slowly. They store high amounts of energy reserves (Schulz, 2006; Teske and Nelson, 2006), and since DNA replication requires only about 2% of a cell´s energy budget (Lane and Martin, 2010), these bacteria might be able to afford the transcription of intron-containing genes into RNA. Alternatively, it is tempting to speculated that the high frequency of introns in this group of bacteria is somehow connected to the extreme polyploidy in these organisms, which contain up to several thousands nucleoids (Lane and Martin, 2010). We postulate that intron persistence does not always imply a positive evolutionary cause.

More important than their functional aspects is at this state the mere presence of introns in 16S rRNA genes. This observation brings up the question whether the large sulfur bacteria are the only group of prokaryotes, which accumulate introns in this gene, or whether also other groups of bacteria contain such introns and are likewise discriminated in clone

libraries. Introns in the 23S rRNA genes have been reported for other bacteria (Everett et al., 1999; Nesbø and Doolittle, 2003; Seshadri et al., 2003; Raghavan et al., 2007) and this gene bears the same evolutionary importance and conservation as the 16S rRNA gene. Among the *Archaea*, introns occur even more frequently in rRNA genes and are also found in 16S rRNA genes (Burggraf et al., 1993; Itoh et al., 1998). Nevertheless, until today, there were no entries in public databases reporting bacterial 16S rRNA genes with introns (Cannone et al., 2002; Jackson et al., 2002; Simon et al., 2008). This could be a coincidence or the consequence of a bias produced by the applied techniques, including even the interpretation of results.

We suggest that 16S rRNA genes, which are enlarged due to the presence of introns, are unlikely to be detected using commonly applied techniques for retrieving 16S rRNA sequences. This was confirmed experimentally by mixing the same amount of intron-containing (2350 nt) and intron-lacking (1500 nt) 16S rRNA genes and applying a universal PCR. We observed a systematic discrimination of the longer gene over the homologue being shorter by 850 nt (Figure 6, lane 3). In case the two homologues differ slightly less in length (700 nt), longer amplicons may be produced in low concentrations (Fig. 6, lane 4). This PCR bias can be explained by the kinetics of the reaction (Cardinale et al., 2004). The amplification of longer 16S rRNA genes from natural mixed samples, in which the amount of intron-containing genes is likely much lower than intron-lacking genes, may thus be affected much stronger by this length heterogeneity bias. Accordingly, it can be assumed that intron-containing 16S rRNA gene sequences have been strongly suppressed in environmental clone libraries mainly due to this length heterogeneity bias. Furthermore, it seems likely that PCR amplicons significantly larger than the expected size were excluded from further analysis in earlier studies.

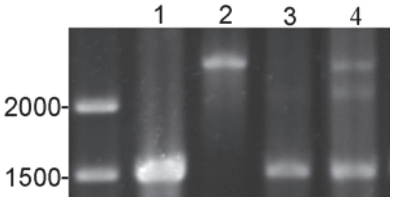


Figure 6. In a universal PCR approach, a vector containing a 16S rRNA gene without intron (1500 nt, lane 1) and a vector containing a 16S rRNA gene with S1396 (2350 nt, lane 2) was amplified. In a PCR reaction containing the same amount of both these target genes, only the non-intron-containing 16S rRNA gene is amplified (lane 3). A PCR reaction containing equal amounts of the intron-free 16S rRNA gene and a gene enlarged by 700 nt, also larger amplicons are produced simulatenously with the shorter homologue (lane 4).

The sole reason why it was noticed that 16S rRNA gene sequences of large sulfur bacteria were frequently absent from universal clone libraries (Angert et al., 1998; Gillan et al., 1998; Edgcomb et al., 2002; Lopez-Garcia et al., 2003; Sekar et al., 2006; Stevens and Ulloa,

2008) was the fact that their presence in an environmental sample could be unambiguously verified. A biased retrieval based on length differentiation certainly also needs to be considered for microbes other than large sulfur bacteria and which do not possess a conspicuous morphology. The analysis of large metagenomic datasets could in the future reveal whether introns are even a quite common trait of bacterial 16S rRNA genes.

Acknowledgements

This study was funded by the Max Planck Society.

References

Adamidi, C., Fedorova, O., and Pyle, A.M. (2003) A group II intron inserted into a bacterial heat-shock operon shows autocatalytic activity and unusual thermostability. *Biochemistry* **42**: 3409-3418.

Angert, E.R., Northup, D.E., Reysenbach, A.L., Peek, A.S., Goebel, B.M., and Pace, N.R. (1998) Molecular phylogenetic analysis of a bacterial community in Sulphur River, Parker Cave, Kentucky. *American Mineralogist* **83**: 1583-1592.

Bhattacharya, D., Friedl, T., and Helms, G. (2002) Vertical evolution and intragenic spread of lichen-fungal group I introns. *J Mol Evol* **55**: 74-84.

Birgisdottir, A.B., and Johansen, S. (2005) Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. *Nucl Acids Res* **33**: 2042-2051.

Britschgi, T.B., and Cangelosi, G.A. (1995) Detection of rifampin-resistant bacteria using DNA probes for precursor rRNA. *Mol. Cell. Probes* **9**: 19-24.

Burggraf, S., Larsen, N., Woese, C.R., and Stetter, K.O. (1993) An Intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci U S A* **90**: 2547-2550.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D´Souza, L.M., Du, Y. et al. (2002) The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 1-31.

Cardinale, M., Brusetti, L., Quatrini, P., Borin, S., Puglia, A.M., Rizzi, A. et al. (2004) Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Appl Environ Microbiol* **70**: 6147-6156.

Chandler, D.P., Fredrickson, J.K., and Brockman, F.J. (1997) Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. *Mol Ecol* **6**: 475-482.

Chien, M.F., Tosa, S., Huang, C.C., and Endo, G. (2009) Splicing of a bacterial group II intron from *Bacillus megaterium* is independent of intron-encoded protein. *Microbes Environ* **24**: 28-32.

Cilia, V., Lafay, B., and Christen, R. (1996) Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol* **13**: 451-461.

Coenye, T., and Vandamme, P. (2003) Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* **228**: 45-49.

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M. et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucl Acids Res* **33**: D294-D296.

Dai, L., and Zimmerly, S. (2002) Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucl Acids Res* **30**: 1091-1102.

Dai, L., and Zimmerly, S. (2003) ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA* **9**: 14-19.

Dennis, P.P., Ziesche, S., and Mylvaganam, S. (1998) Transcription analysis of two disparate rRNA operons in the halophilic archaeon *Haloarcula marismortui*. *J Bacteriol* **180**: 4804-4813.

Edgcomb, V.P., Kysela, D.T., Teske, A., Gomez, A.D., and Sogin, M.L. (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc Natl Acad Sci U S A* **99**: 7658-7662.

Everett, K.D.E., Kahane, S., Bush, R.M., and Friedman, M.G. (1999) An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J Bacteriol* **181**: 4734-4740.

Foley, S., Bruttin, A., and Brussow, H. (2000) Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. *Journal of Virology* **74**: 611-618.

Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A., and Olsen, G.J. (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* **74**: 2461-2470.

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

Gillan, D.C., Speksnijder, A.G.C.L., Zwart, G., and De Ridder, C. (1998) Genetic diversity of the biofilm covering *Montacuta ferruginosa* (*Mollusca, Bivalvia*) as evaluated by denaturing gradient gel electrophoresis analysis and cloning of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ Microbiol* **64**: 3464-3472.

Goddard, M.R., and Burt, A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A* **96**: 13880-13885.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **No. 41**: 95-98.

Haugen, P., Simon, D.M., and Bhattacharya, D. (2005) The natural history of group I introns. *Trends Genet* **21**: 111-119.

Hong, S.H., Bunge, J., Leslin, C., Jeon, S., and Epstein, S.S. (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J* **3**: 1365-1373.

Itoh, T., Suzuki, K., and Nakase, T. (1998) Occurrence of introns in the 16S rRNA genes of members of the genus *Thermoproteus*. *Arch Microbiol* **170**: 155-161.

Jackson, S.A., Cannone, J.J., Lee, J.C., Gutell, R.R., and Woodson, S.A. (2002) Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J Mol Biol* **323**: 35-52.

Lambowitz, A.M., and Belfort, M. (1993) Introns as mobile genetic elements. *Annu Rev Biochem* **62**: 587-622.

Lane, N., and Martin, W. (2010) The energetics of genome complexity. *Nature* **467**: 929-934.

Lopez-Garcia, P., Duperron, S., Philippot, P., Foriel, J., Susini, J., and Moreira, D. (2003) Bacterial diversity in hydrothermal sediment and epsilon proteobacterial dominance in experimental microcolonizers at the Mid-Atlantic Ridge. *Environ Microbiol* **5**: 961-976.

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

Muyzer, G., Teske, A., Wirsen, C.O., and Jannasch, H.W. (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel-electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164**: 165-172.

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S. et al. (2002) Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res* **9**: 123-130.

Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L.L., Hangaishi, A. et al. (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Research* **65**: 6071-6079.

Nesbø, C.L., and Doolittle, W.F. (2003) Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc Natl Acad Sci U S A* **100**: 10806-10811.

Nielsen, H., and Johansen, D. (2009) Group I introns. Moving in new directions. *RNA Biol* **6**: 375-383.

Nozaki, H., Takahara, M., Nakazawa, A., Kita, Y., Yamada, T., Takano, H. et al. (2002) Evolution of rbcL group IA introns and intron open reading frames within the colonial *Volvocales* (*Chlorophyceae*). *Molecular Phylogenetics and Evolution* **23**: 326-338.

Oerther, D.B., Pernthaler, J., Schramm, A., Amann, R., and Raskin, L. (2000) Monitoring precursor 16S rRNAs of *Acinetobacter* spp. in activated sludge wastewater treatment systems. *Appl Environ Microbiol* **66**: 2154-2165.

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

Raghavan, R., Miller, S.R., Hicks, L.D., and Minnick, M.F. (2007) The unusual 23S rRNA gene of *Coxiella burnetii*: Two self-splicing group I introns flank a 34-base-pair exon, and one element lacks the canonical omega G. *J Bacteriol* **189**: 6572-6579.

Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A.M. (1993) Group I and Group II introns. *FASEB J* **7**: 15-24.

Salman, V., Amann, R., Girnth, A.-C., Polerecky, L., Bailey, J.V., Høgslund, S. et al. (2011) A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. *Syst Appl Microbiol (in press)*.

Schmid, M., Schmitz-Esser, S., Jetten, M., and Wagner, M. (2001) 16S-23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: implications for phylogeny and *in situ* detection. *Environ Microbiol* **3**: 450-459.

Schulz, H.N. (2006) The genus *Thiomargarita*. In *The Prokaryotes*. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E. (eds): Springer, pp. 1156-1163.

Schulz, H.N., Brinkhoff, T., Ferdelman, T.G., Marine, M.H., Teske, A., and Jørgensen, B.B. (1999) Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* **284**: 493-495.

Sekar, R., Mills, D.K., Remily, E.R., Voss, J.D., and Richardson, L.L. (2006) Microbial communities in the surface mucopolysaccharide layer and the black band microbial mat of black band-diseased *Siderastrea siderea*. *Appl Environ Microbiol* **72**: 5963-5973.

Seshadri, R., Paulsen, I.T., Eisen, J.A., Read, T.D., Nelson, K.E., Nelson, W.C. et al. (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci U S A* **100**: 5455-5460.

Simon, D.M., Clarke, N.A.C., McNeil, B.A., Johnson, I., Pantuso, D., Dai, L.X. et al. (2008) Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA* **14**: 1704-1713.

Stevens, H., and Ulloa, O. (2008) Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol* **10**: 1244-1259.

Suzuki, M.T., and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625-630.

Teske, A., and Nelson, D.C. (2006) The genera *Beggiatoa* and *Thioploca*. In *The Prokaryotes*. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E. (eds): Springer, pp. 784-810.

Toor, N., and Zimmerly, S. (2002) Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA* **8**: 1373-1377.

Toor, N., Keating, K.S., and Pyle, A.M. (2009) Structural insights into RNA splicing. *Curr Opin Struct Biol* **19**: 260-266.

Toro, N. (2003) Bacteria and archaea group II introns: additional mobile genetic elements in the environment. *Environ Microbiol* **5**: 143-151.

Tourasse, N.J., and Kolstø, A.B. (2008) Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution. *Nucl Acids Res* **36**: 4529-4548.

v. Wintzingerode, F., Göbel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213-229.

Van Camp, G., Chapelle, S., and De Wachter, R. (1993) Amplification and sequencing of variable regions in bacterial 23S ribosomal RNA genes with conserved primer sequences. *Curr Microbiol* **27**: 147-151.

Supplemental Information

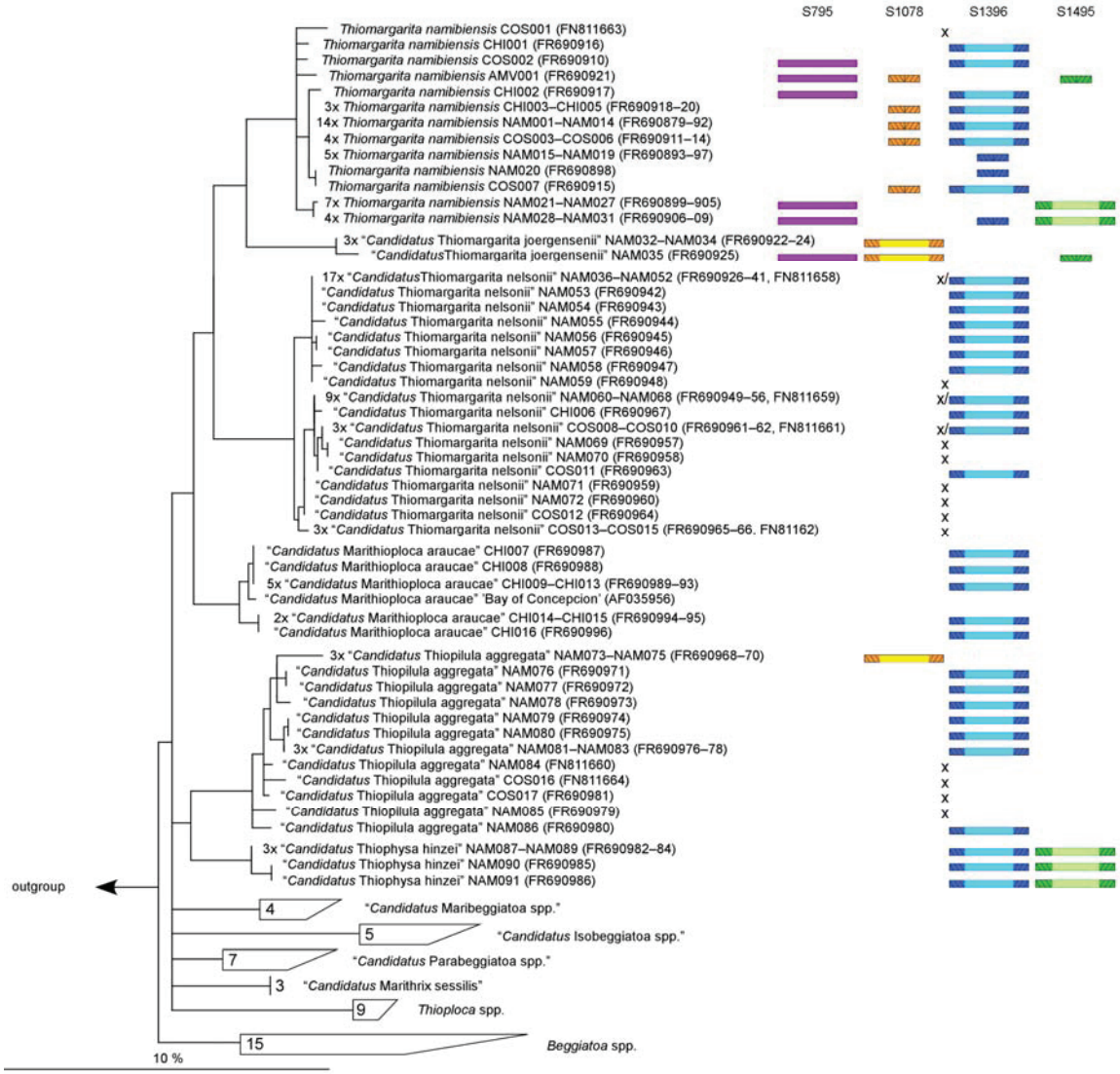


Figure S1. Multifurcation tree adapted from Salman *et al.* (2011) based on nearly full-length 16S rRNA gene sequences of the family *Beggiatoaceae*. Introns were located in 16S rRNA gene sequences subsequent to sequencing of individual cells and filaments. Introns are indicated with horizontal bars aside of each sequence or collection of identical sequences. From left to right, the introns are sorted according to their position in the 16S rRNA gene, from the 5´-end to the 3´-end. ORF-lacking introns are indicated by shorter bars containing diagonal stripes; ORF-containing introns are shown as longer bars with a central inserted non-striped region, representing the ORF. The symbol "x" indicates that no intron was found in the respective sequence(s).

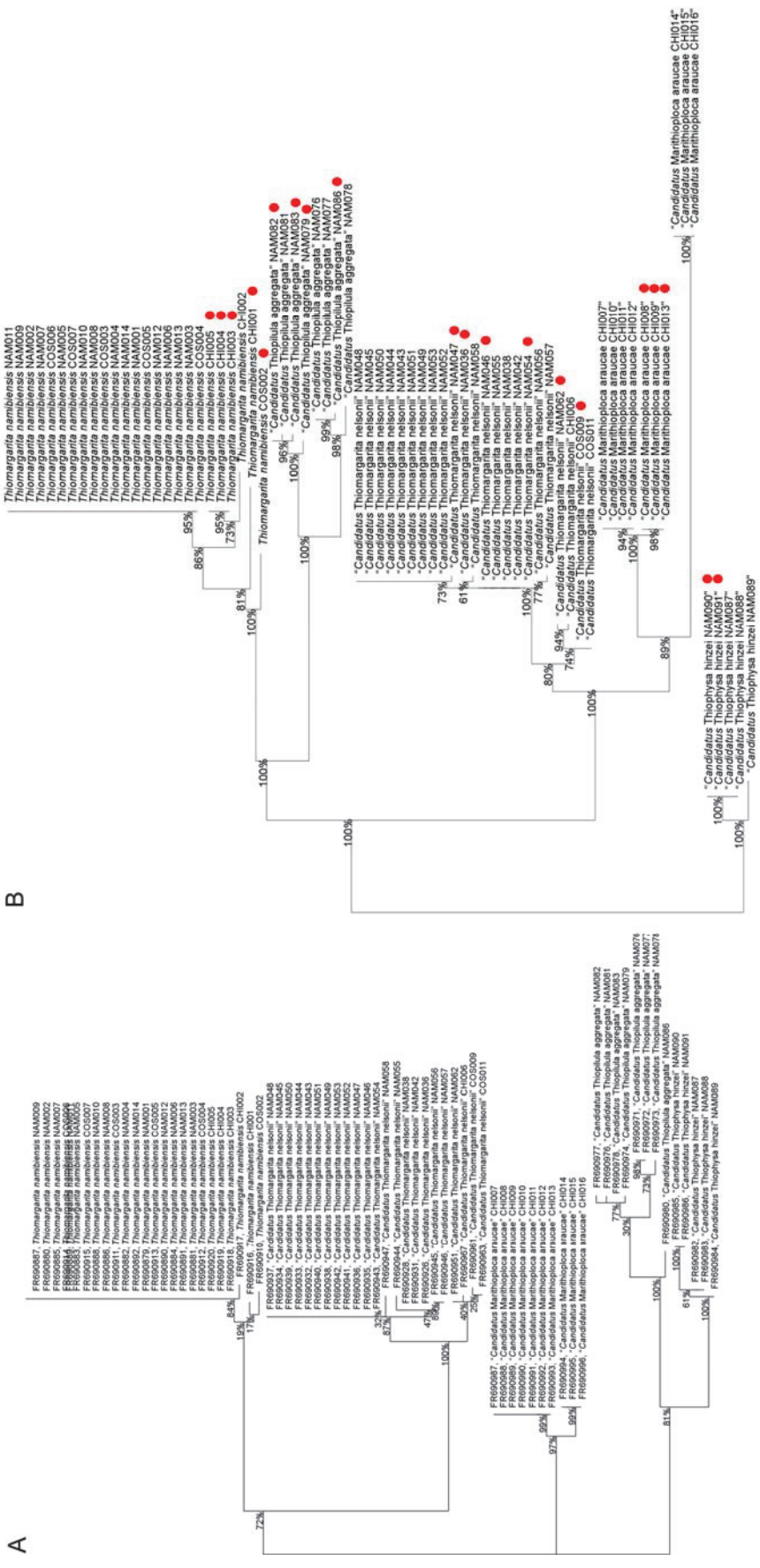


Figure S2. Maximum likelihood trees including bootstraps values of 100 runs of the members of the *Beggiatoaceae* that contain intron S1396 of ~850 nt. (A) Tree calculated with 16S rRNA gene region and (B) tree calculated with S1396 intron region. Within a given species, the tree topologies of (A) and (B) are highly similar and sequence of S1396 are highly conserved, suggesting a vertical heredity of the intron within a species. Red spots highlight minor differences in tree topologies of (A) and (B) within a phylogenetic group. However, the relationship between different species or generic groups is generally different in (A) and (B), indicating a horizontal transfer of S1396 between certain lineages.

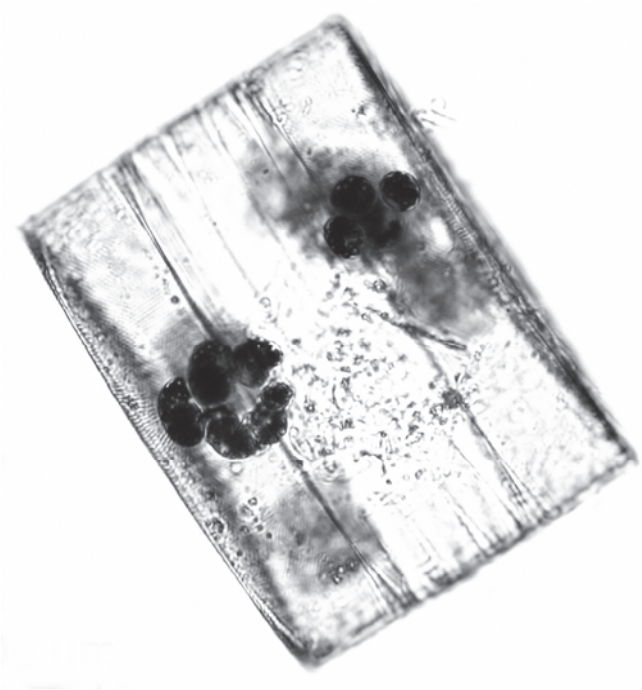
Table S1. Oligonucleotide probes (horseradish peroxidise-conjugated) and helpers (non-labelled) used in this study.

probe	sequence	hybridization temperature	formamide concentration	specificity
VSO1285[a]	5´-cgg act acg agt agt ttt gt-3´	46°C	35%	non-filamentous *Beggiatoaceae*
Thm1061	5´-acg aca cga gct gac gac-3´	46°C	20%	5´-exon adjacent to intron insertion site
Thm1079	5´-gac tta acc caa cat ctc-3´	35°C	20%	3´-exon adjacent to intron insertion site
Thm1078-INT	5´-agt tat tta cga cac g-3´	35°C	10%	5´-exon and 5´-intron
Thm1073	5´-acc caa cat ctc acg aca-3´	46°C	30%	5´-exon and 3´-exon region adjacent to intron insertion site
Thm1380	5´-tgt aca agg ccc ggg aac-3´	46°C	20%	5´-exon adjacent to intron insertion site
Thm1397	5´-atg gtg tga cgg gcg gtg-3´	46°C	20%	3´-exon adjacent to intron insertion site
Thm1397-INT	5´-acg ggc ggt gat tgc g-3´	35°C	10%	5´-exon and 5´-intron
Thm1389	5´-acg ggc ggt gtg tac aag-3´	46°C	35%	5´-exon and 3´-exon region adjacent to intron insertion site
ThmITS30	5´-gat caa gta att tgt gtg g-3´	35°C	10%	ITS1 of *Thiomargarita* spp.
ThmITS174	5´- tct ccc agc tga gct ata-3´	35°C	10%	ITS1 of *Thiomargarita* spp.
ThmITS192	5´-tga atg caa atc agg cgc-3´	35°C	10%	ITS1 of *Thiomargarita* spp.
ThmITS417	5´-cga atg caa tat acc caa g-3´	35°C	10%	ITS1 of *Thiomargarita* spp.
S1078-37[b]	5´-ccg tta taa gct gga ttt-3´	35°C	10%	intron S1078
S1078-52[c]	5´-atc tgt tag ggt ttc acc-3´	35°C	10%	intron S1078
S1078-723[d]	5´-ata aca atc ata aat tta-3´	35°C	10%	intron S1078
S1396-35[e]	5´-gat aga tgg att ttc acc-3´	35°C	10%	intron S1396
S1396-179[f]	5´-ggg ctt tac ggt atg ttg-3´	35°C	10%	intron S1396
S1396-203[g]	5´-tag gct att tct gac att-3´	35°C	10%	intron S1396

helpers	sequence	helpers	sequence
[a]H1285-1267	5´-gag att rgc tcc ccc tcg-3´	[e]HS1396-35-14	5´-ayc tag tct aat aaa ctg ccw-3´
H1285-1249	5´-cgg gtt tgc rac cct ctg-3´	HS1396-35-53	5´-tcm gcy gtc cga aag ck-3´
H1285-1230	5´-trc tac cca ttg tag cac g-3´	[f]HS1396-179-137	5´-agt cac aac gcc ctt tty tg-3´
H1285-1305	5´-gag ttg cag act cca atc-3´	HS1396-179-157	5´-ctt cct yat ggt aca cac atr-3´
[b]HS1078-37-16	5´-gtc tat aga atc gct tct ata-3´	HS1396-179-197	5´-gct att tct gac att ccc agc-3´
HS1078-37-55	5´-catt aat ctg tta ggg ttt ca-3´	[g]HS1396-203-12	5´-agt cta ata aac tgc cwg g-3´
[c]HS1078-52-28	5´-tta taa gct gga ttt gtc tat aga-3´	HS1396-203-122	5´-gcy gag act acc att kg-3´
HS1078-52-71	5´-acg gta ttg ccc tca gca tta-3´	HS1396-203-187	5´-ccc agc ggg ctt tac g-3´
[d]HS1078-723-705	5´-cta aaa tgg atc gat tg-3´	HS1396-203-221	5´-gat atc ccw ggg tca ttt at-3´
HS1078-723-741	5´-aag agc cag cgt gtt-3´		

Chapter 5

General Conclusions



Novel morphologic diversity among large sulfur bacteria

In the past four decades, the exploration of the marine environment led to the discovery of a variety of unknown large sulfur bacteria. Besides the marine, filamentous counterparts to freshwater *Beggiatoa* spp. and *Thioploca* spp., also attached filaments were detected that resembled *Thiothrix* in some traits (Kalanetra *et al.*, 2004). Regarding marine, non-filamentous sulfur bacteria, chain-forming spherical cells were detected and classified as the new genus *Thiomargarita* (Schulz *et al.*, 1999). Soon, other, non-filamentous, marine organisms such as cube-like aggregates (Kalanetra *et al.*, 2005) or spherical single cells (de Beer *et al.*, 2006; Gallardo & Espinoza, 2007; Girnth *et al.*, 2011) were reported. Further diversity studies of Namibian sediments, initiated in my diploma thesis (Gutzke, 2007) and extensively expanded in this PhD thesis, revealed four additional morphotypes of large sulfur bacteria (described in Chapter 2 and 3). None of these novel morphotypes was filamentous, but represented single cells in various states of assembly concomitant with a variety of coatings (Fig. 1).

A morphotype with peculiar features is presented separately in Chapter 2. Here, for the first time, individual attached cells, instead of multicellular attached filaments, were reported within the family *Beggiatoaceae* (Gammaproteobacteria). A sessile state is generally very common in unicellular, benthic bacteria and usually emerges from a free-living, planktonic state (Marshall, 2006). In the group of prosthecate Alphaproteobacteria a more complex life cycle is known, which resembles that proposed for the novel, attached large sulfur bacterium. In the prosthecate Alphaproteobacteria, a rod-shaped swarmer cell attaches to a surface and elongates by producing a stalk. This cell is reproductive and undergoes asymmetric cell division, producing the planktonic, motile swarmer cells (Poindexter, 2006). Also the novel, attached sulfur bacterium is considered to result from a smaller, possibly motile, cell. Subsequent to attachment, the cell is regarded to elongate and form apical buds, which are released as free-living cells into the sediment (Chapter 2). This process of asymmetric division is furthermore reminiscent of gonidia-production in filaments of the closely related "*Candidatus* Marithrix" within the same family or of more distantly related organisms like *Thiothrix*, *Leucothrix* and certain cyanobacteria. Possibly, the discovered attached cells and their filamentous equivalents follow a similar strategy with a comparable dimorphic lifestyle. Apparently, the ability to develop attached cells or filaments with distinct poles during a certain life stage evolved more than once in evolution.

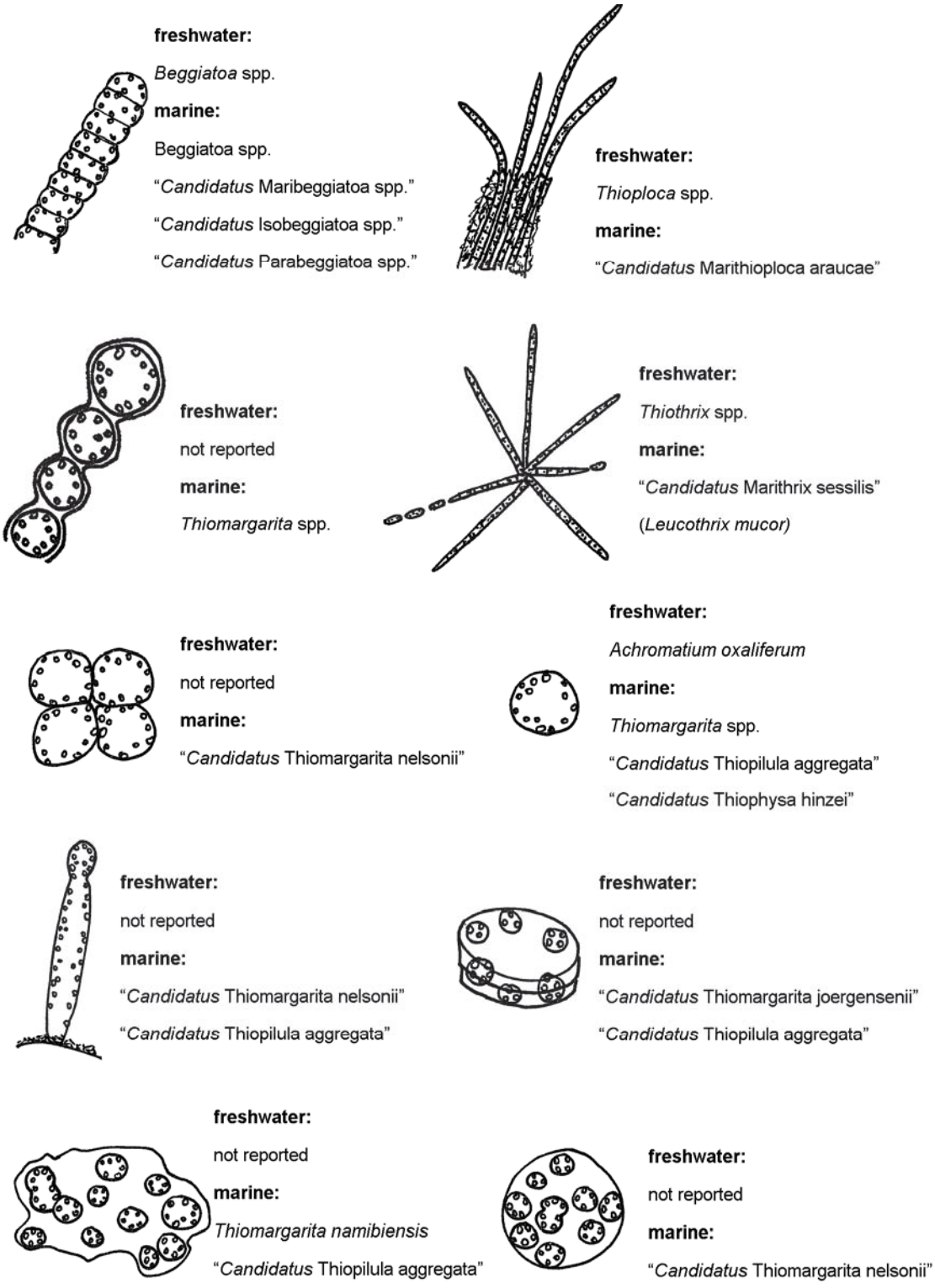


Figure 1. Ten different morphotypes were identified so far among the large sulfur bacteria. The upper six types have been reported prior to this thesis. The lower four were newly characterized in this thesis. The phylogeny of the different morphotypes and their nomenclature are not correlating. Instead, the same morphology can be found in several taxa. Six morphotypes have so far been exclusively detected in marine environments.

Apart from this attached morphotype, three additional morphologies among the non-filamentous sulfur bacteria were described (Chapter 3). These types comprise spherical cells residing inside dead diatom frustules, a thick mucus layer or a rigid envelope (Fig. 1). In each of the three novel aggregation types, division stages and cells with varying diameters were found. Sequencing of the 16S rRNA gene resulted in identical sequences for different cells of a given aggregate (data not shown). This suggests that all cells of an aggregate are clonal and, by differences in growth and reproduction, may differ in cell size. This is in accordance to non-filamentous large sulfur bacteria arranged in chains, in which individual cells may have different sizes, but feature identical 16S rRNA gene sequence as well (data not shown).

Moreover, during this thesis a sporadic rolling motility was observed among many different morphotypes of non-filamentous large sulfur bacteria (Chapter 3). This form of motility was nearly unrecognized until today and was described only in the genus *Achromatium* (*Thiophysa*) (Hinze, 1903; Schewiakoff 1892). Now, it could be revealed that this characteristic is a common feature among the non-filamentous large sulfur bacteria, including *Thiomargarita* spp. and two other generic groups (Chapter 3).

The diversity of large sulfur bacteria in marine environments, as revealed in this thesis, seems to be greater than previously recognized. It remains to be determined whether there exist respective morphological analogs in freshwater habitats (Fig. 1). Moreover, results presented in Chapter 3 showed that morphological analogs frequently represent two or even multiple taxa (e.g. gliding, free-living filaments). In certain taxa, morphologies may even be intermixed (e.g. *Thiomargarita namibiensis*). This implies that morphological analogs most likely did not evolve from a common ancestor featuring the respective morphological trait, but that they possibly evolved several times independently in the course of sulfur bacterial evolution. Therefore, the putative morphology of the ancestral cell in the group of large sulfur bacteria remains ambiguous.

Phylogenetic diversity among large sulfur bacteria

By inferring relationships and phylogenetic ancestry from the molecular data analyzed during this PhD thesis, it could be shown that the classification within the group of large sulfur bacteria was in need for revision and that several new taxa had to be introduced (Chapter 3). In the past six years, the novel family name *Thiotrichaceae* (Garrity *et al*., 2005) has already had a major impact on current taxonomic applications. It was

validated by the IJSEM (Euzéby, 2005) and is therefore used for sequences submitted to the GenBank/EMBL/DDBJ databases, which are imported into frequently used SILVA database and ARB software. Nevertheless, in Chapter 3 it is proposed that the family name *Beggiatoaceae* should be revived because it is the historically and phylogenetically correct name for the monophyletic family of organisms related to the genus *Beggiatoa* (Chapter 3). Furthermore, at least ten genera could be determined within the *Beggiatoaceae*, which is an astonishing result regarding that a classification of the family based on traditional, morphology-derived characteristics would include merely three genera (*Beggiatoa*, *Thioploca* and *Thiomargarita*, see Chapter 1, Fig. 17). The new classification proposed in Chapter 3 is well supported by the modern guidelines: the taxa are classified according to 16S rRNA gene identity values corresponding to standard cut-off values (see Table 1 in Chapter 1 and Table 1 in Chapter 3), and each defined taxon shows a clear monophyletic relationship in the reconstructed phylogenetic tree (Chapter 3, Fig. 1).

The reclassification presented in this thesis aimed for the creation of a system, which can be easily applied and expanded in the future. Therefore, morphological traits were disregarded as parameters for classification because they were misleading (Chapter 3). Instead, a new nomenclature was proposed based on reconstructed tree topology and calculated identity values (Chapter 3). A branch not being monophyletic with a branch including either a type species or with the first sequence published for a given genus, received a new name. The nomenclature was mostly guided by morphological characteristics or historical information, however, the classification itself was unambiguously based on molecular information. A putative classification on species level was generally avoided because a species differentiation only based on 16S rRNA gene data is doubtful (introduced in Chapter 1, Fox *et al.*, 1977b; Fox *et al.*, 1992; Stackebrandt & Goebel, 1994; Tindall *et al.*, 2010). The genus *Thiomargarita* was an exception, because the amount of data was extremely high and the proposed species differentiation was also supported by ITS sequences.

Implications on the future taxonomy of large sulfur bacteria

As presented above, there are two major outcomes of the present study: first, there is an extensive morphological diversity among the large sulfur bacteria in marine environments and second, also a vast phylogenetic diversity among these bacteria could be detected. Accordingly, both characteristics addressed in this thesis turned out to be much

more diverse than previously recognized. Within the *Beggiatoaceae*, ten morphotypes are described instead of six (Fig. 1) and ten genera are classified instead of three (compare Fig. 17 in Chapter 1 and Fig. 1 in Chapter 3). Generally, the finding of a greater morphological diversity among the large sulfur bacteria does not necessarily simplify the handling of this group of bacteria, but rather demonstrates that this feature is evermore tending to mislead the observer with respect to classification and identification of the organism. On the other hand, the detected vast phylogenetic diversity indicates a much higher and clearer resolution for future applications.

Already in earlier studies, morphologically identical large sulfur bacteria were reported to phylogenetically segregate into different putative genera (Ahmad *et al.*, 1999; Ahmad *et al.*, 2006). Now, also the opposite is revealed: organisms of extremely diverse morphology can be assigned to the exact same species when analyzing their 16S rRNA genes and ITS regions (Chapter 3, e.g. *Thiomargarita namibiensis*). Furthermore, earlier attempts to classify large sulfur bacteria according to their diameters or habitat salinity remains difficult when considering the blending of these features within certain taxonomic groups (Chapter 3). So far, only members of the *Beggiatoaceae*, which populate freshwater habitats, can be clearly distinguished based the duality of habitat salinity and morphological characteristic (certain *Beggiatoa* spp. and *Thioploca* spp.).

These conflicts underpin the need for new categories, on which identification and classification of large sulfur bacteria should be based on in the future. Apparently, the evolutionary differences among the different types of large sulfur bacteria might not be fully reflected by the information gained when comparing morphological features, in contrast to the insight gained when analyzing certain physiological features or genomic regions. Regarding the lack of pure cultures for most representatives of the *Beggiatoaceae*, it is proposed to henceforth base the classification of the large sulfur bacteria mainly on molecular data (16S rRNA and ITS sequences, Chapter 3) as it is the current standard in the classification of bacteria. For confirmation and surely also improvement of the novel classification, other features should be examined, according to the multiphasic approach (Feil, 2004; Gevers *et al.*, 2005; Stackebrandt & Goebel, 1994; Stackebrandt *et al.*, 2002; Tindall *et al.*, 2010). This can include sequencing of additional genomic regions like the ITS region as presented in Chapter 3 or other functional genes like those required for sulfide-oxidation. Furthermore, it can involve characterization of

distinct tolerance thresholds and flux requirements for oxygen and sulfide or the determination of storage capacities for nitrate, polyphosphate or other electron donors and acceptors.

Certainly, the here-proposed systematic is not the ultimate truth and following the rules of the Bacteriological Code (Lapage *et al.*, 1992), it can unfortunately not be contributed more than proposing *Candidatus* taxa. Nevertheless, the proposal is surely a great improvement to the previous, morphology-based classification and marks the first serious step towards a reclassification of the large sulfur bacteria already requested several times before (Ahmad *et al.*, 1999; Ahmad *et al.*, 2006; Mussmann *et al.*, 2003). The novel nomenclature will prevent confusion and misinterpretation of the different types of sulfur bacteria in the future and it simplifies reference to a certain taxonomic group. The large amount of novel, high-quality phylogenetic data aids in identifying or classifying newly detected large sulfur bacteria.

In addition, an improved classification is necessary to specifically assign the diverse ecotypes, which are frequently detected and characterized in different habitats, to a distinct phylogenetic group. A fast and easy tool to identify organisms or specific phylogenetic groups in their natural environment along with direct information about their abundance is fluorescence *in situ* hybridization (FISH, Amann *et al.*, 1995). Here, probes that are specific for one or several of the novel sulfur bacterial taxa were designed (Table 1) and can be applied in future studies for the general detection of these groups in a given habitat. Their specific abundance at certain locations or depths can thus be detected. The obtained results can be combined with a detailed characterization of the respective habitat, for example when using microsensors. A combination of these data could support the determination of putative ecological preferences for certain phylogenetic groups of large sulfur bacteria. Eventually, the direct assignment of specific ecological parameters to a certain phylogenetic group may enable a specific description of the organisms subsumed in that group and could finally lead to specifically modulated approaches for successful culturing.

Table 1. FISH probes targeting the 16S rRNA in members of the *Beggiatoaceae*. The specificities of previously published probes were recalculated including also the phylogenetic data obtained in this thesis. Calculation of probe sequences was performed with the ARB ProbeDesign tool and specificities were tested against the RDP database. The group containing *Beggiatoa* spp., Cluster 'XII' in Chapter 3, is highly diverse and thus a common probe cannot be designed. Nearly complete specificity for this group as well as for the entire *Beggiatoaceae*, can be achieved when mixing different group-specific probes.

probe	probe specificity	probe sequence (5´− 3´)	T_h /FA (%)[a]	reference
VSO673	"*Candidatus* T. nelsonii", 50% of "*C*. Parabeggiatoa spp.", "*C*. Marithioploca sp.", "*C*. Maribeggiatoa spp.", "*C*. Thiopilula sp.", "*C*. Thiophysa sp." and 5 additional non-target sequences	CGCTTCCCTCTACTGTAC	46 / 35	(Kalanetra *et al.*, 2004)
VSO1284	non-filamentous *Beggiatoaceae*	CGGACTACGAGTAGTTTTGT	NA / NA	Chapter 4
VSO1287	*Thiomargarita* spp., "*C*. Thiopilula sp.", "*C*. Thiophysa sp.", "*C*. Maribeggiatoa spp.", "*C*. Parabeggiatoa spp.", and 1 additional non-target sequence, excluding "*Candidatus* Maribeggiatoa sp." 'Limfjorden', "*Candidatus* Parabeggiatoa sp." 'Carmel Canyon'	CCGGACTACGAGTAGTTT	NA / NA	this thesis
Thm465	~30% of *T. namibiensis*	GTCAAGACTCTAGGGTAT	46 / 35	(Girnth *et al.*, 2011)
TMnam1135	*T. namibiensis*	TCTAGAGTTCCCGACCTA	NA / NA	this thesis
TMjoe460	"*C*. T. joergensenii"	GATTCTAGGGTATTGGCC	NA / NA	this thesis
Thm482	~86% of "*C*. T. nelsonii"	CTTCTTCTATTGCTGATG	46 / NA	(Grünke *et al.*, 2010)
TMG-831	70% of *Thiomargarita* spp. and 2 additional non-target sequences	GGATCAATTTCCCCCAAC	NA / NA	this thesis
Biwa829	*Thioploca* spp.	AGGTATACCCTTCCAACGTC	46 / 20	(Kojima *et al.*, 2003)
829-Thioploca	"*C*. Marithioploca sp." some "*C*. T. joergensenii"	GGATTAATTTCCCCCAACATC	46 / 20	(Teske *et al.*, 1995)
MTPL176	"*C*. Marithioploca sp."	CGGACTACGAGTAGCTTT	NA / NA	this thesis
MBeg176	"*C*. Maribeggiatoa spp." and 1 aditional non-target sequence	CTCCCATAGGACTTATGC	NA / NA	this thesis
Blim193	AF532771 and AF532775 in "*C*. Isobeggiatoa spp."	AAAAGACGCCCTCTTTCC	46 / NA	(Mussmann *et al.*, 2003)
IBeg1167	"*C*. Isobeggiatoa spp."	TCACCTTCCGTCGGTTTG	NA / NA	this thesis
Blim575	"*C*. Parabeggiatoa spp." and 36 additional non-target sequences	CTAGCCGCCTACATACGC	46 / NA	(Mussmann *et al.*, 2003)
PBeg845	"*C*. Parabeggiatoa spp." and 1 additional non-target sequence	GCTACGACACTAAAAGGC	NA / NA	this thesis
TPI844	"*C*. Thiopilula sp."	CTGCGACACTAAAGGATC	NA / NA	this thesis
TPH997	"*C*. Thiophysa sp."	TTCAGCAAGGTTCCGAGG	NA / NA	this thesis
MTX834	"*C*. Marithrix sessilis"	AAAGGGCAAAATCCCTCC	NA / NA	this thesis
BegF830	freshwater *Beggiatoa* spp., excluding *Beggiatoa* strains LPN, 1401-13 and AA5A	AGACAGTCTCTCCCAACA	NA / NA	this thesis
BegAM184	marine *Beggiatoa* spp. and Arauama sequences, excluding *Beggiatoa* strain MS-81-1c	CCATCTTTTACCCGTAGG	NA / NA	this thesis
BegC730	*Beggiatoa* sp. 'Chiprana'	ATCAGTCCAGATAGTCGC	NA / NA	this thesis

[a] T_h=hybridization temperature, FA=formamide concentration in hybridization buffer, NA=not available

Large sulfur bacteria lack a confined biogeography but specialized morphologies reflect adaptation to habitats

Within a phylogenetically defined species, sequence data are highly identical. However, from a close genetic relationship the origin of organisms from nearby geographic sites cannot be directly deduced. At least within the *Beggiatoaceae*, it is often observed that organisms belonging to the same species actually originate from extremely distant habitats (Chapter 3). For example, nearly identical 16S rRNA genes (>99.8%) among thioplocas originating from freshwater lakes in Germany and Japan (Kojima *et al.*, 2003) and from Danish fjords (Høgslund *et al.*, 2010) were detected. These closely related organisms apparently occur in habitats located on different continents, not being directly connected. Furthermore, certain marine filaments contain 16S rRNA genes that are by 95−99% identical ("*Candidatus* Isobeggiatoa spp.") and occur in coastal marine regions of Germany, Japan and Chile (Kojima & Fukui, 2003; Mussmann *et al.*, 2003) or in fjords of Denmark and Svalbard (Jørgensen *et al.*, 2010; Mussmann *et al.*, 2003). The finding of cosmopolitan taxa is not new in the current understanding of microbial biogeography (Fenchel, 2003) and meets the postulate of 'everything is everywhere, but, the environment selects' (Baas Becking, 1934).

One reason for this putative absence of a confined biogeographical distribution of sulfur bacterial taxa suggests a frequent dispersal of these organisms with respect to evolution. This dispersal can be explained fairly easy for marine habitats, even on a global scale, because water masses are frequently displaced by natural currents and, to a lesser extent, by ballast water in container ships. Upwelling areas like in the Gulf of Mexico or at the Namibian coast are connected by currents along the equatorial Atlantic belt and the spreading of even larger benthic organisms than bacteria along this current has been suggested (Olu-Le Roy et al. 2007). Evermore, Mediterranean Sea and North Sea water masses are naturally connected with the Atlantic Ocean via the Gulf Stream. Therefore, recent dispersal of benthic organisms including small and large bacteria across the oceans is highly possible. Furthermore, the finding that at least some of the large sulfur bacteria are able to produce smaller life-forms, which are dispersed much more efficiently, strongly supports the hypothesis of currently ongoing spreading events. For freshwater habitats, however, a spatial connection remains speculative, and the only recent ways of dispersal might be through aerosols and feathers of birds or fur of animals.

A second scenario, explaining the relative similarity of widely distributed organisms, is an extremely slow mutation rate, subsequent to a single or few dispersal events in ancient times. This implies that the niches, which are currently inhabited by almost identical organisms, are similar in their overall composition, because the natural selection seemingly resulted in a similar genetic pool. In both postulated scenarios, possible adaptations to micro-niches at the various geographic sites could have happened, however, these adaptations may not have manifested as mutations in the 16S rRNA gene sequences. This assumption is based on recent findings of other closely related strains that do not show speciation, as inferred from 16S rRNA comparison, according to their geographic origin. However, when analyzing additional features, like protein-coding genes in *Sulfolobus* sp. (Whitaker *et al.*, 2003) or the metabolite spectrum in *Salinibacter* sp. (Rossello-Mora *et al.*, 2008), a correlation with the organisms' geographic location could be obtained.

Inversely, not only extremely similar genes at distant sites are found, but also a high genetic diversity at a single sampling site was detected (Chapter 3). Obviously, whereas certain lineages remained nearly unchanged in their 16S rRNA gene sequence, others show cumulative speciation events. This finding supports the assumption that a given site does have a selective pressure on sulfur bacterial evolution, and that only in some cases these adaptations seem to be reflected in the rRNA genes. When detecting a divergence in 16S rRNA genes at a given site, an overall genetic variance can also be postulated for coding genes. Thus, the observed speciation within sulfur bacterial communities might reflect certain physiological properties or other phenotypic traits that are not yet documented. Certainly, much information for a detailed physiological characterization of the respective ecotypes is still missing, only allowing speculation. Cells might exhibit unique physiological properties, like different ranges of tolerance towards oxygen or sulfide, varying storage capacities of energy reserves or differing modes or speed of mobility. These kinds of adaptation to certain niches surely are manifested in molecular differences between the organisms and it remains to be tested, which genetic regions are affected. Just as well, detailed morphological traits are at the current state not yet fully examined and, surely, additional knowledge about features like intracellular compartments, membrane and cell wall composition or extracellular substances can reveal further differences between the respective types of cells.

For most previously reported types of large sulfur bacteria, adaptation of morphology to the natural habitat was postulated, involving means of mobility, levels of tolerance to oxygen and sulfide as well as the degree of aggregation (Girnth *et al.*, 2011; Kalanetra *et al.*, 2004; La Riviere & Schmidt, 2006; Schulz, 2006; Teske & Nelson, 2006). When speculating about possible evolutionary developments regarding morphological complexity within this group of bacteria, it is tempting to suggest that a simple form like a single, spherical cell is the basic form. Cellular aggregates or chains appear more advanced and a filament may be even more complex, also in terms of coordination of cell division and motility. Bundles of multiple filaments thus probably constitute the most complex life form. However, it cannot be clearly stated that certain sulfur bacteria are more advanced than others because some species exhibit life cycles that interconnect cellular types and aggregation modes. Morphologies can merge from one into another, as exemplified by the formation of free-living single cells of attached filaments (Kalanetra *et al.*, 2004; Larkin & Shinabarger, 1983) or of a free-living spherical cell from an attached elongated cell (Chapter 2). Moreover, the finding of putative reductive division in "*Candidatus* Thiomargarita nelsonii" represents a change in morphology, in this case in size. The life strategy of large sulfur bacteria is considered to be not opportunistic and the increased tolerance towards oxygen and sulfide in non-filamentous types reflects their diminished capabilities in mobility. Apparently, these non-filamentous large sulfur bacteria, which cannot adjust their position in a constantly changing habitat efficiently by fast movement, possibly developed a great spectrum of morphologies, which are to some extent interchangeable to enable adjustment.

Many large sulfur bacteria occur in aggregates and have the ability to produce acertain kind of mucous sheath. A close association of cells or filaments can surely represent a selective advantage in certain situations or habitats. An advantage of bundle-formation in filamentous representatives, which might be the highest form of aggregation, certainly is the constant orientation within the substrate gradients allowing a more efficient movement over longer distances (Schulz *et al*., 1996). Among the non-filamentous sulfur bacteria, aggregation could serve as a protection against drift of single cells or it could help in burying the aggregate into deeper layers of the sulfidic sediment subsequent to sediment resuspension, because a large aggregate will settle faster than a single cell. Generally, certain types of coatings can serve as diffusion boundary regarding the putative sharing of exoenzymes or siderophores. The close proximity of

the cells might also enable a better communication, e.g. for performing a coordinated cell division, which may be mediated via quorum sensing.

Large sulfur bacteria populating empty diatom frustules (Chapter 3) surely represent a unique example for aggregation. So far, the inhabitation of only completely empty frustules was observed and most likely the cells invaded the silicate remains when the viable part of the diatom itself has already died. The most stunning observation in this respect is that the populated frustules are still intact with the two thecae being still perfectly nested. Thus, the mechanism of how the cells entered the frustule seems unclear. Possibly, there exist smaller variants of the cells, which can penetrate the pores (<1 µm in diameter). The ecological reason for the inhabitation of empty silicate frustules remains unknown. It can be speculated that these cells cannot produce any kind of coating themselves and thus populate the remains of diatoms and foraminifera (data not shown) for protection.

Many ecological adaptations in certain types of large sulfur bacteria have already been determined, and there are definitely many more to be detected. Although not yet fully unraveled, it can be speculated that minor ecological differences allow the frequent co-occurrence of different groups of large sulfur bacteria frequently in certain habitats. There are examples of a co-occurring free-living and bundle-forming filaments (Schulz *et al.*, 1996) and also nearly all types of non-filamentous types were found at one site (Namibia, Chapter 3). Although exploiting the same resources, these bacteria do not live a competitive life-mode, which can be explained by the fact that large sulfur bacteria are not entirely dependent on constant nutrient uptake. They have the ability to produce intracellular storage compounds and to sustain a very slow metabolism (Schulz and Jørgensen, 2001). After all, highly restrictive habitats in terms of fluidity of the sediment or fluxes of oxygen and sulfide, are selecting for certain types of large sulfur bacteria, which is mostly determined by their different levels of tolerance for these compounds, but moderate conditions may allow for the co-occurrence of different types.

Introns in the 16S rRNA genes

Sequences of the small ribosomal subunit (16S rRNA) are nowadays used as the basis for the classification of all living organisms and thus large databases exist, including a vast amount of sequences from environmental samples. This gene was considered particu-larly useful as it is present in all living organisms, is not subject to horizontal

gene transfer and, most importantly, has a conserved length (recently reviewed in Pace, 2009). However, the last attribute has now come into question as 16S rRNA genes of prokaryotes can apparently be much longer than usual due to the presence of large introns (Chapter 4). This fact is one among many others that weakens the previously considered 'universal' suitability of 16S rRNA genes for diversity studies. This conflict is caused by the fact that the study of the small ribosomal subunit usually involves polymerase-chain-reaction (PCR). However, this technique is frequently affected by systematic biases (Chandler *et al.*, 1997; Cilia *et al.*, 1996; Coenye & Vandamme, 2003; Dennis *et al.*, 1998; Frank *et al.*, 2008; Hong *et al.*, 2009; Suzuki & Giovannoni, 1996; v. Wintzingerode *et al.*, 1997), and now also the bias caused by length heterogeneity of homologous genes (Chandler *et al.*, 1997) needs to be considered when amplifying 16S rRNA genes with PCR. The group of large sulfur bacteria is the first, in which numerous, large introns were detected in the 16S rRNA gene and thus comprises a group, which is systematically discriminated in universal PCRs during diversity studies. The absence of their sequences from clone libraries was shown several times (Angert *et al.*, 1998; Edgcomb *et al.*, 2002; Gillan *et al.*, 1998; Lopez-Garcia *et al.*, 2003; Sekar *et al.*, 2006; Stevens & Ulloa, 2008) and although it is unknown, how many additional bacteria contain introns in their 16S rRNA genes, it seems plausable that other groups of microorganisms are as well regularly discriminated in microbial diversity studies.

The presence of introns in 16S rRNA genes (and also in 23S rRNA genes as shown in one case) is a common feature among the large sulfur bacteria. This fact together with the systematic discrimination of longer template genes over shorter ones in a mixed sample (Chapter 4) offer a straight forward explanation for the common difficulty or even failure in producing full-length 16S rRNA sequences for *Thiomargarita* or "*Candidatus* Mari-thioploca" in previous studies. In these two genera, only the first half of the gene was successfully sequenced before (Girnth *et al.*, 2011; Kalanetra *et al.*, 2005; Schulz *et al.*, 1999; Teske *et al.*, 1995), which is to current knowledge the intron-free part of the gene. Although amplification in these studies was similar to the single-cell PCR approach presented here, sequencing was only partially successful. Reasons can be a failure of amplification due to the length heterogeneity bias when not using a specific primer or the rejection of an amplicon, which was longer than expected.

The mechanisms of dispersal of bacterial introns within the host genome are well studied and considered to be highly efficient (recently reviewed in Nielsen & Johansen, 2009; Toor *et al.*, 2009). In sulfur bacteria, it can only be speculated that introns are present at other sites in the genome than the *rrn* operon. The only sequenced genome of a large sulfur bacterium, which is publicly available (Mussmann *et al.*, 2007) does not contain introns. An intron usually has a certain target sequence, in which it inserts via a process termed homing (Lambowitz & Belfort, 1993). Assuming that large sulfur bacteria contain multiple rRNA operons, a given intron could have at least inserted into every such target sequence present in the rRNA gene copies. In this case, a possible length heterogeneity PCR bias has to be also considered, because the different gene copies could also be heterogeneous with respect to carrying introns. Accordingly, a PCR would yield a single, intron-containing amplicon only when all intron target sites in the host are occupied, and it can be assumed that as long as the host includes intron-containing and intron-lacking gene copies, the shorter molecules would be preferentially amplified. This implies that whenver introns were detected in this thesis, the different types of introns have already inserted into every target sequence of the rRNA operons. In those cases, where a certain type of intron was not present in the amplicon, the cell either did not contain this intron in the respective genes or it contained a mix of inserted and non-inserted intron types, which resulted in the amplification of the shortest possible molecule from a given cell. In order to test this hypothesis, a full genome sequence of a single, intron-containing cell is surely needed.

Introns can be considered as ‘selfish mobile elements’ that are capable of mediating their own mobility in and out of host genomes. Ever since introns were detected in bacteria, questions about their biological function arose. The basic requirement for persistence within a host genome is seemingly to not have any negative impact on the host´s reproductivity. This is especially relevant for host organisms that continuously compete with other organisms for niche succession and that multiply rapidly. For at least two reasons, the establishment of introns in these organisms is very unlikely. First, an invasion by an intron is possibly noticed quickly as DNA replication and DNA repair happen on a fast scale, thus foreign DNA is likely detected at an early stage. Second, the replication of excess DNA requires more energy, and competitive organisms can most likely not afford this. Therefore, it is not surprising that bacteria with short generation times tend to reduce their genomes to the minimum information. The large sulfur

bacteria clearly do not fit into this category of microorganisms as they are capable of storing both their electron donor and acceptor in high amounts and accomplish a life strategy, which includes long periods of persistence and a seemingly slow metabolism (Schulz & Jørgensen, 2001; Schulz, 2006). They are considered to have long generation times (possibly weeks to months) and they do not appear to live a competitive life mode because they are found frequently co-occurring in one habitat. Instead of minimizing their genomic content, they rather multiply it and contain several thousand nucleoids (Lane and Martin, 2010), which could lead to a minimized repair of single gene copies. For all these reasons, large sulfur bacteria might be destined to accumulate foreign DNA like introns. Possibly, also other bacteria with comparable life modes could contain such introns. An example supporting this hypothesis is *Coxiella burnetii*, a bacterium, which features multiple introns in its 23S rRNA gene. It is a potential pathogenic parasite experiencing a surplus of nutrients and no competitors in infected host cells, and it is very robust when persisting outside of the host. It has a genome far larger than expected as parasites usually undergo reductive evolution and minimize their genomes (Raghavan *et al.*, 2007).

There are few studies, which report positive effects of intron presence in a host organism. Some group I introns were shown to assists in the correct folding of tRNAs in *Azoarcus* (Rangan *et al.*, 2004) and of rRNAs in *Tetrahymena* (Cao & Woodson, 1998; Cao & Woodson, 2000). A group I intron in a *Bacillus* phage is considered to be beneficial to the host because it destroys the genomes of competing phages (Goodrich-Blair & Shub, 1996). Efficient splicing of a group II intron resulted in the ligation of two exon messages, which are then translated into a protein that mediates conjugation of a plasmid in *Lactococci* (Mills *et al.*, 1996). Apart of a putative co-evolution of introns and their hosts (Nielsen & Johansen, 2009) also negative impacts of intron presence on host fitness are observed. A group I intron in the 23S rRNA gene of *Simkania negevensis* was reported to not splice from precursor RNA, which resulted in a delayed growth by affecting protein biosynthesis (Everett *et al.*, 1999). Furthermore, free RNA introns were shown to bind non-covalently to their target sequence in a native ribosome and thus prevented protein synthesis by direct inhibition, also resulting in slow cell growth (Nikolcheva & Woodson, 1997). Implications of introns on the host organism, as those presented here as an excerpt of possibilities, remain to be determined for the intron-containing large sulfur bacteria.

The insertion of introns into rRNA genes is not the only example for the extension of these genes. Access nucleotides, called intervening sequences, were reported from bacterial 16S rRNA genes (Linton *et al.*, 1994a; Linton *et al.*, 1994b; Rainey *et al.*, 1996; Springer *et al.*, 1993) and 23S rRNA genes (Burgin *et al.*, 1990; Raghavan *et al.*, 2007; Ralph & McClelland, 1993). They are usually 100−450 nt long, but can be up to 759 nt and may contain an open reading frame (Ralph & McClelland, 1993). In contrast to introns, intervening sequences are not self-catalytic and are thus removed via cellular enzymes, like RNAse III, commonly performing rRNA precursor maturation (Burgin *et al.*, 1990). During this process, the two exons are not ligated and the respective rRNAs have been frequently found to remain fragmented (Burgin *et al.*, 1990; Linton *et al.*, 1994a; Raghavan *et al.*, 2007; Rainey *et al.*, 1996; Ralph & McClelland, 1993; Springer *et al.*, 1993). The fragmented molecule is considered fully functional because no negative effect on ribosome formation or protein synthesis was observed (Burgin *et al.*, 1990; Raghavan *et al.*, 2007; Ralph & McClelland, 1993). Nevertheless, the detection of long intervening sequences has never provoked a reconsideration of PCR-based studies of microbial diversity, as has the detection of introns in this thesis (Chapter 4).

Enlarged 16S rRNA genes containing intervening sequences were so far only found in pure cultures. In one case, cells of such a culture contained several *rrn* operons with two types of 16S rRNA genes that differed by 150 nt. Both these homologues could be simultaneously amplified from the heterogeneous sample (Linton *et al.*, 1994b). However, as presented in Chapter 4, when amplifying homologous genes that differ by 850 nt, a discrimination of the longer gene is observed. This suggests that there exists a threshold regarding length difference, which determines the discrimination of the longer homologue. This hypothesis was further tested using equal amounts of homologous genes that differ by 700 nt and here, a faint band of the longer amplicon was visible along with the smaller band on a gel (Chapter 4). The larger amplicon was absent again, when providing the shorter gene in a ten times higher concentration (data not shown). This moreover suggests, that the absolute ratio of the two homologues might also be relevant for the discrimination effect. This concentration-dependent PCR bias is known for genes with equal lengths, resulting in the increased amplification of the less concentrated homologue (Suzuki & Giovannoni, 1996). The result presented here, however, reveals the opposite. Supposedly, the length heterogeneity bias can be dominating over the concentration bias.

16S rRNA introns were, interestingly, not first detected in a pure culture, even though sequencing of clones bears principally the highest possibility to overcome a length heterogeneity bias. Instead, a single-cell PCR approach led to the detection of the introns. This single-cell PCR on representatives of the large sulfur bacteria was initially applied for a different purpose, namely the correlation of cellular morphology and phylogenetic affiliation (Chapter 3). Therefore, the PCR reaction did not literally contain a single bacterial cell, but a single large sulfur bacterium, which was usually contaminated with many other bacteria. Then, a specific primer was used to ensure targeting the 16S rRNA gene of the large sulfur bacterium. In fact, the combination of both these actions was fundamental for the detection of introns in this group. Not every individual of a sulfur bacterial population contains introns in the 16S rRNA genes (Chapter 3). If a bulk DNA extract of a sulfur bacteria-containing sample was used for amplification, even when using a group-specific primer, sequences lacking introns can be expected to be preferentially amplified due to the length heterogeneity bias. In turn, if a single sulfur bacterium was amplified using two universal primers, again only the shortest gene without an intron would be retrieved, which would be likely those of the contaminants. This last assumption was tested by removing several clonal *Thiomargarita* cells from a chain, washing them for trying to eliminate all possible contaminants and applying universal PCRs. In few cases, only the expected, larger 16S rRNA gene sequence containing the intron was amplified (sequences NAM056, NAM057), but in most cases, the regular band of ~1500 nt was retrieved, which needed to be cloned and sequences affiliated with large sulfur bacteria could not be found (data not shown).

Besides the above named prerequisites, which enable the detection of introns in 16S rRNA genes, it is also important include PCR products of unexpected lengths in the downstream analyses. As introns in 16S rRNA genes have not been known until today and because the scientific community is not commonly aware of intervening sequences, PCR amplicons with aberrant lengths are usually not expected. In the course of this thesis, the 16S rRNA genes were amplified along with the ITS region and since ITS regions in bacteria are known to vary from about 140 to 1530 nt (Gürtler & Stanisich, 1996), there was no expectation for a certain amplicon length. This approach eventually led to the detection of introns in 16S rRNA genes and shows very nicely that having less expectations on the outcome of an experiment can open possibilities for finding extremely interesting novelties.

Perspectives

Based on the proposed reclassification, this thesis suggests several FISH probes for the detection of the novel taxa in their natural environments. Once the probes are experimentally optimized, they can be applied on populations of large sulfur bacteria revealing their specific abundance and spatial distribution in different natural habitats. As suggested earlier, the localization of certain taxa of large sulfur bacteria could be combined with a specific characterization of the inhabited site, guiding towards a profound description of growth conditions and other relevant features, which could eventually enable successful culturing of some yet uncultured types.

Thiobacterium bovistum is considered to be taxonomically related to the large sulfur bacteria (Garrity *et al.*, 2005), but phylogenetic data are so far absent for its classification (Grünke *et al.*, 2010). Attempts for 16S rRNA gene sequencing and classification included universal clone libraries and revealed several sequences affiliated to the family *Beggiatoaceae* (Grünke *et al.*, 2010). Using the data presented in this thesis, it could be found that some of these sequences in fact affiliate to "*Candidatus* Thiopilula aggregata", being spherical and larger in cells size and forming smaller but similar aggregates like *Thiobacterium*. So far, no sequence-specific FISH probes were designed to test whether some of the detected sequences originated from the *Thiobacterium* mat, however, their affiliation to the Gammaproteobacteria was revealed using the FISH probe Gam42a (Grünke *et al.*, 2010). Now, it could be tested if a "*Candidatus* Thiopilula"-specific probe or a collection of probes presented in Table 1 hybridizes with the cells in *Thiobacterium* aggregates. If this was the case, single or few cells from an aggregate could be amplified with specific primers deduced from probe sequences to achieve classification.

It will be challenging to test the hypothesis regarding the presence of 16S rRNA introns in other bacteria, because in most cases the discrepancy of bacteria being present but their sequences being absent from clone libraries remains unnoticed. Metagenomics could give insights soon, but also the continous application of regular PCR approaches could reveal further enlarged bacterial 16S rRNA genes. Sequencing of pure cultures of which a 16S rRNA gene sequence could not yet be determined or the performance of single-cell PCR with smaller bacteria after separation with FACS or microfluidic devices could be promosing.

References for Chapters 1 and 5

(1999). Validation list no 71. *Int J Syst Bacteriol* **49**: 1325-1326.

AHMAD A, BARRY JP, NELSON DC (1999). Phylogenetic affinity of a wide, vacuolate, nitrate-accumulating *Beggiatoa* sp. from Monterey Canyon, California, with *Thioploca* spp. *Appl Environ Microbiol* **65**: 270-277.

AHMAD A, KALANETRA KM, NELSON DC (2006). Cultivated *Beggiatoa* spp. define the phylogenetic root of morphologically diverse, noncultured, vacuolate sulfur bacteria. *Can J Microbiol* **52**: 591-598.

AMANN RI, LIN CH, KEY R, MONTGOMERY L, STAHL DA (1992). Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. *Syst Appl Microbiol* **15**: 23-31.

AMANN RI, LUDWIG W, SCHLEIFER KH (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-169.

ANGERT ER, NORTHUP DE, REYSENBACH AL, PEEK AS, GOEBEL BM, PACE NR (1998). Molecular phylogenetic analysis of a bacterial community in Sulphur River, Parker Cave, Kentucky. *Am Mineral* **83**: 1583-1592.

BAAS BECKING LGM (1934). Geobiologie of inleiding tot de milieukunde Diligentia Wetensch. Serie 18/19, van Stockum's Gravenhange.

BEGGIATO FS (1833). Delle Terme Euganee. University of Padova: Padova.

BERSA E (1920). Über das Vorkommen von kohlensaurem Kalk in einer Gruppe von Schwefelbakterien. Sitzungsbericht Akademie der Wissenschaften, mathematisch-naturwissenschaftliche Klasse, I Abteilung: Wien.

BLAND JA, STALEY JT (1978). Observations on the biology of *Thiothrix*. *Arch Microbiol* **117**: 79-87.

BLAND JA, BROCK TD (2005). Genus IV. *Leucothrix* Oersted. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). Bergey´s Manual of Systematic Bacteriology, 2 edn. Springer: New York. pp 162-168.

BROCK J (2011). Impact of sulfide-oxidizing bacteria on the phosphorus cycle in marine sediments. Ph.D. thesis, University of Bremen, Bremen.

BUCHANAN RE, GIBBONS NE (1974). Bergey´s Manual of Determinative Bacteriology The Williams & Wilkins Company: Baltimore.

BURGIN AB, PARODOS K, LANE DJ, PACE NR (1990). The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* **60**: 405-414.

CAO Y, WOODSON SA (1998). Destabilizing effect of an rRNA stem-loop on an attenuator hairpin in the 5 ' exon of the *Tetrahymena* pre-rRNA. *RNA* **4**: 901-914.

CAO Y, WOODSON SA (2000). Refolding of rRNA exons enhances dissociation of the *Tetrahymena* intron. *RNA* **6**: 1248-1256.

CHANDLER DP, FREDRICKSON JK, BROCKMAN FJ (1997). Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. *Mol Ecol* **6**: 475-482.

CHERNOUSOVA E, GRIDNEVA E, GRABOVICH M, DUBININA G, AKIMOV V, ROSSETTI S, KUEVER J (2009). *Thiothrix caldifontis* sp. nov. and *Thiothrix lacustris* sp. nov., gammaproteobacteria isolated from sulfide springs. *Int J Syst Evol Microbiol* **59**: 3128-3135.

CILIA V, LAFAY B, CHRISTEN R (1996). Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol* **13**: 451-461.

COENYE T, VANDAMME P (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* **228**: 45-49.

COHN F (1865). Zwei neue Beggiatoen. *Hedwigia* **4**: 81-84.

COHN F (1875). Untersuchungen über Bacterien. Beiträge zur Biologie der Pflanzen. J. U. Kern´s Verlag (Max Müller), Band I. Heft 3.

COLLINS MD, RODRIGUES U, ASH C, AGUIRRE M, FARROW JAE, MARTINEZMURCIA A, PHILLIPS BA, WILLIAMS AM, WALLBANKS S (1991). Phylogenetic analysis of the genus *Lactobacillus* and related lactic-acid bacteria as determined by reverse transcriptase sequencing of 16S ribosomal RNA. *FEMS Microbiol Lett* **77**: 5-12.

DANDO PR, HUGHES JA, THIERMANN F (1995). Preliminary observations on biological communities at shallow hydrothermal vents in the Aegean Sea. In: Parson LM, Walker CL, Dixon DL (eds). Hydrothermal Vents and Processes: London. pp 303-317.

DARWIN C (1859). On the origin of species by means of natural selection, 1 edn. Murray, J.: London.

DE ALBUQUERQUE JP, KEIM CN, LINS U (2010). Comparative analysis of *Beggiatoa* from hypersaline and marine environments. *Micron* **41**: 507-517.

DE BEER D, SAUTER E, NIEMANN H, KAUL N, FOUCHER JP, WITTE U, SCHLUTER M, BOETIUS A (2006). In situ fluxes and zonation of microbial activity in surface sediments of the Hakon Mosby Mud Volcano. *Limnol Oceanogr* **51**: 1315-1331.

DENNIS PP, ZIESCHE S, MYLVAGANAM S (1998). Transcription analysis of two disparate rRNA operons in the halophilic archaeon *Haloarcula marismortui*. *J Bacteriol* **180**: 4804-4813.

DEVIDE Z (1952). Zwei neue farblose Schwefelbakterien: *Thiogloea ruttneri* n. gen., n. sp. und *Thiogloea ragusina* n. sp. *Hydrobiologia* **14**: 446-455.

EDGCOMB VP, KYSELA DT, TESKE A, GOMEZ AD, SOGIN ML (2002). Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc Natl Acad Sci U S A* **99**: 7658-7662.

EUZÉBY J (2005). Validation of publication of new names and new combinations previously effectively published outside the IJSEM: validation list no. 106. *Int J Syst Evol Microbiol* **55**: 2235-2238.

EVERETT KDE, KAHANE S, BUSH RM, FRIEDMAN MG (1999). An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J Bacteriol* **181**: 4734-4740.

FEIL EJ (2004). Small change: keeping pace with microevolution. *Nat Rev Microbiol* **2**: 483-495.

FENCHEL T (2003). Biogeography for bacteria. *Science* **301**: 925-926.

FOSSING H, GALLARDO VA, JØRGENSEN BB, HUTTEL M, NIELSEN LP, SCHULZ H, CANFIELD DE, FORSTER S, GLUD RN, GUNDERSEN JK, KUVER J, RAMSING NB, TESKE A, THAMDRUP B, ULLOA O (1995). Concentration and transport of nitrate by the mat-forming sulfur bacterium *Thioploca*. *Nature* **374**: 713-715.

FOX GE, MAGRUM LJ, BALCH WE, WOLFE RS, WOESE CR (1977a). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci U S A* **74**: 4537-4541.

FOX GE, PECHMAN KR, WOESE CR (1977b). Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. *Int J Syst Bacteriol* **27**: 44-57.

FOX GE, WISOTZKEY JD, JURTSHUK P (1992). How close is close: 16S ribosomal RNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**: 166-170.

FRANK JA, REICH CI, SHARMA S, WEISBAUM JS, WILSON BA, OLSEN GJ (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* **74**: 2461-2470.

GALLARDO VA (1977). Large benthic microbial communities in sulfide biota under Peru-Chile subsurface countercurrent. *Nature* **268**: 331-332.

GALLARDO VA, ESPINOZA C (2007). New communities of large filamentous sulfur bacteria in the eastern South Pacific. *Int Microbiol* **10**: 97-102.

GARRITY GM, BELL JA, LILBURN T (2005). Family I. *Thiotrichaceae* fam. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). Bergey´s Manual of Systematic Bacteriology, 2 edn. Springer: New York. p 131.

GEVERS D, COHAN FM, LAWRENCE JG, SPRATT BG, COENYE T, FEIL EJ, STACKEBRANDT E, VAN DE PEER Y, VANDAMME P, THOMPSON FL, SWINGS J (2005). Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**: 733-739.

GILLAN DC, SPEKSNIJDER AGCL, ZWART G, DE RIDDER C (1998). Genetic diversity of the biofilm covering *Montacuta ferruginosa* (*Mollusca, Bivalvia*) as evaluated by denaturing gradient gel electrophoresis analysis and cloning of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ Microbiol* **64**: 3464-3472.

GIRNTH A-C, GRÜNKE S, LICHTSCHLAG A, FELDEN J, KNITTEL K, WENZHÖFER F, DE BEER D, BOETIUS A (2011). A novel, mat-forming *Thiomargarita* population associated with a sulfidic fluid flow from a deep-sea mud volcano. *Environ Microbiol* **13**: 495-505.

GOLDHAMMER T, BRÜCHERT V, FERDELMAN TG, ZABEL M (2010). Microbial sequestration of phosphorous in anoxic upwelling sediemnts. *Nature Geoscience* **3**: 557-561.

GOODRICH-BLAIR H, SHUB DA (1996). Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell* **84**: 211-221.

GRÜNKE S, LICHTSCHLAG A, DE BEER D, KUYPERS M, LÖSEKANN-BEHRENS T, RAMETTE A, BOETIUS A (2010). Novel observations of *Thiobacterium*, a sulfur-storing Gammaproteobacterium producing gelatinous mats. *ISME J*: 1-13.

GUEZENNEC J, ORTEGA-MORALES O, RAGUENES G, GEESEY G (1998). Bacterial colonization of artificial substrate in the vicinity of deep-sea hydrothermal vents. *FEMS Microbiol Ecol* **26**: 89-99.

GÜRTLER V, STANISICH VA (1996). New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiol* **142**: 3-16.

GUTZKE V (2007). Morphologische und physiologische Untersuchungen an *Thiomargarita* spp. Diploma thesis, University of Hannover, Hannover.

HAROLD R, STANIER RY (1955). The genera *Leucothrix* and *Thiothrix*. *Bacteriol Rev* **19**: 49-64.

HEAD IM, GRAY ND, HOWARTH R, PICKUP RW, CLARKE KJ, JONES JG (2000). *Achromatium oxaliferum* - understanding the unmistakable. In: Schink B (ed). Advances in microbial ecology. Kluwer Academic/Plenum Publishers: New York. pp 1-40.

HEIJS SK, DAMSTE JSS, FORNEY LJ (2005). Characterization of a deep-sea microbial mat from an active cold seep at the Milano mud volcano in the Eastern Mediterranean Sea. *FEMS Microbiol Ecol* **54**: 47-56.

HINCK S, NEU TR, LAVIK G, MUSSMANN M, DE BEER D, JONKERS HM (2007). Physiological adaptation of a nitrate-storing *Beggiatoa* sp. to diel cycling in a phototrophic hypersaline mat. *Appl Environ Microbiol* **73**: 7013-7022.

HINZE G (1901). Ueber den Bau der Zellen von *Beggiatoa mirabilis* Cohn. *Berichte der Deutschen Botanischen Gesellschaft* **19**: 369-374.

HINZE G (1903). *Thiophysa volutans*, ein neues Schwefelbakterium, vol. Heft 6. Berichte der deutschen botanischen Gesellschaft, Heft 6.

HØGSLUND S, NIELSEN JL, NIELSEN LP (2010). Distribution, ecology and molecular identification of *Thioploca* from Danish brackish water sediments. *FEMS Microbiol Ecol* **73**: 110-120.

HONG SH, BUNGE J, LESLIN C, JEON S, EPSTEIN SS (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J* **3**: 1365-1373.

HOPPE-SEYLER F (1886). Ueber die Gährung der Cellulose mit Bildung von Methan und Kohlensäure. *Zeitschrift für Physiologische Chemie* **10**: 401-440.

HOWARTH R, UNZ RF, SEVIOUR EM, SEVIOUR RJ, BLACKALL LL, PICKUP RW, JONES JG, YAGUCHI J, HEAD IM (1999). Phylogenetic relationships of filamentous sulfur bacteria (*Thiothrix* spp. and Eikelboom type 021N bacteria) isolated from wastewater-treatment plants and description of *Thiothrix eikelboomii* sp. nov., *Thiothrix unzii* sp. nov., *Thiothrix fructosivorans* sp. nov. and *Thiothrix defluvii* sp. nov. *Int J Syst Bacteriol* **49**: 1817-1827.

IMHOFF JF, SULING J, PETRI R (1998). Phylogenetic relationships among the *Chromatiaceae*, their taxonomic reclassification and description of the new genera *Allochromatium*, *Halochromatium*, *Isochromatium*, *Marichromatium*, *Thiococcus*, *Thiohalocapsa* and *Thermochromatium*. *Int J Syst Bacteriol* **48**: 1129-1143.

JACQ E, PRIEUR D, NICHOLS P, WHITE DC, PORTER T, GEESEY GG (1989). Microscopic examination and fatty-acid characterization of filamentous bacteria colonizing substrata around subtidal hydrothermal vents. *Arch Microbiol* **152**: 64-71.

JANNASCH HW, WIRSEN CO (1981). Morphological survey of microbial mats near deep-sea thermal vents. *Appl Environ Microbiol* **41**: 528-538.

JANNASCH HW, NELSON DC, WIRSEN CO (1989). Massive natural occurrence of unusually large bacteria (*Beggiatoa* sp.) at a hydrothermal deep-sea vent site. *Nature* **342**: 834-836.

JØRGENSEN BB (1977). Distribution of colorless sulfur bacteria (*Beggiatoa* spp.) in a coastal marine sediment. *Mar Biol* **41**: 19-28.

JØRGENSEN BB, REVSBECH NP (1983). Colorless sulfur bacteria, *Beggiatoa* spp. and *Thiovulum* spp. in O_2 and H_2S microgradients. *Appl Environ Microbiol* **45**: 1261-1270.

JØRGENSEN BB, GALLARDO VA (1999). *Thioploca* spp.: filamentous sulfur bacteria with nitrate vacuoles. *FEMS Microbiol Ecol* **28**: 301-313.

JØRGENSEN BB, TESKE A, AHMAD A (2005). Genus VII. *Thioploca* Lauterborn. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). Bergey´s Manual of Systematic Bacteriology, 2 edn. Springer: New York. pp 171-178.

JØRGENSEN BB, DUNKER R, GRÜNKE S, ROY H (2010). Filamentous sulfur bacteria, *Beggiatoa* spp., in arctic marine sediments (Svalbard, 79 degrees N). *FEMS Microbiol Ecol* **73**: 500-513.

KALANETRA KM, HUSTON SL, NELSON DC (2004). Novel, attached, sulfur-oxidizing bacteria at shallow hydrothermal vents possess vacuoles not involved in respiratory nitrate accumulation. *Appl Environ Microbiol* **70**: 7487-7496.

KALANETRA KM, JOYE SB, SUNSERI NR, NELSON DC (2005). Novel vacuolate sulfur bacteria from the Gulf of Mexico reproduce by reductive division in three dimensions. *Environ Microbiol* **7**: 1451-1460.

KALANETRA KM, NELSON DC (2010). Vacuolate-attached filaments: highly productive *Ridgeia piscesae* epibionts at the Juan de Fuca hydrothermal vents. *Mar Biol* **157**: 791-800.

KEIL F (1912). Beiträge zur Physiologie der farblosen Schwefelbakterien. *Beiträge zur Biologie der Pflanzen* **11**: 335-372.

KELLY DP, WOOD AP (2000). Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov and *Thermithiobacillus* gen. nov. *Int J Syst Evol Microbiol* **50**: 511-516.

KLAS Z (1937). Über den Formenkreis von *Beggiatoa mirabilis*. *Archives of Microbiology: Aus dem Ozeanographischen Institut Split und den Botanischen Institut Zagreb* **8**: 312-320.

KOJIMA H, FUKUI M (2003). Phylogenetic analysis of *Beggiatoa* spp. from organic rich sediment of Tokyo Bay, Japan. *Water Res* **37**: 3216-3223.

KOJIMA H, TESKE A, FUKUI M (2003). Morphological and phylogenetic characterizations of freshwater *Thioploca* species from Lake Biwa, Japan, and Lake Constance, Germany. *Appl Environ Microbiol* **69**: 390-398.

KOLKWITZ R (1912). Über die Schwefelbakterie *Thioploca ingrica* Wislouch. *Botanische Gesellschaft* **Band 30**: 662-666.

KOLKWITZ R (1955). Über die Schwefelbakterie *Thioploca ingrica* Wislouch. *Berichte der Deutschen Botanischen Gesellschaft* **68**: 374-380.

KOPPE F (1923). Die Schlammflora der ostholsteinischen Seen und des Bodensees. *Arch Hydrobiol* **Band XIV**: 619-675.

LA RIVIERE JWM, SCHMIDT K (2006). Morphologically conspicuous sulfur-oxidizing bacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds). The Prokaryotes. Springer. pp 941-954.

LAMBOWITZ AM, BELFORT M (1993). Introns as mobile genetic elements. *Annu Rev Biochem* **62**: 587-622.

LANE N, MARTIN W (2010) The energetics of genome complexity. *Nature* **467**: 929-934:

LAPAGE SP, SNEATH PHA, LESSEL EF, SKERMAN VBD, SEELIGER HPR, CLARK WA (1992). International Code of Nomenclature of Bacteria: Bacteriological Code. ASM Press: Washington (DC).

LARKIN JM, SHINABARGER DL (1983). Characterization of *Thiothrix nivea*. *Int J Syst Bacteriol* **33**: 841-846.

LARKIN JM, HENK MC (1996). Filamentous sulfide-oxidizing bacteria at hydrocarbon seeps of the Gulf of Mexico. *Microsc Res Tech* **33**: 23-31.

LAUTERBORN R (1907). Eine neue Gattung der Schwefelbakterien (*Thioploca schmidlei* nov. gen. nov. spec.). *Berichte der deutschen botanischen Gesellschaft* **25**: 238-242.

LEADBETTER ER (1974). Family II. *Beggiatoaceae*. In: Buchanan RE, Gibbons NE (eds). Bergey´s Manual of Determinative Bacteriology, 8 edn. The Williams & Wilkins Company: Baltimore. pp 112-116.

LINNAEUS C (1735). Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Theodorum Haak.

LINTON D, CLEWLEY JP, BURNENS A, OWEN RJ, STANLEY J (1994a). An intervening sequence (IVS) in the 16S ribosomal RNA gene of the eubacterium *Helicobacter canis*. *Nucl Acids Res* **22**: 1954-1958.

LINTON D, DEWHIRST FE, CLEWLEY JP, OWEN RJ, BURNENS AP, STANLEY J (1994b). Two types of 16S ribosomal RNA genes are found in *Campylobacter helveticus*: analysis, applications and characterization of the intervening sequence found in some strains. *Microbiol* **140**: 847-855.

LOPEZ-GARCIA P, DUPERRON S, PHILIPPOT P, FORIEL J, SUSINI J, MOREIRA D (2003). Bacterial diversity in hydrothermal sediment and epsilon proteobacterial dominance in experimental microcolonizers at the Mid-Atlantic Ridge. *Environ Microbiol* **5**: 961-976.

LUDWIG W, SCHLEIFER KH (1994). Bacterial phylogeny based on 16S and 23S ribosomal RNA sequence analysis. *FEMS Microbiol Rev* **15**: 155-173.

MAIER S, PREISSNER WC (1979). Occurrence of *Thioploca* in Lake Constance and Lower Saxony, Germany. *Microb Ecol* **5**: 117-119.

MAIER S (1980). Growth of *Thioploca ingrica* in a mixed culture system. *Ohio J Sci* **80**: 30-32.

MAIER S (1984). Description of *Thioploca ingrica* sp. nov., nom. rev. *Int J Syst Bacteriol* **34**: 344-345.

MAIER S, GALLARDO VA (1984a). *Thioploca araucae* sp. nov. and *Thioploca chileae* sp. nov. *Int J Syst Bacteriol* **34**: 414-418.

MAIER S, GALLARDO VA (1984b). Nutritional characteristics of two marine thioplocas determined by autoradiography. *Arch Microbiol* **139**: 218-220.

MARSHALL K (2006). Planktonic versus sessile life of prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds). The Prokaryotes. Springer. pp 3-15.

MARTINEZ-MURCIA AJ, BENLLOCH S, COLLINS MD (1992). Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridizations. *Int J Syst Bacteriol* **42**: 412-421.

MCHATTON SC, BARRY JP, JANNASCH HW, NELSON DC (1996). High nitrate concentrations in vacuolate, autotrophic marine *Beggiatoa* spp. *Appl Environ Microbiol* **62**: 954-958.

MIGULA W (1894). Ueber ein neues System der Bakterien. *Arbeit aus dem bakteriologischen Institut der technischen Hochschule zu Karlsruhe* **1**: 235-238.

MILLS DA, MCKAY LL, DUNNY GM (1996). Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J Bacteriol* **178**: 3531-3538.

MULLIS K, FALOONA F, SCHARF S, SAIKI R, HORN G, ERLICH H (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Sym* **51**: 263-273.

MURRAY RGE, SCHLEIFER K-H (1994). Taxonomic notes - a proposal for recording the properties of putative taxa of prokaryotes. *Int J Syst Bacteriol* **44**: 174-176.

MURRAY RGE, STACKEBRANDT E (1995). Taxonomic note - implementation of the provisional status *Candidatus* for incompletely described prokaryotes. *Int J Syst Bacteriol* **45**: 186-187.

MUSSMANN M, SCHULZ HN, STROTMANN B, KJAER T, NIELSEN LP, ROSSELLÓ-MORA RA, AMANN RI, JØRGENSEN BB (2003). Phylogeny and distribution of nitrate-storing *Beggiatoa* spp. in coastal marine sediments. *Environ Microbiol* **5**: 523-533.

MUSSMANN M, HU FZ, RICHTER M, DE BEER D, PREISLER A, JORGENSEN BB, HUNTEMANN M, GLOCKNER FO, AMANN R, KOOPMAN WJH, LASKEN RS, JANTO B, HOGG J, STOODLEY P, BOISSY R, EHRLICH GD (2007). Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol* **5**: 1923-1937.

NADSON GA (1913). Ueber Schwefelmikroorganismen des Hapsaler Meerbusens. *Bulletin du Jardin Impériale Botanique de St-Pétersbourg* **13**: 106-112.

NELSON DC, CASTENHOLZ RW (1981). Use of reduced sulfur compounds by *Beggiatoa* sp. *J Bacteriol* **147**: 140-154.

NELSON DC, WATERBURY JB, JANNASCH HW (1982). Nitrogen-fixation and nitrate utilization by marine and freshwater *Beggiatoa*. *Arch Microbiol* **133**: 172-177.

NELSON DC, WIRSEN CO, JANNASCH HW (1989). Characterization of large, autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas Basin. *Appl Environ Microbiol* **55**: 2909-2917.

NELSON WC, JANNASCH HW (1983). Chemoautotrophic growth of a marine *Beggiatoa* in sulfide-gradient cultures. *Arch Microbiol* **136**: 262-269.

NIELSEN H, JOHANSEN D (2009). Group I introns. Moving in new directions. *RNA Biol* **6**: 375-383.

NIKOLCHEVA T, WOODSON SA (1997). Association of a group I intron with its splice junction in 50S ribosomes: implications for intron toxicity. *RNA* **3**: 1016-1027.

ODINTSOVA EV, WOOD AP, KELLY DP (1993). Chemolithoautotrophic growth of *Thiothrix ramosa*. *Arch Microbiol* **160**: 152-157.

OERSTED AS (1844). De Regionibus marinis. Elementa Topographiae Historiconaturalis Freti Oeresund. Ph D thesis, University of Copenhagen, Copenhagen.

OLSEN GJ, LARSEN N, WOESE CR (1991). The ribosomal RNA database project. *Nucl Acids Res* **19**: 2017-2021.

OLU-LE ROY K, VON COSEL R, HOURDEZ S, CARNEY SL, JOLLIVET D (2007). Ampi-Atlantic cold seep *Bathymodiolus* species complexes across the equatorial belt. *Deep-Sea Res* **54**: 1890-1911.

OTTE S, KUENEN G, NIELSEN LP, PAERL HW, ZOPFI J, SCHULZ HN, TESKE A, STROTMANN B, GALLARDO VA, JØRGENSEN BB (1999). Nitrogen, carbon, and sulfur metabolism in natural *Thioploca* samples. *Appl Environ Microbio* **65**: 3148-3157.

PACE NR, OLSEN GJ, WOESE CR (1986). Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**: 325-326.

PACE NR (2009). Mapping the Tree of Life: Progress and Prospects. *Microbiol Mol Biol Rev* **73**: 565-576.

PEPLIES J, KOTTMANN R, LUDWIG W, GLÖCKNER FO (2008). A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. *Syst Appl Microbiol* **31**: 251-257.

POINDEXTER JS (2006). Dimorphic prostecate bacteria: the genera *Caulobacter*, *Asticcacaulis*, *Hyphomicrobium*, *Pedomicrobium*, *Hyphomonas* and *Thiodendron*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds). The Prokaryotes, 3 edn. Springer. pp 72-90.

PRINGSHEIM EG (1957). Observations on *Leucothrix mucor* and *Leucothrix cohaerens* nov. sp.: with a survey of colorless filamentous organisms. *Bacteriol Rev* **21**: 69-76.

RABENHORST L (1865). Flora europaca algarum aquae dulcis submannae. Section II. Kummer, Leipzig: 94.

RAGHAVAN R, MILLER SR, HICKS LD, MINNICK MF (2007). The unusual 23S rRNA gene of *Coxiella burnetii*: Two self-splicing group I introns flank a 34-base-pair exon, and one element lacks the canonical omega G. *J Bacteriol* **189**: 6572-6579.

RAINEY FA, WARD-RAINEY NL, JANSSEN PH, HIPPE H, STACKEBRANDT E (1996). *Clostridium paradoxum* DSM 7308(T) contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiol-UK* **142**: 2087-2095.

RAJ HD (1977). Leucothrix. *Crit Rev Microbiol* **5**: 271-304.

RAJANIEMI P, HROUZEK P, KASTOVSKA K, WILLAME R, RANTALA A, HOFFMANN L, KOMAREK J, SIVONEN K (2005). Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (*Nostocales*, Cyanobacteria). *Int J Syst Evol Microbiol* **55**: 11-26.

RALPH D, MCCLELLAND M (1993). Intervening sequence with conserved open reading frame in eubacterial 23S ribosomal RNA genes. *Proc Natl Acad Sci U S A* **90**: 6864-6868.

RANGAN P, MASQUIDA B, WESTHOF E, WOODSON SA (2004). Architecture and folding mechanism of the *Azoarcus* group I pre-tRNA. *J Mol Biol* **339**: 41-51.

ROSSELLO-MORA R, LUCIO M, PENA A, BRITO-ECHEVERRIA J, LOPEZ-LOPEZ A, VALENS-VADELL M, FROMMBERGER M, ANTON J, SCHMITT-KOPPLIN P (2008). Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *Isme Journal* **2**: 242-253.

SAIKI RK, SCHARF S, FALOONA F, MULLIS KB, HORN GT, ERLICH HA, ARNHEIM N (1985). Enzymatic Amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anemia. *Science* **230**: 1350-1354.

SAIKI RK, GELFAND DH, STOFFEL S, SCHARF SJ, HIGUCHI R, HORN GT, MULLIS KB, ERLICH HA (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

SAYAMA M, RISGAARD-PETERSEN N, NIELSEN LP, FOSSING H, CHRISTENSEN PB (2005). Impact of bacterial NO_3^- transport on sediment biogeochemistry. *Appl Environ Microbiol* **71**: 7575-7577.

SCHEWIAKOFF W (1892). Ueber einen neuen bacterienähnlichen Organismus des Süsswassers. habilitation thesis, University Heidelberg, Heidelberg.

SCHULZ HN, JØRGENSEN BB, FOSSING HA, RAMSING NB (1996). Community structure of filamentous, sheath-building sulfur bacteria, *Thioploca* spp., off the coast of Chile. *Appl Environ Microbiol* **62**: 1855-1862.

SCHULZ HN, BRINKHOFF T, FERDELMAN TG, MARINE MH, TESKE A, JØRGENSEN BB (1999). Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* **284**: 493-495.

SCHULZ HN, JØRGENSEN BB (2001). Big bacteria. *Annu Rev Microbiol* **55**: 105-137.

SCHULZ HN, DE BEER D (2002). Uptake rates of oxygen and sulfide measured with individual *Thiomargarita namibiensis* cells by using microelectrodes. *Appl Environ Microbiol* **68**: 5746-5749.

SCHULZ HN, JØRGENSEN BB (2005). Genus VI. *Thiomargarita*. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). Bergey´s Manual of Systematic Bacteriology. Springer: New York. pp 169-171.

SCHULZ HN, SCHULZ HD (2005). Large sulfur bacteria and the formation of phosphorite. *Science* **307**: 416-418.

SCHULZ HN (2006). The genus *Thiomargarita*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds). The Prokaryotes, 3 edn. Springer. pp 1156-1163.

SEKAR R, MILLS DK, REMILY ER, VOSS JD, RICHARDSON LL (2006). Microbial communities in the surface mucopolysaccharide layer and the black band microbial mat of black band-diseased *Siderastrea siderea*. *Appl Environ Microbiol* **72**: 5963-5973.

SKERMAN VBD, MCGOWAN V, SNEATH PHA (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* **30**: 225-420.

SKERMAN VBD, MCGOWAN V, SNEATH PHA (1989). Approved lists of bacterial names (amended edition). American Society for Microbiology: Washington, DC. pp 1-298.

SOUTAR A, CRILL PA (1977). Sedimentation and climatic patterns in the Santa Barbara Basin during 19th and 20th centuries. *Geol Soc Am Bull* **88**: 1161-1172.

SPRINGER N, LUDWIG W, AMANN R, SCHMIDT HJ, GORTZ HD, SCHLEIFER KH (1993). Occurrence of fragmented 16S ribosomal RNA in an obligate bacterial endosymbiont of *Paramecium caudatum*. *Proc Natl Acad Sci U S A* **90**: 9892-9895.

STACKEBRANDT E, GOEBEL BM (1994). A place for DNA-DNA reassociation and 16S ribosomal RNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**: 846-849.

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. *Int J Syst Evol Microbiol* **52**: 1043-1047.

STACKEBRANDT E, EBERS J (2006). Taxonomic parameters revisited: tarnished gold standards. *Micobiology Today* **33**: 152-155.

STALEY JT, BRYANT MP, PFENNIG N, HOLT JG (1989). Bergey´s Manual of Systematic Bacteriology. Williams & Wilkins Company: Baltimore.

STEIN JL (1984). Subtidal gastropods consume sulfur-oxidizing bacteria - evidence from coastal hydrothermal vents. *Science* **223**: 696-698.

STEVENS H, ULLOA O (2008). Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol* **10**: 1244-1259.

STROHL WR, LARKIN JM (1978). Enumeration, isolation, and characterization of *Beggiatoa* from freshwater sediments. *Appl Environ Microbiol* **36**: 755-770.

STROHL WR, CANNON GC, SHIVELY JM, GUDE H, HOOK LA, LANE CM, LARKIN JM (1981). Heterotrophic carbon metabolism by *Beggiatoa alba*. *J Bacteriol* **148**: 572-583.

STROHL WR (1989). Family I. *Beggiatoaceae*. In: Staley JT, Bryant MP, Pfennig N, Holt JG (eds). Bergey´s Manual of Systematic Bacteriology, 1 edn. Williams & Wilkins Company: Baltimore. pp 2089-2106.

STROHL WR (2005). Genus III. Beggiatoa. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). Bergey´s Manual of Systematic Bacteriology
Springer: New York. pp 148-161.

SUZUKI MT, GIOVANNONI SJ (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625-630.

TEMARA A, DERIDDER C, KUENEN JG, ROBERTSON LA (1993). Sulfide-oxidizing bacteria in the burrowing echinoid, *Echinocardium cordatum* (Echinodermata). *Mar Biol* **115**: 179-185.

TESKE A, RAMSING NB, KÜVER J, FOSSING H (1995). Phylogeny of *Thioploca* and related filamentous sulfide-oxidizing bacteria. *Syst Appl Microbiol* **18**: 517-526.

TESKE A, NELSON DC (2006). The genera *Beggiatoa* and *Thioploca*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds). The Prokaryotes, 3 edn. Springer. pp 784-810.

THORNTON HG (1953). Sergei Nicholaevitch Winogradsky. *Obituary Notices of Fellows of the Royal Society* **8**: 635-644.

TINDALL BJ, KAMPFER P, EUZEBY JP, OREN A (2006). Valid publication of names of prokaryotes according to the rules of nomenclature: past history and current practice. *Int J Syst Evol Microbiol* **56**: 2715-2720.

TINDALL BJ, ROSSELLO-MORA R, BUSSE HJ, LUDWIG W, KAMPFER P (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* **60**: 249-266.

TOOR N, KEATING KS, PYLE AM (2009). Structural insights into RNA splicing. *Curr Opin Struct Biol* **19**: 260-266.

TREVISAN V (1842). Prospetto della Flora Euganea. Coi Tipi del Seminario: Padua.

UNZ RF, HEAD IM (2005). Genus I. *Thiothrix* In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). Bergey´s Manual of Systematic Bacteriology, 2 edn. Springer: Baltimore. pp 131-141.

V. WINTZINGERODE F, GÖBEL UB, STACKEBRANDT E (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213-229.

VAN NIEL CB (1948). Family A. *Achromatiaceae* Massart. In: Breed RS, Murray EGD, Hitchens AP (eds). Bergey´s Manual of Determinative Bacteriology, 6 edn. The Williams and Wilkins Company: Baltimore. pp 997-999.

VAUCHER JP (1803). Histoire des conferves d´eau douce, contenant leurs différents modes de reproduction, et la description de leurs principales espèces: Paschoud, Geneva.

WEST GS, GRIFFITHS BM (1909). *Hillhousia mirabilis*, a giant sulphur bacterium. *P R Soc Lond B-Conta* **81**: 398-405.

WHITAKER RJ, GROGAN DW, TAYLOR JW (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**: 976-978.

WILLIAMS LA, REIMERS C (1983). Role of bacterial mats in oxygen-deficient marine basins and coastal upwelling regimes: preliminary report. *Geology* **11**: 267-269.

WILLIAMS TM, UNZ RF (1985). Filamentous sulfur bacteria of activated sludge - characterization of *Thiothrix*, *Beggiatoa*, and Eikelboom Type-021N strains. *Appl Environ Microbiol* **49**: 887-898.

WILSON KH, BLITCHINGTON RB, GREENE RC (1990). Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. *J Clin Microbiol* **28**: 1942-1946.

WINOGRADSKY S (1890). Sur le organismes de la nitrification. Comptes rendus de séances l'Academie des Sciences. 1013-1016: Band 110.

WINOGRADSKY S (1892). Recherches sur les organismes de la nitrification. *Annales de l'Institut Pasteur* **4**: 213-231.

WINOGRADSKY SN (1887). Ueber Schwefelbacterien. *Bot Zeitung* **45**: 489-610.

WINOGRADSKY SN (1888). Beiträge zur Morphologie und Physiologie der Bacterien, Heft 1. *Zur Morphologie und Physiologie der Schwefelbacterien*: Leipzig, 1-120.

WIRSEN CO, JANNASCH HW, MOLYNEAUX SJ (1993). Chemosynthetic microbial activity at mid-Atlantic Ridge hydrothermal vent sites. *J Geophys Res-Sol Ea* **98**: 9693-9703.

WISLOUCH SM (1912). *Thioploca ingrica* nov. sp. *Berichte der deutschen botanischen Gesellschaft* **30**: 470-474.

WOESE C, SOGIN M, STAHL D, LEWIS BJ, BONEN L (1976). Comparison of 16s ribosomal RNAs from mesophilic and thermophilic bacilli - some modifications in Sanger method for RNA sequencing. *J Mol Evol* **7**: 197-213.

WOESE CR, FOX GE, ZABLEN L, UCHIDA T, BONEN L, PECHMAN K, LEWIS BJ, STAHL D (1975). Conservation of primary structure in 16S ribosomal RNA. *Nature* **254**: 83-86.

WOESE CR, FOX GE (1977). Phylogenetic structure of prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**: 5088-5090.

WOESE CR, MAGRUM LJ, GUPTA R, SIEGEL RB, STAHL DA, KOP J, CRAWFORD N, BROSIUS J, GUTELL R, HOGAN JJ, NOLLER HF (1980). Secondary structure model for bacterial 16S ribosomal RNA - phylogenetic, enzymatic and chemical evidence. *Nucl Acids Res* **8**: 2275-2293.

WOESE CR (1987). Bacterial Evolution. *Microbiol Rev* **51**: 221-271.

WOESE CR, KANDLER O, WHEELIS ML (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**: 4576-4579.

XIAO SH, YUAN XL, KNOLL AH (2000). Eumetazoan fossils in terminal proterozoic phosphorites? *Proc Natl Acad Sci U S A* **97**: 13684-13689.

YARZA P, RICHTER M, PEPLIES J, EUZEBY J, AMANN R, SCHLEIFER KH, LUDWIG W, GLOCKNER FO, ROSSELLO-MORA R (2008). The all-species living tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* **31**: 241-250.

ZABLEN L, BONEN L, MEYER R, WOESE CR (1975a). Phylogenetic status of *Pasteurella pestis*. *J Mol Evol* **4**: 347-358.

ZABLEN LB, KISSIL MS, WOESE CR, BUETOW DE (1975b). Phylogenetic origin of chloroplast and prokaryotic nature of its ribosomal RNA. *Proc Natl Acad Sci U S A* **72**: 2418-2422.

ZUCKERLAND E, PAULING L (1965). Molecules as documents of evolutionary history. *J Theoret Biol*.

Acknowledgements

I very much thank my reviewers, Dr. Heide Schulz-Vogt and Prof. Rudolf Amann for evaluating my PhD thesis and guiding me through the years at the MPI. My kind regards also go to Dr. Bernhard Fuchs, Prof. Annette Becker, Anne Schwedt and Johanna Radziejewski for being part of my defense committee.

My special thanks goes to Heide for being my supervisor through the past years, for introducing me to the marvelous and evermore exciting *Thiomargarita* and for always having an open ear for any professional or private concern.

A big thank you also goes to the ecophysiology group, you are great colleagues in every aspect, who have turned into friends with time and who are always there for each other!

Thank you also to Lubos for the intense discussions and challenging questions that always motivated me to achieve the best results. Furthermore, I appreciate very much the help of all people from the microbiology and molecular ecology department for introducing me into various techniques, and thank you to the 'ARB people', including Prof. Ludwig and Elmar Pruesse, for solving my sometimes big, sometimes small hadicaps with that program.

To my feeling, the past one year was especially intense with regards to ups and downs, friendship and loyalty, frustration and success. In this time I especially enjoyed being together with my dear office mates, who supported me just as much as I (hopefully) supported them. Thank you girls!

Much encouragement and lots of joy I also experienced with the girls of our regular 'Mädelsabend'. Thanks girls for the great time, the cooking, chatting, laughing, cats kuscheln, Lebkuchen baking and all the other crazy stuff we did...

Mein letztes aber auch liebstes Dankeschön geht an meine Familie. Mama, Papa, Isabell, Oma, Opa und Diana, ihr habt meine Uni- und Doktorandenzeit stets mit Interesse verfolgt und mich in jeglicher Hinsicht liebevoll unterstützt. Jawad, du hast immer das Beste aus mir heraus geholt und hast mir gezeigt was ich alles kann. Du glaubst immer an mich und hast mich durch deine Liebe bis hier her gebracht. Danke an euch!

Curriculum vitae

First name	Verena
Surname, maiden name	Salman, Gutzke
Date of birth	21 July 1983
Place of birth	Hannover, Germany
Nationality	German

01/2008 – 06/2011	Ph.D. candidate at the University of Bremen, Germany
04/2007 – 12/2007	Diploma thesis "Morphological and Physiological Studies of *Thiomargarita* spp." at the Leibniz University of Hannover, Germany
10/2003 – 12/2007	Study of Biological Sciences at the Leibniz University of Hannover, Germany
06/2003	Diploma at German secondary school qualifying for university admission, Hannover, Germany
07/2001	Diploma at Craig High School, Janesville, Wisconsin, USA

Participation at conferences

07/2010 Gordon Research Conference (GRC), Tilton, New Hampshire.
Poster: "Reclassification of the giant sulfur bacteria and the discovery of large insertions in their 16S rRNA genes"

03/2009 Verein Allgemeiner und Angewandter Mikrobiologie (VAAM), Bochum, Germany.
Poster: "Unexpected morphological diversity and 16S rDNA abnormality in the giant sulfur bacterium Thiomargarita"

08/2008 International Society of Microbial Ecology (ISME), Cairns, Australia.
Poster: "Morphological studies of *Thiomargarita* spp."

Publication List

Salman V, Amann R, Girnth A-C, Polerecky L, Bailey J, Høgslund S, Jessen G, Pantoja S, Schulz-Vogt HN. A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. Syst Appl Microbiol, in press
doi: 10.1016/j.syapm.2011.02.001

Bailey JV, Salman V, Rouse G, Schulz-Vogt HN, Levin L, Orphan V. Dimorphic development in methane seep dwelling ecotypes of the largest known bacteria. Submitted for publication.

Hinck S, Mussmann M, Salman V, Neu TR, Lenk S, de Beer D, Jonkers HM. Vacuolated *Beggiatoa*-like filaments from different hypersaline environments form a novel genus. Submitted for publication.

Manuscripts in preparation

Salman V, Amann R, Schulz-Vogt HN. Large and frequent self-splicing introns in the 16S rRNA genes of large sulfur bacteria.

Brock J, Rhiel E, Beutler M, Salman V, Schulz-Vogt HN. Unusual polyphosphate inclusions observed in a marine *Beggiatoa* strain.

Grünke S, Lichtschlag A, de Beer D, Felden J, Salman V, Ramette A, Schulz-Vogt HN, Boetius A. Diversity of deep-sea thiotrophic bacterial mats associated with cold seeps of the Norwegian continental margin.

Anlage zur Dissertation

Name:_____ Ort, Datum:_____

Anschrift:_____

Erklärung

gem. § 6 (5) Nr. 1 - 3 PromO

Ich erkläre, dass ich

1. die Arbeit ohne unerlaubte fremde Hilfe angefertigt habe,
2. keine anderen, als die von mir angegebenen Quellen und Hilfsmittel benutzt habe und
3. die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

_____

(Unterschrift)