Molecular basis of inter- and intraspecific multicellularity in prokaryotes

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

vorgelegt von

Roland Wenter

am 04. Februar 2010

1. Gutachter: Prof. Dr. Jörg Overmann

2. Gutachter: Prof. Dr. Dirk Schüler

Tag der mündlichen Prüfung: 12. Mai 2010

„In der Natur ist die Bedeutung des unendlich Kleinen unendlich groß"

Louis Pasteur, französischer Mikrobiologe (1822-1895)

Meinen Eltern gewidmet

Publications originating from this thesis

Chapter 3:

Vogl, K., **Wenter, R.**, Dressen, M., Schlickenrieder, M., Plöscher, M., Eichacker, L.A., Overmann, J. (2008) Identification and analysis of four candidate symbiosis genes from *"Chlorochromatium aggregatum"*, a highly developed bacterial symbiosis. *Environ Microbiol* 10: 2842-2856

Chapter 4:

Wenter, R., Hütz, K.A., Dibbern, D., Plöscher, M., Li, T., Reisinger, V., Plöscher, M., Eichacker, L.A., Eddie, B., Hanson, T., Bryant, D.A., Overmann, J. (2009) Expression-based identification of genetic determinants of the bacterial symbiosis *"Chlorochromatium aggregatum"*. *Environ Microbiol* (published online ahead of print on April 01, 2010; doi:10.1111/j.1462-2920.2010.02206.x)

Chapter 5:

Wenter, R., Wanner, G., Schüler, D., Overmann, J. (2009) Ultrastructure, tactic behavior and potential for sulfate reduction of a novel multicellular magnetotactic prokaryote from North Sea sediments. *Environ Microbiol* 11: 1493-1505

Contributions of Roland Wenter to the publications listed in this thesis

Chapter 3:

Roland Wenter performed the inital Northern blot analyses of ORFs Cag_1919 and 1920 as well as the long range RT-PCR of Cag_1919. He also prepared the protein bands for mass spectrometry and conducted the bioinformatics data analysis. The cultures of *Chlorobium chlorochromatii* CaD and "*Chlorochromatium aggregatum*" were maintained together with Kajetan Vogl. Roland Wenter wrote the respective experimental procedure and results sections and created Figure 1B.

Chapter 4:

Roland Wenter performed the *in silico* subtractive hybridisation analysis, prokaryotic cDNA suppression subtractive hybridisation, reverse transcription, quantitative real-time PCR and phylogenetic analysis of ORF Cag_1285. He conducted the cross-linking studies of consortia membrane proteins as well as the isolation of chlorosomes and analysis of chlorosomal proteins together with Katharina Hütz. Two-dimensional difference gel electrophoresis of the cytosolic proteome was performed together with Veronika Reisinger and Dörte Dibbern. Roland Wenter prepared the proteins for mass spectrometry as well as the cDNA for Illumina whole transcriptome sequencing and subsequently conducted the bioinformatics data analyses. *Chlorobium chlorochromatii* CaD and "*Chlorochromatium aggregatum*" were cultivated together with Katharina Hütz and Dörte Dibbern. Roland Wenter and Jörg Overmann created all figures and tables and wrote the publication.

Chapter 5:

Field sampling of the magnetic multicellular prokaryotes (MMPs) was done together with Prof. Dr. Dirk Schüler. Roland Wenter conducted the magnetotactic enrichment of the MMPs, phase contrast microscopy and the chemotaxis assays. Furthermore he performed fluorescence *in situ* hybridisation and phylogenetic analyses of the MMPs 16S rRNA genes as well as the DsrAB and AprA gene products. Scanning electron microscopy was carried out together with Prof. Dr. Gerhard Wanner. Roland Wenter and Jörg Overmann created all figures and wrote the publication.

I herby confirm the above statements

_____   _____

         Roland Wenter   Prof. Dr. Jörg Overmann

Contents

Chapter 1

Summary

Although the molecular basis of inter- and intraspecific multicellularity in prokaryotes has important implications for the understanding of bacterial interactions with human or plant hosts, functional studies are limited because bacterial associations are difficult to maintain in laboratory cultures. Therefore the molecular determinants underlying the interactions between the partner bacteria leading to prokaryotic multicellularity in most cases have remained unknown. The phototrophic consortium "*Chlorochromatium aggregatum*" represents the first model system to identify the molecular basis as well as physiological properties of the symbiotic interaction between non-related prokaryotes since it recently could be cultivated (Pfannes *et al.*, 2007) and its epibiont *Chlorobium chlorochromatii* was isolated in axenic culture (Vogl *et al.*, 2006).

Four putative symbiosis genes from the epibiont *Chl. chlorochromatii* were recovered by suppression subtractive hybridisation of genomic DNA and bioinformatics approaches. These genes do not occur in the free-living relatives of the epibiont. Two putative hemagglutinin-like gene products were unusually large, whereas the other two encoded a putative hemolysin and a putative RTX toxin-like protein predicted to form a C-terminal Ca^{2+}-binding beta roll structure. A series of Northern blot analyses was conducted to determine the transcripts lengths of the symbiosis genes. No signals could be detected with the specific probes, indicating a low abundance of the transcripts. A subsequent RT-PCR approach revealed their constitutive transcription. Unlike other RTX toxins, a gene product of the RTX-like protein could not be detected by $^{45}Ca^{2+}$-autoradiography, indicating a low abundance of the corresponding protein in the cells. The RTX-type C-terminus exhibited a significant similarity to RTX modules of various proteins from proteobacterial pathogens providing the first indication that putative symbiosis genes have been acquired laterally via horizontal gene transfer and subsequently were employed in multicellular interactions between different species of prokaryotes.

Moreover, genes and proteins potentially involved in this symbiotic interaction were identified on the genomic, transcriptomic and proteomic level. Only a limited number of ORFs were found to be unique to the epibiont genome as compared to all available genomes of free-living relatives. 2-D differential gel electrophoresis (2-D DIGE) of the cytoplasmic proteomes recovered proteins that were detected exclusively in consortia but not in pure epibiont cultures. The most intense of these spots could be attributed to the epibiont *Chl. chlorochromatii* using mass spectrometry. Analyses of the membrane proteins of consortia, of consortia treated with cross-linkers, and of pure cultures

indicated that a branched chain amino acid ABC-transporter binding protein is only expressed in the symbiotic state of the epibiont and possibly is located at the cell contact site to the central bacterium. Furthermore, analyses of chlorosomal proteins revealed that an uncharacterised, small epibiont protein is only expressed during symbiosis. This protein may be involved in the intracellular sorting of chlorosomes. The application of a novel prokaryotic cDNA suppression subtractive hybridisation technique led to identification of 14 differentially regulated genes. The subsequent transcriptomic comparison of symbiotic and free-living epibionts by Illumina whole transcriptome sequencing indicated that 328 genes were differentially transcribed. The three approaches were mostly complementary and thereby yielded a first inventory of 352 genes which are likely to be involved in the bacterial interaction in "*C. aggregatum*". Most notably, the majority of the regulated genes encoded components of central metabolic pathways whereas only very few (7.5%) of the unique symbiosis genes turned out to be regulated under the experimental conditions tested. This pronounced regulation of central metabolic pathways may serve to fine-tune the symbiotic interaction in "*C. aggregatum*" in response to environmental conditions.

To provide the basis for the future development of a novel model system for intraspecific multicellularity of bacteria that complements the studies on the interspecific interaction in the phototrophic consortium "*C. aggregatum*", a novel type of multicellular magnetotactic prokaryote (MMP) was analysed using methods that already proved of value in establishing "*C. aggregatum*" as a cultivable model system. As yet, the uncultured MMPs representing highly structured, intraspecific bacterial aggregates have only been observed at several sampling sites in North and South America. In the present study, a novel type of MMP was discovered for the first time in Europe. Furthermore, the open intertidal sand flats of the North Sea investigated represent a novel type of habitat of these multicellular bacteria, which so far have only been found in salt marshes or coastal lagoons indicating that MMPs occur in different types of habitats over a broader geographical range than previously assumed. Ultrastructural analysis of the MMPs from North Sea revealed that the MMP harboured bullet-shaped magnetosome crystals composed of an iron sulfide mineral and therefore are unique with respect to their specific combination of morphology and chemical composition. Within each aggregate, the magnetosome chains of individual cells were aligned in a highly ordered array of several parallel chains oriented in the same direction. This particular morphology and arrangement of magnetosomes has so far not been visualised in other MMPs and provides an explanation for the observed magnetic optimisation of these multicellular bacterial aggregates achieved by intraspecific microbial communication. Besides the distinct cytological features, the MMP from the North Sea represents a novel phylotype related to the dissimilatory sulfate reducing *Desulfobacteraceae* (*Deltaproteobacteria*) and, based on its phylogenetic distance to known sequence types, a new genus and species, for which the name

'*Candidatus* Magnetomorum litorale' is proposed. Fluorescence *in situ* hybridisation with a specific oligonucleotide probe revealed that all MMPs in the tidal flat sediments studied belonged to the novel phylotype. Within each MMP, all bacterial cells showed a hybridisation signal, indicating that the aggregates are composed of cells of the same phylotype. Genes for dissimilatory sulfite reductase (*dsr*AB) and dissimilatory adenosine-5´-phosphate reductase (*apr*A) could be detected in purified MMP samples, suggesting that MMPs are capable of sulfate reduction. Additionally, chemotaxis assays yielded a strong response of MMPs toward acetate and propionate, which are typical substrates for sulfate-reducing bacteria and thus represent the first known potential growth substrates for a successful cultivation, a prereqisite to establishing the MMP as a new model system to study intraspecific multicellularity in prokaryotes.

Chapter 2

Introduction

Inter- and intraspecific interactions and symbioses between prokaryotes

In their natural environment prokaryotes mostly live in close spatial proximity to other prokaryotes resulting in a great variety of bacterial interactions. In general, the interaction of living organisms is understood as the impact of the interaction partners on each other including their mutual relations and can be categorised based on their effect or mechanism (Abrams, 1987). Classes of such biological interactions include predation, parasitism, competition as well as symbiosis (Krebs, 2001).

Symbiosis designates the living together of different species of organisms predominantly for the benefit of all partners (Moran, 2006). Symbiotic relationships represent an important form among prokaryotic interactions. During the course of evolution symbiosis became a typical way of prokaryotic life exhibiting a striking diversity. Most bacterial symbioses discovered so far involve eukaryotes and thus were in the focus of investigations (Overmann, 2006). In contrast, only a few symbioses between prokaryotes are currently known. Compared to the numerous well documented inter- and intraspecific prokaryotic interactions in unstructured populations of autoinducer-producing bacteria (e.g. Bassler, 2002) or myxobacteria (e.g. Reichenbach and Dworkin, 1992) much less is known about the various forms of symbioses between different bacterial species which have been studied mainly with respect to syntrophic cooperations (Schink, 2002).

Syntrophy exists at least in some of the interspecific interactions between prokaryotic cells, in which all bacterial partners entirely depend on each other to catalyze metabolic processes they can not perform alone. The syntrophic cooperation of prokaryotes in methanogenic degradation is understood in sufficient detail to permit a functional understanding of this specific interaction. It comprises the fermentation of amino acids and sugars by the acetogenic bacterial partner of the association and the methanogenic degradation of electron rich substrates like short chain fatty acids, primary alcohols and aromatics to acetate, carbonate and molecular hydrogen by the methanogenic archaea (Schink, 2002). These fermentative reactions are endergonic and thus only proceed under very low concentration of the reaction products hydrogen and acetate which are consumed in parallel by the associated methanogenic archaea. Metabolites like acetate, formate, sulfur or molecular hydrogen are transferred as electron carrier between oxidative and reductive metabolic processes from one bacterial partner to the other. A stable syntrophic association was also observed for green sulfur bacteria and sulfate-reducing bacteria in which sulfur compounds act as electron

carrier in a sulfur cycle (Pfennig, 1980). Sulfide is photooxidised by the green sulfur bacteria to sulfur or sulfate, which is subsequently reduced back by sulfur- and sulfate-reducing bacteria generating sulfide. The performance of the metabolite transfer depends on the diffusion distance facilitated by spatial vicinity of the single partner cells (Schink, 1991), which also intensifies the syntrophic cooperation due to its energy limitation (Schink, 2002). Consequently a close, permanent association of bacteria with different, but complementary metabolism is of selective advantage. Thus bacteria interacting by the exchange of soluble compounds are subject to a strong selective pressure in their natural habitats to develop mechanisms to maintain a close proximity to each other because the diffusion decreases sharply with distance. This effect has lead to the frequent formation of bacterial microcolonies, biofilms and aggregates. However, an increasing number of cases are known in which different bacterial species are capable of forming highly structured associations (Overmann, 2002).

These purely prokaryotic, morphologically conspicuous multicellular aggregates are called "consortia" (L. v. *consere*, to assemble) (Buder, 1914). Consortia typically consist of two or more dissimilar types of prokaryotes, which are capable to maintain a permanent cell-cell contact (Trüper and Pfennig, 1971; Schink, 1991) and seem to have an obligatory mutual interdependence (Overmann and Schubert, 2002). With regard to their highly ordered structure, prokaryotic consortia represent the most developed type of all bacterial symbioses known today, but only little is known about their occurrence, environmental significance, specificity as well as their molecular and physiological basis of interaction. Consortia are not only relevant for maintaining biogeochemical cycles in different environments by catalysing key metabolic processes (e.g. Mobarry *et al.*, 1996; Boetius *et al.*, 2000), but also are of medical relevance due to their involvement in dental plaque (Kolenbrander and London, 1993; Whittaker *et al.*, 1996) and technological significance for example in anaerobic sludge bed reactors for wastewater treatment (Harmsen *et al.*, 1996; de Bok *et al.*, 2004). Similarly, the molecular mechanisms and the evolution of reciprocal adaption in bacteria-bacteria interactions have implications for the understanding of bacterial interactions with human or plant hosts and for the endosymbiosis of eukaryotic cell organelles.

Bacterial consortia occur in substantially different habitats like the digestive tract, oral cavity, chemocline of freshwater lakes or deep sea sediments. The different morphological types of consortia are recognised based on the taxonomy as well as the arrangement of the participating prokaryotes (Overmann, 2006) and can be characterised after functional categories. The following description of different consortia was ordered according to the utilisation of sulfur, carbon or nitrogen compounds by at least one partner bacterium involved in the physiological interaction.

Structured associations of two different hyperthermophilic archaeal species consisting of a large spherical *Ignicoccus hospitalis* cell covered by small cocci attached to its surface by thin fibres were discovered in hot submarine vent microbial communities (Fig. 1; Huber *et al.*, 2002; Burkhardt *et al.*, 2009). The crenarchaeote *I. hospitalis* has a chemoautotrophic metabolism that couples anaerobic sulfur respiration using molecular hydrogen with CO_2-fixation (Paper *et al.*, 2007). The epibionts belong to the previously unknown phylum "Nanoarchaeota" which is represented so far only by the respective species called "*Nanoarchaeum equitans*" (Huber *et al.*, 2002). In contrast to *I. hospitalis*, "*N. equitans*" is not able to grow in pure culture suggesting that its interaction with *I. hospitalis* is obligatory (Jahn *et al.*, 2008). "*N. equitans*" has the smallest archaeal genome (0.49 Mb) known to date (Waters *et al.*, 2003), thus its biosynthetic and catabolic capacity is very limited, but "*N. equitans*" was shown to obtain metabolites like lipids (Jahn *et al.*, 2004) and amino acids (Jahn *et al.*, 2008) from *I. hospitalis*. These observations indicate that the relationship of "*N. equitans*" to *Ignicoccus* is parasitic, but the exact nature of these unique microbial interactions so far remained elusive.

Consortia of filamentous sulfate-reducing bacteria with sulfide-oxidising *Thioploca* sp. have been observed in marine upwelling sediments of the South Pacific Ocean (Fig. 1; Jørgensen and Gallardo, 1999). The polysaccaride sheets of *Thioploca* are densely colonised on the in- and outside by filamentous sulfate-reducing bacteria of the genus *Desulfonema* (Jørgensen and Gallardo, 1999; Teske *et al.*, 2009). *Thioploca* takes up and stores nitrate on the sediment surface, then moves in its sheets downwards into deeper, anoxic sediment layers where sulfide is taken up and oxidized using the stored nitrate as electron acceptor. The sulfate-reducing *Desulfonema* epi- and endobionts are producing sulfide, which could be of advantage to *Thioploca* under sulfide limitation (Jørgensen and Gallardo, 1999). In turn, the chemolithoautotrophic *Thioploca* is thought to produce organic carbon sources and sulfur compounds which may be taken up by *Desulfonema*.

Spherical, immotile consortia also consisting of sulfate-reducing eubacteria, but in this case associated with methanogenic archaea occur in methane-hydrate-rich deep sea sediments and mediate the anaerobic oxidation of methane (ANME), one of the major sinks of this substantial greenhouse gas in marine environments (Fig. 1; Boetius *et al.*, 2000). In these structured consortia, up to 100 archaeal cells of the order *Methanosarcinales* form a central aggregate which is surrounded by a few layers of sulfate-reducers belonging to the *Desufosarcina*/*Desulfococcus* group. The oxidation of methane is thought to be catalyzed by the methanogens employing anaerobic methane oxidation with sulfate as terminal electron acceptor which yields carbonate, sulfide and H_2O as products (Boetius *et al.*, 2000). The sulfate-reducing partner bacteria make this reaction possible by scavenging intermediates and using acetate as carbon source. Microbial consortia consisting of methanogenic archaea or rod shaped Gram-negative bacteria orientated

longitudinally along a filamentous Gram-positive bacterium were found to be associated with the hindgut wall of the termite *Reticulitermes flavipes* (Fig. 1; Breznak and Pankratz, 1977; Leadbetter and Breznak, 1996). The attachment of the epibionts to the central bacterium is mediated by fibrous material (Breznak and Pankratz, 1977).
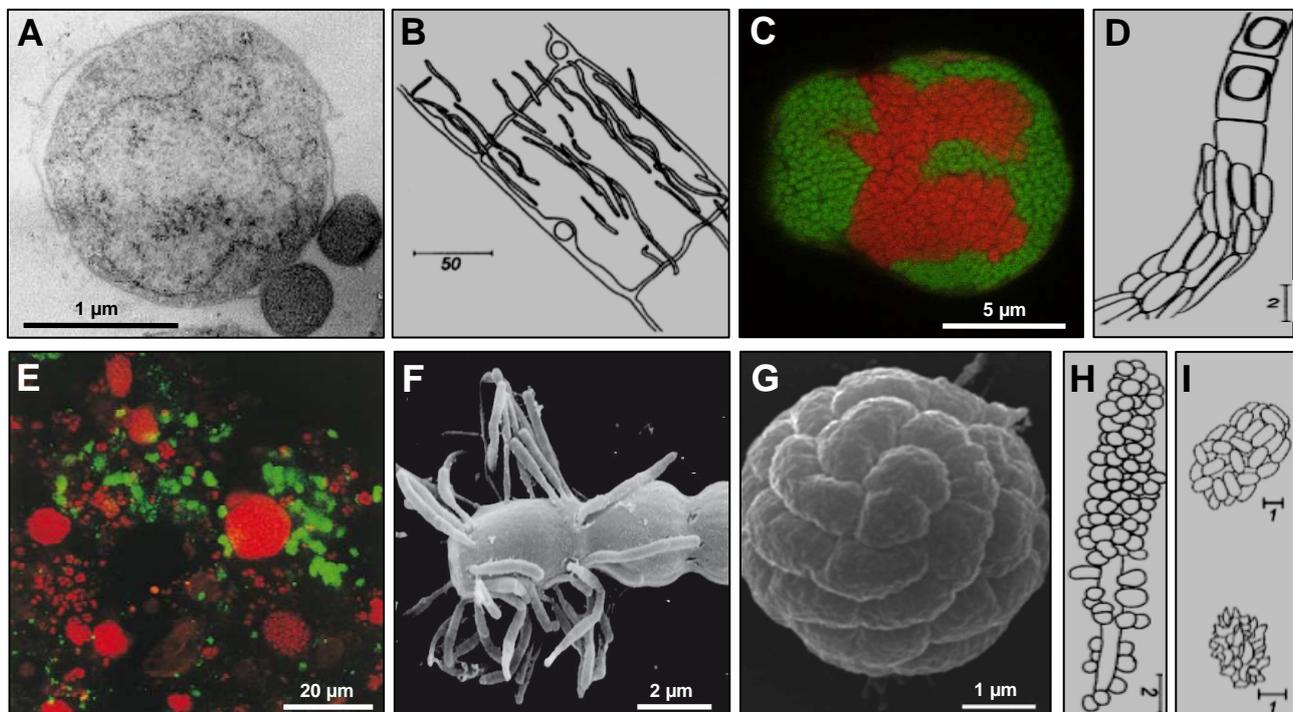
Fig. 1. Different types of prokaryotic consortia. **A.** Archaeal aggregate of *Ignicoccus hospitalis and* "*Nanoarchaeum equitans*" (Huber *et al.*, 2002). **B.** *Thioploca* sp. covered with filamentous *Desulfonema* sp. (scale bar in µm; modified after Overmann, 2001). **C.** Archaeal-bacterial ANME consortium (red: methane-oxidising archaea, green: sulfate-reducing eubacteria; Boetius *et al.*, 2000). **D.** Consortium from the hindgut microbial community of *Reticulitermes flavipes* (scale bar in µm; modified after Overmann, 2001). **E.** *Nitrobacter* sp. closely associated with clusters of *Nitrosomonas* sp. (red: ammonia-oxidising bacteria, green: nitrite-oxidising bacteria; Okabe *et al.*, 1999). **F.** *Anabaena* sp. heterocyst covered by *Rhizobium* sp. **G.** Mulicellular magnetotactic prokaryote (Winklhofer *et al.*, 2007). **H.** Corn-cob formation from dental plaque and **I.** Examples of different photothrophic consortia (scale bar in µm; modified after Overmann, 2001).

Nitrite-oxidising consortia composed of *Nitrobacter* sp. closely associated with larger cell clusters of ammonia-oxidizing *Nitrosomonas* sp. occur in nitrifying activated sludge (Fig. 1; Mobarry *et al.*, 1996). The ammonia-oxidising bacteria grow first as microcolonies, which produce nitrite and then are colonised from the outside by nitrite-consuming *Nitrobacter* aggregates. This close spatial arrangement correlates with the oxygen partial pressure gradients inside the aggregates, because molecular oxygen is important for ammonia oxidation as well as for nitrification (Wilén *et al.*,

2004). Physiological basis of this close symbiosis is the synthrophic exchange of the intermediate nitrite, which is toxic to the ammonia oxidiser, but removed by the nitrite-oxidising *Nitrobacter*.

Filamentous, limnic cyanobacteria belonging to the genus *Anabaena* were found to be colonized by numerous heterotrophic bacteria like *Rhizobium* species specifically attached to their heterocysts (Fig. 1; Paerl and Kellar, 1978). These epibionts were shown to be chemotactic towards nitrogen and carbon excretion products of the cyanobacterial heterocysts (Pearl and Pickney, 1996). Recent radiolabeling experiments suggest that carbon and nitrogen compounds fixed by the cyanobacterium are assimilated by the epibiont (Behrens *et al.*, 2008). On the other hand, the respiratory activity of the epibionts is believed to decrease the local oxygen concentration and to produce CO_2 which in turn stimulates the photosynthetic growth of the cyanobacterium (Schiefer and Caldwell, 1982). Thus this type of interspecific interaction has been described as mutualistic with benefits for both associated species (Paerl, 1984).

In other bacterial consortia, the type of physiological interaction is yet unknown. In the dental plaque of the human oral cavity "corn-cob" formations are present (Jones, 1972), which constitute of spherical streptococci arranged around the rod-shaped central corynebacterium *Bacterionema matruchotii* (Mouton *et al.*, 1977) in an ordered fashion (Fig. 1). The "corn-cob"-consortia are structurally similar to so-called phototrophic consortia which occur in freshwater environments (Fig. 1; Lauterborn, 1906).

Phototropic consortia represent unique, highly structured associations composed of a fixed number of tightly packed green or brown coloured green sulfur bacterial epibiont cells utilising sulfide as electron donor associated with a colourless, heterotrophic central bacterium in an ordered fashion. Barrel-shaped, motile phototrophic consortia were first described in the early 20th century (Lauterborn, 1906; Lauterborn, 1913; Buder, 1914). To date, eight different motile and two immotile morphotypes containing gas vesicles have been described based on their overall morphology as well as the colour and arrangement of their epibionts (Overmann and Schubert, 2002; Glaeser and Overmann, 2004). Due to their high degree of mutual interdependence, the motile phototrophic consortia, especially their only cultured representative "*Chlorochromatium aggregatum*" (Fröstl and Overmann, 1998; Pfannes *et al.*, 2007), are regarded as the most highly developed interaction between different types of prokaryotes known so far (Schink, 1991; Overmann, 2006). Since phototrophic consortia consist of two different bacterial species, their binary names are without standing in nomenclature and are accordingly given in quotation marks (Trüper and Pfennig, 1971). Phototrophic consortia occur in the chemocline of numerous stratified lakes and ponds worldwide where light reaches sulfide rich water layers at low light intensities and low sulfide concentrations (Overmann *et al.*, 1998; Glaeser and Overmann, 2004). The biomass of the anaerobic, phototrophic consortia can amount to two-thirds of the total bacterial biomass of the

chemocline (Gasol *et al.*, 1995; Glaeser and Overmann, 2003a). This dominance of green sulfur bacteria in an associated state indicates that the close contact with the central bacterium is of selective advantage for occupying this narrow, well-defined ecological niche in which the yet unknown metabolic interactions of the bacterial partners are likely to influence the biogeochemical sulfur and carbon cycling.

Interestingly, aggregates consisting of a single species of magnetotactic, marine sediment inhabiting *Deltaproteobacteria* have been shown to encompass highly coordinated cell-cell interactions as so-called magnetotactic multicellular prokaryotes (MMPs) (Keim *et al.*, 2004a). MMPs were found in the anoxic zone of various marine environments (e.g. Farina *et al.*, 1983; DeLong *et al.*, 1993; Simmons *et al.*, 2004). At present, nothing is known about the physiology and the molecular basis of the intraspecific interaction of yet uncultivated MMPs.

Although the molecular basis of symbiosis between prokaryotes has implications for the understanding of bacterial interactions with human or plant hosts, functional studies are limited because associations between prokaryotes are difficult to maintain in laboratory cultures. So far, only *Anabaena* sp.-*Rhziobium* sp. (Stevenson and Waterbury, 2006), *Ignicoccus-Nanoarchaeum* (Huber *et al.*, 2002) and ANME consortia (Nauhaus *et al.*, 2007) could be cultivated, but all with significant limitations like too slow growth rates or inaxenic culturability of one or both partners. Therefore the type of physiological interaction, molecular mechanisms of cell-cell recognition, adhesion and signal exchange as well as the degree of mutual interdependence between the partner bacteria in most cases have remained unknown. A suitable model system for a detailed investigation of highly structured consortia is thus needed. The phototrophic consortium "*C. aggregatum*" represents the first model system to elucidate the physiological and molecular basis of symbiotic interaction between non-related prokaryotes since it recently became available in fast growing enrichment culture forming an almost pure, dense monolayer biofilm (Pfannes *et al.*, 2007). Its respective epibiont *Chlorobium chlorochromatii* could be isolated in axenic culture (strain CaD; Vogl *et al.*, 2006) and the genome has been sequenced. Moreover, annotation and gap closure of the newly sequenced central bacterium genome is in progress (unpublished data). On the other hand, the magnetotactic multicellular prokaryote represent an ideal microorganism for the future development of a novel model system for bacterial multicellularity that complements the characterisation of molecular mechanisms and basic principles underlying the interspecific interaction in "*C. aggregatum*" with the investigation of a highly structured, but intraspecific multicellular association of prokaryotes.

The phototrophic consortium "*Chlorochromatium aggregatum*" and multicelluar magnetotactic prokaryotes as model systems for bacterial multicellularity

"*C. aggregatum*" consists of a highly organized assemblage of 16±3 green sulfur bacterial epibionts surrounding a central, motile betaproteobacterium which exhibits a single, polar flagellum (Fig. 1A/B; Wanner *et al*., 2008). The rod-shaped central bacterium is most likely chemoheterotrophic (Fröstl and Overmann, 2000), but could not be cultured separately from its epibionts. It represents a novel and phylogenetically isolated lineage within the *Comamonadaceae* related to *Rhodoferax* sp., *Polaromonas vacuolata* and *Variovorax paradoxus* (Kanzler *et al.*, 2005). The immotile epibiont *Chl. chlorochromatii* is an obligate anaerobic, photolithoautotrophic green sulfur bacterium representing a unique phylotype (Vogl *et al*., 2006). Its capability to grow in pure culture (Vogl *et al*., 2006) indicates that *Chl. chlorochromatii* is not obligate symbiotic. Like all known green sulfur bacteria it is utilising sulfide as electron donor and incorporates CO_2 via the reverse tricarboxylic acid cycle (Glaeser and Overmann, 2003a). The overall physiology of the epibiont resembles other free-living green sulfur bacterial species and did not reveal conspicuous differences (Vogl *et al*., 2006), but in consortia it appears to have novel, so far unknown metabolic capabilities. Thus this special kind of symbiosis is based on a so far neglected physiological interaction.
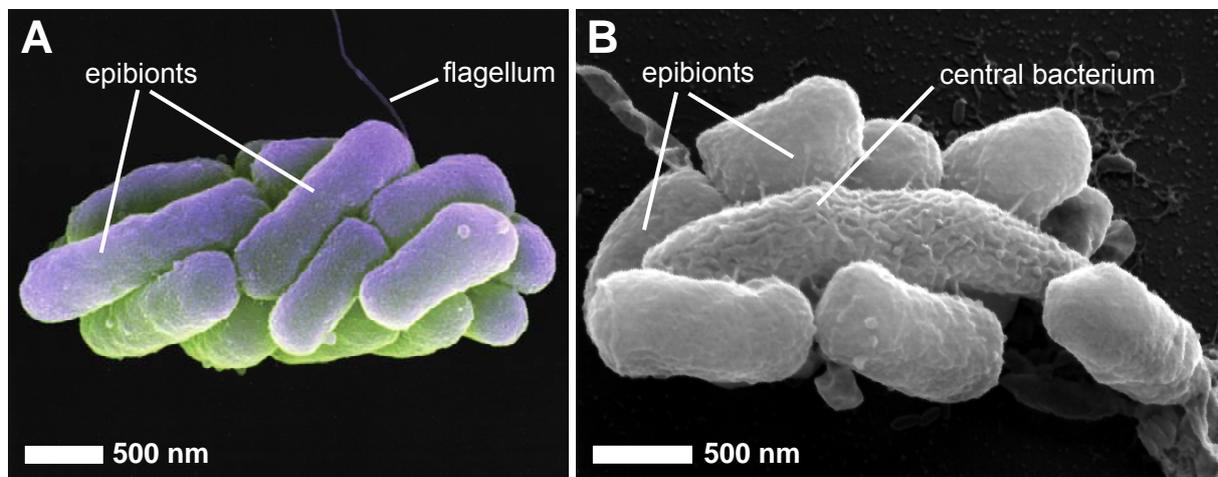
Fig. 2. Scanning electron micrographs of an **A.** intact (modified after Wanner *et al*., 2008) and **B.** disaggregated phototrophic consortium "*C. aggregatum*" showing the epibionts and the central bacterium (modified after Overmann, 2006).

A distinct type of phototrophic consortium constitutes from a single, unique phylotype of the epibiont (Fröstl and Overmann, 2000; Glaeser and Overmann, 2004) as well as of the central bacterium (Pfannes *et al*., 2007). The epibiont and central bacterium 16S rRNA gene sequences so far have only been detected in associated, but not in free-living state in their natural habitat (Glaeser and Overmann, 2004). Therefore the bacterial partners seem to be specifically adapted to life in an

obligate symbiosis with an almost unknown selective advantage and molecular basis.

However, several independent findings indicate the presence of specific mechanisms of mutual recognition and co-adaption of the partner bacteria. Phototrophic consortia show the most specialised cell division of all consortia known to date representing a coordinated, simultaneous fission of all epibionts together with the central rod during which their spatial arrangement and permanent contact is maintained (Overmann *et al.*, 1998). This observation suggests highly synchronised cell cycles. Therefore also the number and length of epibionts per consortium always stays constant as observed in enrichment cultures as well as in natural assemblages. The phototrophic consortium "*C. aggregatum*" accumulates in the light at wavelengths which correspond to the absorption maxima of the bacterial chlorophylls present in its epibionts (Fröstl and Overmann, 1998). It obviously exhibits a scotophobic response, whereby the monopolar flagellated central bacterium (Overmann *et al.*, 1998; Wanner *et al.*, 2008) confers motility to the consortium. The bacteriochlorophylls of the epibiont possibly function as light sensors (Fröstl and Overmann, 1998; Glaeser and Overmann, 2003a), but there are recent indications deriving from the central bacterium draft genome sequence that a bacteriophytochrome could represent the photoreceptor for the observed phototaxis of the consortium (unpublished data). Additionally, intact consortia exhibit a phobic reaction (Buder, 1914) when suddenly illuminated with light of high intensity. Moreover, "*C. aggregatum*" shows chemotaxis toward sulfide and 2-oxoglutarate (Fröstl and Overmann, 1998; Glaeser and Overmann, 2003b). Therefore a rapid, interspecific signal exchange must occur between the epibionts and its central bacterium. In an ecological context, the scotophobic response and chemotactic behavior enables the motile consortia to maintain their vertical position in the chemocline and probably allows them to accumulate in a zone of the water column favourable for their growth which leads to their narrow stratification observed in the chemocline of freshwater habitats. Thus one likely selective advantage of the symbiotic state for the autotrophic, but immotile epibionts is the motility conferred by the central bacterium. In turn, the central bacterium possibly uses 2-oxoglutarate produced by the epibiont as a carbon source whereas its 2-oxoglutarate uptake seems to be controlled by the physiological state of the epibiont (Glaeser and Overmann, 2003b).

Conspicuous subcellular structures of the epibionts were observed exclusively in association with the central bacterium (Vogl *et al.*, 2006; Wanner *et al.*, 2008) which most likely play an important role in the interaction of the bacterial partners in phototrophic consortia. In free-living epibionts the chlorosomes were equally distributed over the inner face of the cytoplasmatic membrane. In contrast, gaps devoid of chlorosomes are present in each symbiotic epibiont cell at its contact site to the central bacterium in intact consortia (Fig. 3A; Vogl *et al.*, 2006; Wanner *et al.*, 2008). A 17 nm-thick, layered structure composed of regularly arranged elements is attached to the

inner side of the cytoplasmatic membrane at all contact points (Fig. 3B). Since chlorosomes were absent only at the positions of this so-called epibiont contact layer (ECL) it is most likely responsible for the intracellular sorting of the chlorosomes in the epibiont (Vogl *et al.*, 2006; Wanner *et al.*, 2008). This observation indicates that the ECL represents a specific adaptation of the epibiont to its life in symbiosis which is essential for the interaction of the partner bacteria.
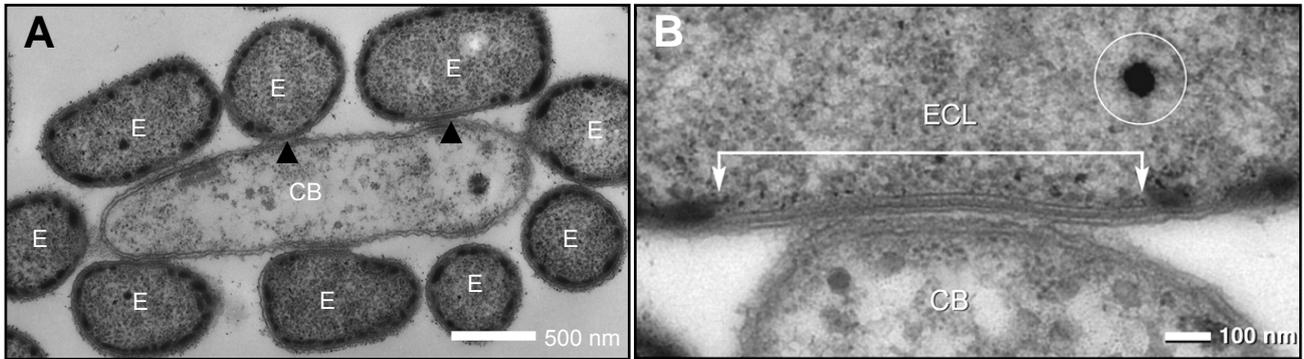
Fig. 3. Transmission electron micrographs of intact "*C. aggregatum*" consortia. **A.** Longitudinal ultrathin section of a central bacterium (CB) surrounded by its epibiont cells (E); at the attachment site, the cytoplasmatic membrane of the epibionts is free of chlorosomes (arrowheads) (Wanner, G., Overmann, J., unpublished). **B.** Epibiont contact layer (ECL) at the contact area of the epibiont cell and the central bacterium which is characterised by a layered structure (arrow bar) and the absence of chlorosomes (circle: polyphosphate globule).

The most striking feature of the central bacterium is the presence of one to three hexagonally packed, 35 nm thick and up to 1 µm long paracrystalline structures named central bacterium crystal (CBC) per cell (Fig. 4A; Wanner *et al.*, 2008). Interestingly, striated structures identical to the CBC regarding the overall shape and dimension were also observed in a multicellular magnetotactic prokaryote (Fig. 4B; Silva *et al.*, 2007). The specific occurrence of these very similar structures in nonrelated bacteria forming multicellular assemblages indicates that they could be relevant for prokaryotic cell-cell interactions in general.
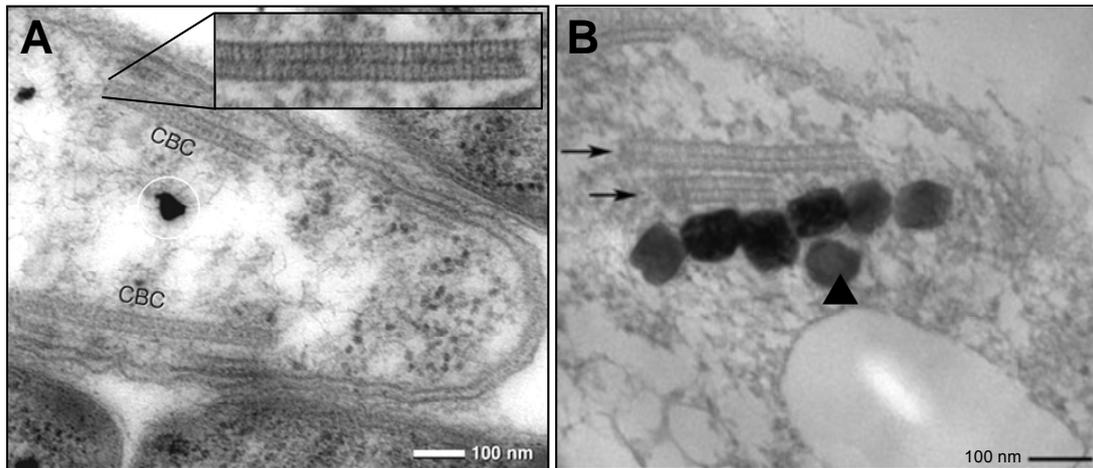
Fig. 4. Transmission electron micrographs of **A.** the central bacterium of "*C. aggregatum*" (circle: osmophilic granule; Wanner *et al.*, 2008) and **B.** a single cell of a multicellular magnetotactic prokaryote (arrowhead: magnetosome cluster; Silva *et al.*, 2007) revealing similar paracrystalline structures called central bacterium crystal (CBC in **A.**) or "striated structure" (arrows in **B.**).

Multicellular magnetotactic prokaryotes (MMPs) represent highly organized, mulberry-shaped motile aggregates (Fig. 5A) of 10-40 clonal single cells most closely related to *Desulfosarcina variabilis* (DeLong *et al.*, 1993; Abreu *et al.*, 2007; Simmons and Edwards, 2007). The flagellated single cells of the MMP are arranged around a central, acellular compartment and show a pyramidal shape (Fig. 5B). Within each MMP, the outer membranes of adjacent Gram-negative cells form contact zones resembling eukaryotic cell junctions (Fig. 5B; Rodgers *et al.*, 1990; Keim *et al.*, 2004a). MMPs move as an entire unit of 3-12 μm in diameter (Fig. 5A; Keim *et al.*, 2004a) exhibiting a conspicuous behaviour called magnetotaxis (Blakemore, 1975).
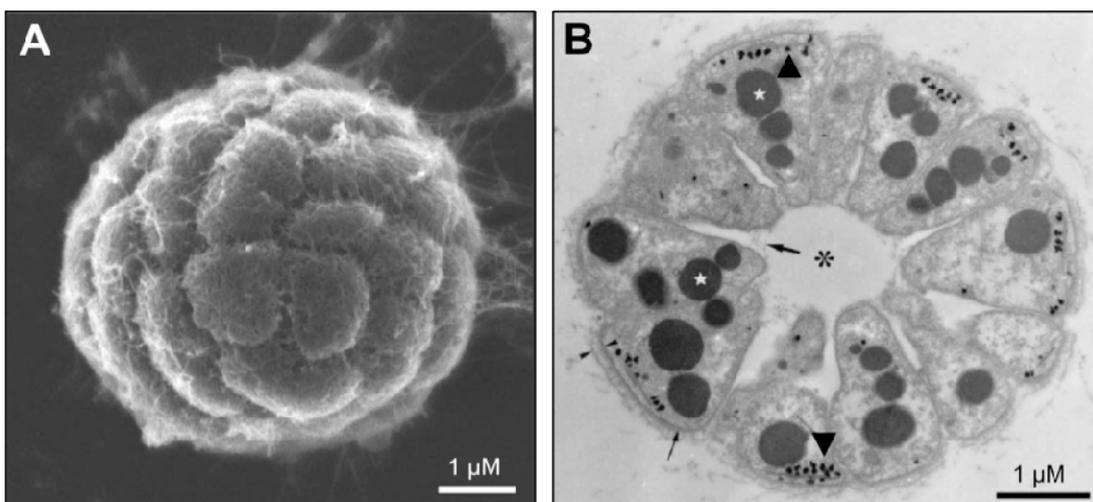
Fig. 5. Ultrastructure of a MMP analysed by **A.** scanning and **B.** transmission electron microscopy (Keim *et al.*, 2004a; black asterisk: acellular compartment, white asterisk: PHA inclusions; arrowheads: magnetosome clusters).

The swimming direction of the motile magnetotactic bacteria is influenced by magnetic field lines (Blakemore, 1975), which is based upon so-called magnetosomes representing intracellular synthesised magnetic crystals consisting of iron oxide or iron sulfide (Balkwill *et al.*, 1980) enclosed by a lipid bilayer membrane (Gorby *et al.*, 1988). These single domain crystals consist of magnetite (Fe_3O_4) or greigite (Fe_3S_4) and represent elementary magnets exhibiting a permanent dipole moment. Each of the individual MMP cells has its own magnetosomes (Fig. 4B, 5B), which typically contain 75 to 120 nm long crystals with cubical or rectangular shape predominantly composed of the magnetic iron-sulfur mineral greigite (Fe_3S_4) (Mann *et al.*, 1990; Pósfai *et al.*, 1998). In magnetotactic bacteria the net magnetic moment of the cell corresponds to the sum of dipoles of all magnetosomes allowing them to passively orient themselves like little compass needles parallel to magnetic field lines. A commonly accepted hypothesis for the relevance of this cellular magnetism for these ubiquitously distributed bacteria is the orientation in their typical micro-habitat (Blakemore, 1975), the oxic-anoxic transition zones of marine and limnic sediments or stratified water bodies (Spring and Bazylinski, 2000), which exhibits vertical chemical and redox gradients. Magnetotaxis is likely to be coupled with chemo- and aerotaxis (Frankel *et al.*, 1997) to enable the microaerophilic or anaerobic bacteria to locate themselves in areas with optimal oxygen and nutrient concentrations (Balkemore and Blakemore, 1990; Schüler *et al.*, 1999). Maintenance of their position in the habitat is most likely regulated only by chemotaxis (Blakemore, 1982; Mann *et al.*, 1990). The advantage of magnetotaxis compared to a purely chemotactic and/or aerotactic response is the reduction of the given three-dimensional orientation problem to a one-dimensional one by using the earth magnetic field to reach sediment zones exhibiting ideal oxygen conditions more quickly than non-magnetotactic bacteria. The magnetotactic bacteria are currently thought to respond to high oxygen levels by swimming along the earth magnetic field lines downward into areas with low or no oxygen toward the geomagnetic north in the Northern (north-seeking) and vice versa south-seeking in the Southern Hemisphere.

Although consisting of individual cells, the MMPs orient themselves within a magnetic field like single-celled magnetotactic bacteria. MMPs move in an oriented, rotating fashion (Greenberg *et al.*, 2005) indicating that the flagellar motion is highly coordinated by communication between the interacting cells of the multicellular aggregate. MMPs exhibit clockwise and counter-clockwise rotation as well as a "ping-pong" motion (Rodgers *et al.*, 1990), also called escape motility (Keim *et al.*, 2007). This latter type of movement is typical for MMPs, but unique among all magnetotactic bacteria. It consists of extended phases of swimming along the magnetic field lines interrupted by excursions in the reverse direction during which the aggregates swim at about double the speed as during forward motion (Greenberg *et al.*, 2005). The excursion distance seems to depend on the applied artificial magnetic field (Greenberg *et al.*, 2005). The reason for this unique magnetotactic

behaviour is yet unknown and can not be explained by the conventional model of magnetotaxis which gave rise to the hypothesis, that MMPs posses a complex magnetoreception controlled by diverse molecular mechanisms (Kirschvink *et al.*, 2001; Greenberg *et al.*, 2005).

Also the life cycle of the MMPs seems to be very elaborate. It comprises a synchronous division of all cells, leading to two identical daughter aggregates without the formation of a single celled stage (Fig. 6; Keim *et al.*, 2004b, 2007).
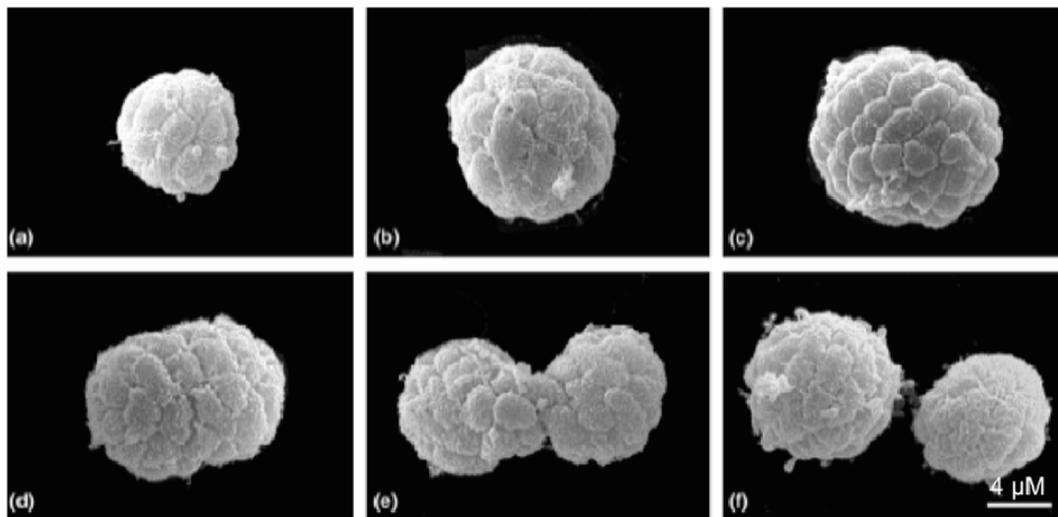
Fig. 6. Multicellular division of individual MMPs in different stages of the division cycle visualised by scanning electron microscopy (Keim *et al.*, 2004b). An elliptical (d) is followed by an eight-shaped structure (e) in which the single cells maintain their helicoidal arrangement.

If the magnetosomes of the individual cells within one MMP were oriented in a random fashion, the resulting magnetic moment of the entire aggregate would be significantly decreased. A continuously multicellular lifestyle is thus necessary for the MMP to preserve its magnetotactically optimized state during reproduction and to retain its capability to swim along magnetic field lines (Keim *et al.*, 2004 a,b; Winklhofer *et al.*, 2007). The optimised magnetotactic moment allowing the fast escape motility along geomagnetic field lines as well as the sophisticated multicellular reproduction mechanism of the MMP seem to be the consequence of an evolutionary pressure to prevent predation from flagellates or ciliates (Martins *et al.*, 2007). In the competition with other environmental bacteria, avoiding predation simply by size possibly represents a major structural advantage of multicellularity in prokaryotes despite their efficient exchange of soluble compounds by close cell contact.

The required intercellular communication for the coordinated multicellular movement and cell division of the MMP has been suggested to proceed via its internal acellular compartment by soluble molecules or membrane vesicles as well as via the cell-cell contact sites (Fig. 5B; Keim *et al.*, 2004a). The outer membranes of opposing cells in the aggregate present extensive, 2 nm wide

regions of constant distance indicative of a specialised structure joining the two cells (Rodgers *et al.*, 1990; Keim *et al.*, 2004a) resembling eukaryotic cell junctions, which mediate intercellular adhesion and communication. Despite these hypotheses deducted from ultrastructural observations nothing is known so far about communication and interaction between the single cells. Furthermore, a close cell-cell interaction and interdependence is suggested by the fact that individual cells of disaggregated MMPs rapidly loose viability (Abreu *et al.*, 2006) and free-living single cells were never observed (e.g. Rodgers *et al.*, 1990; Lins and Farina, 1999) as it is also the case with the phototrophic consortium "*C. aggregatum*" (Glaeser and Overmann, 2004). If the MMP and "*C. aggregatum*" in fact are multicellular organisms based upon the above indications depends on the concept applied to define multicellularity. The traditional concept derived from studies on eukaryotes includes the presence of different cell types and thus precludes any known prokaryotes to be defined as multicellular as they generally represent microorganisms without cell differentiation. New findings on the molecular mechanisms of intercellular interactions between bacteria challenged this concept of prokaryotes being defined as unicellular and independent microorganismsn and novel concepts of multicellularity were generated as exemplified by the proposal that multicellular organisms are such that at least have many cells in close contact that coordinate movements, metabolic activities and growth (Shapiro, 1998; Kaiser, 2001). A more conservative concept to consider an organism multicellular requires a characteristic shape and cell organisation as well as the lack cell autonomy and competition (Carlile, 1980). According to this definition, highly organised prokaryotes like filamentous cyanobacteria, actinomycetes or myxobacteria are not multicellular because their single cells are relatively autonomous, which also applies at least for the epibiont of the phototrophic consortium "*C. aggregatum*". In fact, there is presently a clear tendency to consider also such microorganisms as multicellular according to Kaiser (2001). On the other hand, based on the present knowledge, the MMP fulfils all requirements for multicellularity proposed by Carlile (1980), Shapiro (1998) and Kaiser (2001): it has a specific, well defined cellular organisation including specialised membrane structures, the single cells are not autonomous and do not seem to compete as they coordinate their highly complex synchronous cell division and movement as one multicellular microorganism. Therefore MMP represents the most highly evolved type of all intraspecific bacterial associations known so far and would be an ideal future model system for studying the basis of prokaryotic multicellularity, which not yet could be assed further due to the lack of culturability. On the other hand, the culturable consortium "*C. aggregatum*" represents an established model system to analyse the basic principles and molecular determinants of interspecific multicellularity of bacteria.

Aims of the present study

Interactions between prokaryotes in unstructured populations have been studied intensively, whereas much less is known about the basis of inter- and intraspecific multicellularity in highly structured bacterial consortia. This stands in sharp contrast to the important role of the highly structured phototrophic consortium "*C. aggregatum*" as the first cultivable model system for the investigation of close interactions between different types of prokaryotes. "*C. aggregatum*" provides the unique opportunity to dissect the molecular basis of this interspecific bacterial symbiosis and to elucidate general principles of interspecific multicellularity in bacteria. To date, the molecular level of the interaction between the epibionts with the central bacterium is not understood in sufficient detail. Since the genome of the epibiont *Chl. chlorochromatii* recently became available, the main focus of the present study has been a detailed molecular characterisation of the factors that underlie its symbiotic properties.

Based on its symbiotic way of life and considerable phylogenetic distance to all other green sulfur bacteria, symbiosis-specific genes were expected to occur in the epibiont genome. In order to identify genes involved in the symbiotic interaction, selected strains of green sulfur bacteria covering different physiological types were included in a genomic DNA suppression subtractive hybridisation approach which appeared promising since the members of this group are phylogenetically rather closely related, but the major difference between the free-living strains and the epibiont is its symbiotic state. The presence of symbiosis-specific genes in the epibiont genome was analysed further by an *in silico* subtractive hybridisation of all other available green sulfur bacterial genomes, which represent the whole physiological and phylogenetic diversity of the group. To extend the molecular analysis of "*C. aggregatum*" beyond putative symbiosis genes, specific symbiosis depended regulation of the epibiont in symbiotic compared to its free-living state should be identified on the transcriptomic and proteomic level by novel, comparative functional genomics techniques including cDNA suppression subtractive hybridisation, whole transcriptome sequencing and two-dimensional difference gel electrophoresis (2-D DIGE). To provide the basis for the future development of a novel model system for intraspecific multicellularity, a novel type of MMP discovered at the beginning of this study was intended to be analyzed in close detail using methods like chemotaxis assays and analysis of genes involved in potential key metabolic functions that already proved of value in establishing "*C. aggregatum*" as a cultivable model system. Since subcellular structures very similar to those found in "*C. aggregatum*" also occur in the nonrelated MMPs, there is a good indication that they could be relevant for prokaryotic cell-cell interactions in general. Thus the MMP would represent an ideal future model system to complement the characterization of molecular mechanisms and basic principles underlying the interspecific interactions in "*C. aggregatum*" with a highly structured intraspecific multicellular association.

References

Abrams, P.A. (1987) On classifying interactions between populations. *Oecologia* 73: 272-281

Abreu, F., Silva, K.T., Martins, J.L. and Lins, U. (2006) Cell viability in multicellular magnetotactic prokaryotes. *Int Microbiol* 9: 267-271

Abreu, F., Martins, J.L., Silveira, T.S., Keim, C.N., Lins de Barros, H.G.P., Filho, F.J.G. and Lins, U. (2007) 'Candidatus *Magnetoglobus multicellularis*', a multicellular, magnetotactic prokaryote from a hypersaline environment. *Int J Syst Evol Microbiol* 57: 1318-1322

Bassler, B.L. (2002) Small talk. Cell-to-cell communication in bacteria. *Cell* 109: 421-424

Behrens, S., Lösekann, T., Pett-Ridge, J., Weber, P.K., Wing-On, N., Stevenson, B.S., Hutcheon, I.D., Relman, D.A., Spormann, A.M. (2008) Linking microbial phylogeny to metabolic activity at the single-cell level by using enhanced element labeling-catalyzed reporter deposition fluorescence *in situ* hybridization (EL-FISH) and NanoSIMS. *Appl Environ Microbiol* **74:** 3143-3150

Blakemore R.P. (1975) Magnetotactic bacteria. *Science* 190: 377-379

Blakemore R.P. (1982) Magnetotactic bacteria. *Ann Rev Microbiol* 36: 217-238

Blakemore R.P., Blakemore N. (1990) Magnetotactic magnetogens. In *Iron Biominerals*. Frankel, R.B., Blakemore, R.P. (eds.). New York, NY, USA: Plenum Press, pp. 51-68

Breznak, J.A., Pankratz, H.S. In situ morphology of the gut microbiota of wood-eating termites *Reticulitermes flavipes* (Kollar) and *Coptotermes formosanus* Shiraki (1977) *Appl Environ Microbiol* 33: 406-426

Boetius, A., Ravenschlag, K., Schubert, CJ., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jørgensen, BB., Witte, U., Pfannkuche, O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* 407: 623-626

Carlile, M.J. (1980) From prokaryote to eukaryote: gains and losses. In *The eukaryotic microbial cell*. Gooday, G.W., Lloyd, D., Trinci, A.P.J. (eds.). Cambridge, United Kingdom: Cambridge University Press, pp. 1-38

Buder, J. (1914) *Chloronium mirabile. Berichte Dtsch Bot Ges* 31: 80-97

Burghardt, T., Junglas, B., Siedler, F., Wirth, R., Huber, H., Rachel, R. (2009) The interaction of *Nanoarchaeum equitans* with *Ignicoccus hospitalis*: proteins in the contact site between two cells. *Biochem Soc Trans* 37: 127-132

De Bok, F.A., Plugge, C.M., Stams, A.J. (2004) Interspecies electron transfer in methanogenic propionate degrading consortia. *Water Res* 38: 1368-1375

De Long E.F., Frankel R.B., Bazylinski D.A. (1993) Multiple evolutionary origins of magnetotaxis in bacteria. *Science* 259: 803-806

Farina, M., Lins de Barros, H., Motta de Esquivel, D. and Danon, J. (1983) Ultrastructure of a magnetotactic microorganism. *Biol Cell* 48: 85-88

Frankel, R.B., Bazylinski, D.A., Johnson, M.S., Taylor, B.L. (1997) Magneto-aerotaxis in marine coccoid bacteria. *Biophys J* 73: 994-1000

Fröstl, J.M., Overmann, J. (1998) Physiology and tactic response of "*Chlorochromatium aggregatum*". *Arch Microbiol* 169: 129-135

Fröstl, J.M., Overmann, J. (2000) Phylogenetic affiliation of the bacteria that constitute phototrophic consortia. *Arch Microbiol* 174: 50-58

Gasol, J.M., Jiergens, K., Massana, R., Calderón-Paz, J.I., Pedrós-Albió, C. (1995) Mass development of *Daphnia pulex* in a sulfide-rich pond (Lake Ciso). *Arch Microbiol* 132: 279-296

Glaeser, J. and Overmann J. (2003a) Characterization and *in situ* carbon metabolism of phototrophic consortia. *Appl Environ Microbiol* 7: 3739-3750

Glaeser, J. and Overmann J. (2003b) The significance of organic carbon compounds for *in situ* metabolism and chemotaxis of phototrophic consortia. *Environ Microbiol* 5: 1053-1063

Glaeser, J. and Overmann J. (2004) Biogeography, evolution and diversity of epibionts in phototrophic consortia. *Appl Environ Microbiol* 70: 4821-4830

Greenberg, M., Canter, K., Mahler, I., Tornheim, A. (2005) Observation of magnetoreceptive behavior in a multicelllular magnetotactic prokaryote in higher than geomagnetic fields. *Biophys J* 88: 1496-1499

Gorby, Y.A., Beveridge, T.J., Blakemore, R.P. (1988) Characterisation of the bacterial magnetosome membrane. *J Bacteriol* 170: 834-841

Harmsen, H.J., Kengen, H.M., Akkermans, A.D., Stams, A.J., de Vos, W.M. (1996) Detection and localization of syntrophic propionate-oxidizing bacteria in granular sludge by in situ hybridization using 16S rRNA-based oligonucleotide probes. *Appl Environ Microbiol* 62: 1656-1663

Huber, H., Hohn, MJ., Rachel, R., Fuchs, T., Wimmer, VC., Stetter, KO. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417: 63-67

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

Jahn, U., Summons, R., Sturt, H., Grosjean, E., Huber, H. (2004) Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I. *Arch Microbiol* 182: 404-413

Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K.O., Rachel, R., Huber, H. (2008) *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. *J Bacteriol* 190: 1743-1750

Jones, S.J. (1972) Special relationship between spherical and filamentous microorganisms in nature human dental plaque. *Arch Or Biol* 17: 613-616

Kaiser, D. (2001) Building a multicellular organism. *Annu Rev Genet* 35: 103-123

Kanzler, B.E.M., Pfannes, K.R., Vogl, K., Overmann, J. (2005) Molecular characterization of the non-photosynthetic partner bacterium in the consortium *"Chlorochromatium aggregatum"*. *Appl Environ Microbiol* 71: 7434-7441

Keim, C.N., Abreu, F., Lins, U., Lins de Barros, H., Farina, M. (2004a) Cell organisation and ultrastructure of a magnetotactic multicellular organism. *J Struct Biol* 145: 254-262

Keim, C.N., Martins, J.L., Abreu, F., Rosado, A.S., Lins de Barros, H., Borojevic, R., Lins, U., Farina, M. (2004b) Multicellular life cycle of magnetotactic multicellular prokaryotes. *FEMS Microbiol Lett* 240: 203-208

Keim, C.N., Martins, J.L., Lins de Barros, H., Lins, U. and Farina, M. (2007) Structure, behaviour, ecology and diversity of multicellular magnetotactic prokaryotes. In *Magnetoreception and Magnetosomes in Bacteria*. Schüler, D. (ed.). Berlin, Heidelberg, Germany: Springer, pp. 103-132

Kirschwink, J.L., Walker, M.M., Diebel, C.E. (2001) Magnetite based magnetorezeption. *Curr Opin Neurobiol* 11: 462-467

Kolenbrander, P.E., London, J. (1993) Adhere today, here tomorrow: oral bacterial adherence. *J Bacteriol* 175: 3247-3252

Krebs, C.J. (2001) *Ecology: the experimental analysis of distribution and abundance*. USA: Benjamin-Cummings Publishing Company, 5th edition

Lauterborn, R. (1906) Zur Kenntnis der sapropelischen Flora. *Allg Bot* 2: 198-197

Lauterborn, R. (1913) Die sapropelische Lebewelt. *Allg Bot* 19: 97-100

Leadbetter, J.R., Breznak, J.A. Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes* (1996). *Appl Environ Microbiol* 62: 3620-3631

Lins, U. and Farina, M. (1999) Organisation of cells in magnetotactic multicellular aggregates. *Microbiol Res* 154: 9-13

Mann, S., Sparks, N.H.C., Frankel, R.B., Bazylinski, D.A., Jannsch, H.W. (1990) Biomineralisation of ferrimagnetic greigite (Fe_3S_4) and iron pyrite (FeS_2) in a magnetotactic bacterium. *Nature* 343: 258-261

Martins, J.L., Silveira, T.S., Abreu, F., Silva, K.T., da Silva-Neto, I.D. and Lins, U. (2007) Grazing protozoa and magnetosome dissolution in magnetotactic bacteria. *Environ Microbiol* 9: 2775-2781

Mobarry, B.K., Wagner, M., Urbain, V., Rittmann, B.E., Stahl, D.A. (1996) Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria. *Appl Environ Microbiol* 62: 2156-2162

Moran, N.A. (2006) Symbiosis. *Curr Biol* 16: 866-871

Mouton, C., Reynolds, H., Genco, R.J. (1977) Combined micromanipulation, culture and immunofluorescent techniques for isolation of the coccal organisms comprising the "corn cob" configuration of human dental plaque. *J Biol Buccale* 5: 321-332

Nauhaus, K., Treude, T., Boetius, A., Krüger, M. (2005) Environmental regulation of the anaerobic oxidation of methane: a comparison of ANME-I and ANME-II communities. *Environ Microbiol*: 98-106

Paper, W., Jahn, U., Hohn, M.J., Kronner, M., Näther, D.J., Burghardt, T., Rachel, R., Stetter, K.O., Huber, H. (2007) *Ignicoccus hospitalis* sp. nov., the host of 'Nanoarchaeum equitans'. *Int J Syst Evol Microbiol* 57: 803-808

Pfennig, N. (1980) Synthrophic mixed cultures and symbiotic consortia with phototrophic bacteria: a review. In: *Anaerobes and anaerobic infections*. Gottschalk, G., Pfennig, N., Werner, G. (eds.). Stuttgart: Fischer, pp. 127-131

Paerl, H.W., Kellar, P.E. (1978) Significance of bacterial *Anabaena* (*Cyanophyceae*) associations with respect to N_2 fixation in freshwater. *J Phycol* 14:254-260

Paerl, H.W. (1984) Transfer of N_2 and CO_2 fixation products from *Anabaena oscillarioides* to associated bacteria during inorganic carbon sufficiency and deficiency. *J Phycol* 20: 600–608

Paerl, H.W., Pinckney, J.L. (1996) A Mini-review of Microbial Consortia: Their Roles in Aquatic Production and Biogeochemical Cycling. *Microb Ecol* 31: 225-247

Pfannes, KR., Vogl, K., Overmann J. (2007) Heterotrophic symbionts of phototrophic consortia: members of a novel diverse cluster of *Beta-proteobacteria* characterized by tandem rrn operon structure. *Environ Microbiol* 9: 2782-2794

Pósfai, M., Buseck, P.R., Bazylinski, D.A., Frankel, R.B. (1998) Iron sulfides from magnetotactic bacteria: structure, composition and phase transitions. *Am Mineral* 83: 1469-1481

Reichenbach, H., Dworkin, M. (1992) The Myxobacteria. In: *The Prokaryotes*. Trüper, H.G., Balows, A., Dworkin, M., Harder, W., Schleifer, K.-H. (eds.). Berlin, Heidelberg, Germany: Springer, pp. 3416-3487

Rodgers, F.G., Blakemore, R.P., Blakemore, N.A., Frankel, R.B., Bazylinski, D.A., Maratea, D., Rodgers, C. (1990) Intercellular structure in a many-celled magnetotactic prokaryote. *Arch Microbiol* 145: 18-22

Schiefer, G. E., Caldwell D. E. (1982) Synergistic interaction between *Anabaena* and *Zoogloea* spp. in carbon dioxide-limited continuous cultures. *Appl Environ Microbiol* 44: 84–87

Schink, B. (1991) Synthrophism among prokaryotes. In *The Prokaryotes*. Balows, A., Trüper H.G., Dworkin, M., Schleifer, K.H. (eds.). Berlin, Heidelberg, Germany: Springer, pp. 276-299

Schink, B. (2002) Synergistic interactions in the microbial world. *Antonie Van Leeuwenhoek* 81: 257-261

Schüler, D. (1999) Formation of magnetosomes in magnetotactic bacteria. *J Mol Microbiol Biotechnol* 1: 79-86

Shapiro, J.A. (1998) Thinking about bacterial populations as multicellular organisms. *Annu Rev Microbiol* 52: 81-104

Silva, K.T., Abreu, F., Almeida, F.P., Keim, C.N., Farina, M. and Lins, U. (2007) Flagellar apparatus of south seeking many celled magnetotactic prokaryotes. *Microsc Res Tech* 70: 10-17

Simmons, S.L., Sievert, S.M., Frankel, R.B., Bazylinski, D.A. and Edwards, K.J. (2004) Spatiotemporal distribution of marine magnetotactic bacteria in a seasonally stratified costal salt pond. *Appl Environ Microbiol* 70: 6230-6239

Simmons, S.L. and Edwards, K.J. (2007) Unexpected diversity in populations of the many-celled magnetotactic prokaryote. *Environ Microbiol* 9: 206-215

Spring, S., Bazylinski, D.A. (2000) Magnetotactic bacteria. In *The Prokaryotes*. Dworkin *et al.* (eds.). New York, NY, USA: Springer (online)

Stevenson, B.S. and Waterbury, J.B. (2006) Isolation and identification of an epibiotic bacterium associated with heterocystous *Anabaena* cells. *Biol Bull* 210: 73-77

Teske, A., Jørgensen, B.B., Gallardo, V.A. (2009) Filamentous bacteria inhabiting the sheaths of marine *Thioploca* spp. on the Chilean continental shelf. *FEMS Microbiol Ecol* 68: 164-172

Trüper, H.G., Pfennig, N. (1971) New nomenclatural combinations in the phototrophic sulfur bacteria. *Int J Syst Bacteriol* 21: 8-10

Okabe, S., Satoh, H., Watanabe, Y. (1999) *In situ* analysis of nitrifying biofilms as determined by *in situ* hybridization and the use of microelectrodes. *Appl Environ Microbiol* 65: 3182-3191

Overmann, J., Tuschak, C., Sass, H., Fröstl, J.M. (1998) The ecological niche of the consortium "*Pelochromatium roseum*" *Arch Microbiol* 169: 120-128

Overmann, J., Schubert, K. (2002) Phototrophic consortia, model systems for symbiotic interrelations between prokaryotes. *Arch Microbiol* 177: 201-208

Overmann, J. (2006) (ed.) *Molecular basis of symbiosis. Progress in molecular and subcellular biology*, Vol. 41. Berlin, Heidelberg, Germany: Springer, pp. 21-37

Vogl, K., Glaeser, J., Pfannes, K.R., Wanner, G., Overmann, J. (2006) *Chlorobium chlorochromatii* sp. Nov., a symbiotic green sulfur bacterium isolated from the phototrophic consortium *"Chlorochromatium aggregatum"*. *Arch Microbiol* 185: 363-372

Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M., Noordewier, M. (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA* 100: 12984-12988

Wanner, G., Vogl, K., Overmann, J. (2008) Ultrastructural characterization of the prokaryotic symbiosis in *"Chlorochromatium aggregatum"*. *J Bacteriol* 190: 3721-3730

Wilén, B.M., Gapes, D., Blackall, L.L., Keller, J. (2004) Structure and microbial composition of nitrifying microbial aggregates and their relation to internal mass transfer effects. *Water Sci Technol* 50: 213-220

Winkelhofer, M., Abraçado, L.G., Davila, A.F., Keim, C.N. and Lins de Barros, H.G.P. (2007) Magnetic optimization in a mulicellular magnetotactic organism. *Biophys J* 92: 661-670

Whittaker, C.J., Klier, C.M., Kolenbrander, P.E. (1996) Mechanisms of adhesion by oral bacteria. *Annu Rev Microbiol* 50: 513-552

Chapter 3

Identification and analysis of four candidate symbiosis genes from "*Chlorochromatium aggregatum*", a highly developed bacterial symbiosis

Summary

The consortium "*Chlorochromatium aggregatum*" currently represents the most highly developed interspecific association between prokaryotes. It consists of green sulfur bacteria, so-called epibionts, which surround a central, motile, chemotrophic bacterium. Four putative symbiosis genes of the epibiont were recovered by suppression subtractive hybridisation and bioinformatics approaches. These genes are transcribed constitutively and do not occur in the free-living relatives of the epibiont. The hemagglutinin-like putative gene products of ORFs Cag_0614 and Cag_0616 are unusually large and contain repetitive regions and RGD tripeptides. Cag_0616 harbors two $\beta\gamma$-crystallin Greek key motifs. Cag_1920 codes for a putative hemolysin whereas the gene product of Cag1919 is a putative RTX-like protein. Based on detailed analyses of Cag_1919, the C-terminal amino acid sequence comprises six repetitions of the motif GGXGXD predicted to form a Ca^{2+}-binding beta roll. Intact "*C. aggregatum*" consortia disaggregated upon the addition of EGTA or pyrophosphate, but stayed intact in the presence of various lectine-binding sugars or proteolytic enzymes. Unlike other RTX toxins, a gene product of Cag_1919 could not be detected by $^{45}Ca^{2+}$-autoradiography, indicating a low abundance of the corresponding protein in the cells. The RTX-type C-terminus coded by Cag_1919 exhibited a significant similarity to RTX modules of various proteobacterial proteins, suggesting that this putative symbiosis gene has been acquired via horizontal gene transfer from a proteobacterium.

Introduction

Functional studies of microbial symbioses have so far focussed on the associations of bacteria with eukaryotes (Overmann and Schubert, 2002; Overmann, 2006). Meanwhile, an increasing number of spatially close associations of prokaryotes have been discovered (Overmann, 2001a). In so-called consortia, different types of prokaryotes maintain a permanent cell-to-cell contact. Such associations occur in habitats like the dental plaque, the digestive tract, deep sea sediments or in upflow anaerobic sludge bed reactors for wastewater treatment, where they catalyze key metabolic processes. So far, however, only the interspecific syntrophic associations involved in methanogenic degradation are understood in sufficient detail (Schink, 1998, 2002).

Phototrophic consortia are highly structured associations between green sulfur bacteria and a chemotrophic bacterium. Most consist of one motile, colourless *Betaproteobacterium* surrounded by up to 69 green sulfur bacterial epibionts (Fröstl and Overmann, 2000; Overmann, 2001b; Overmann and Schubert, 2002). Phototrophic consortia were discovered a century ago (Lauterborn, 1906) and typically occur in the chemocline of stratified lakes (Overmann *et al.*, 1998; Glaeser and Overmann, 2004) where they can amount to two-thirds of the total bacterial biomass (Gasol *et al.*, 1995). 19 different types of phototrophic consortia have been described to date (Glaeser and Overmann, 2004). Several independent experimental findings indicate that a rapid signal transfer occurs between the epibionts and the central bacterium: (1) the cell division between the partners is highly coordinated (Overmann *et al.*, 1998), (2) consortia accumulate scotophobotactically in the light, whereby the central rod confers motility to the consortium and the epibionts act as light sensors (Fröstl and Overmann, 1998; Glaeser and Overmann, 2003), and (3) the carbon uptake of the central bacterium seems to be controlled by the epibiont (Glaeser and Overmann, 2003). Phototrophic consortia thus represent the most highly developed interspecific association of bacteria which is currently recognized and therefore serve as a valuable model system to study the molecular basis of cell-cell interactions between different prokaryotes.

Phototrophic consortia can be maintained in laboratory cultures (Fröstl and Overmann, 1998). Recently, the epibiont of the phototrophic consortium "*Chlorochromatium aggregatum*", *Chlorobium chlorochromatii* strain CaD, could be isolated in pure culture. Extensive physiological analyses of this strain did not reveal conspicuous differences to free-living green sulfur bacteria (Vogl *et al.*, 2006). Although *Chl. chlorochromatii* is capable of growing in pure culture and hence is not obligately symbiotic, neither it nor any other epibiont of the 19 known types of phototrophic consortia has been detected in a free-living state in the natural habitat (Glaeser and Overmann, 2004). The epibionts therefore seem to be specifically adapted to life in the symbiotic state, but the molecular basis of this adaptation is entirely unknown.

The goal of the present work was to identify symbiosis-specific genes of *Chl. chlorochromatii*.

Putative symbiosis genes of the epibiont were recovered by suppression subtractive hybridisation and bioinformatics approaches. Several candidate genes were detected. At least one of these genes is likely to be involved in a specific Ca^{2+}-dependent cell-cell-adhesion of the symbiotic association.

Results

Identification and localization of putative symbiosis genes

Suppression subtractive hybridisation of genomic DNA of *Chl. chlorochromatii* CaD against that of 16 free-living green sulfur bacteria yielded amplification products of 10 different size classes (Suppl. Fig. 1). Representatives of each size class were sequenced. Sequences of clones EpiSH1 and EpiSH8 were found in two size classes but represented the same sequence type. This was also the case for clones EpiSH6 and EpiSH7. Six of the remaining 8 different sequence types (clones EpiSH2, EpiSH3, EpiSH5, EpiSH6, EpiSH9, EpiSH10) coded for enzymes of the central nucleic acid metabolism, amino acid and protein metabolism, or bacteriochlorophyll biosynthesis. Corresponding genes are also present in genomes of other green sulfur bacteria (Table 1).

In contrast, the sequences of the two clones EpiSH4 and EpiSH8 were previously unknown for green sulfur bacteria or exhibited only a low similarity to the available sequences (Table 1). The corresponding open reading frames (ORFs) in the genome sequence of *Chl. chlorochromatii* CaD were Cag_0616 and Cag_1920, respectively. As annotated, ORF Cag_0616 would code for an extremely large putative protein of 20,646 amino acids (aa) exhibiting low similarity to a hemagglutinin of *Magnetococcus* sp. MC-1. ORF Cag_1920 codes for a 3834 aa long protein which shows little similarity to a BNR glycosyl hydrolase with a hemolysin-type calcium binding region (Table 1). Subsequent genome analyses of the neighbourhood of ORFs Cag_0616 and Cag_1920 revealed the presence of two additional unusual sequences, Cag_0614 and Cag_1919. Cag_0614 shows considerable sequence similarity to downstream ORF Cag_0616 and would code for an even longer (36,805 aa) gene product (Table 1). The 2-kb-long region between Cag_0614 and 0616 harbors a sequence coding for an outer membrane like efflux protein (Cag0615) and a promoter region in the intergenic region upstream of Cag_0616. Another additional sequence, ORF Cag_1919, is located upstream of Cag_1920 and codes for a protein of 1526 aa which exhibits low similarity to a glycosyl hydrolase with bacterial neuraminidase repeat (BNR) and a hemolysin-type calcium binding region. No promoter region was identified in the 67 bp-long intergenic region between Cag_1919 and 1920.

Table 1. Results of BLAST searches of amino acid sequences deduced from SSH clone sequences (for size classes of clone types see Suppl. Fig. 1)

SSH clone/ ORF*	DNA insert/ ORF (bp)	Accession No.	BLAST X hits	Expected value	Similarity of amino acid sequences
EpiSH2/ Cag_1814	1438/ 1623	ZP_00590656.1	Light-independent protochlorophyllide reductase, B subunit, BchB [*Chlorobium clathratiforme* DSMZ 5477^T]	3e-85	Identities = 161/242(66%) Positives = 183/242 (75%)
EpiSH3/ Cag_1237	1126/ 1011	ZP_00589080.1	Arginine kinase, ArgK [*Chlorobium clathratiforme* DSMZ 5477^T]	4e-50	Identities = 99/145 (68%), Positives = 119/145 (82%)
EpiSH4/ Cag_0616	1038/61941	ZP_00606097.1	Hemagglutinin [*Magnetococcus* sp. MC-1]	7e-06	Identities = 77/351 (21%), Positives = 135/351 (38%)
EpiSH5/ Cag_1990	963/864	ZP_00512181.1	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase, DapD [*Chlorobium limicola* DSMZ 245^T]	2e-54	Identities = 104/126 (82%) Positives = 119/126 (94%)
Cag_1991	963/1749	ZP_00512179.1	Serine Peptidase of S41A family, C-terminal protease [*Chlorobium limicola* DSMZ 245^T]	2e-24	Identities = 65/158 (41%), Positives = 95/158 (60%)
EpiSH6/ Cag_0375	877/1896	ZP_00589495.1	ABC transporter [*Chlorobium clathratiforme* DSMZ 5477^T]	8e-141	Identities = 249/292 (85%), Positives = 269/292 (92%)
EpiSH8/ Cag_1920	**753/11505**	ZP_00660950.1	Hemolysin-type calcium-binding region:Glycosyl hydrolase, BNR repeat; Hemagglutinin [*Prosthecochloris vibrioformis* DSMZ 265]	3e-14	Identities = 62/178 (34%), Positives = 92/178 (51%)
EpiSH9/ Cag_0763	521/3384	ZP_00590336.1	Exodeoxyribonuclease V, γ-subunit, RecC [*Chlorobium clathratiforme* DSMZ 5477^T]	3e-54	Identities = 107/167 (64%), Positives = 130/167 (77%)
EpiSH10/ Cag_0941	238/1269	ZP_00590490.1	Zn-metalloprotease Peptidase M48, Ste24p [*Chlorobium clathratiforme* DSMZ 5477^T]	5e-25	Identities = 50/78 (64%), Positives = 68/78 (87%)
Cag_0614	**-/110418**	YP_866103.1	Hemolysin-type calcium-binding region [*Magnetococcus* sp. MC-1]	1e-102	Identities = 338/1124 (30%), Positives = 563/1124 (50%)
Cag_1919	**-/4581**	YP_345893.1	Glycosyl hydrolase, BNR repeat [*Pseudomonas fluorescens* PfO-1]	5e-111	Identities = 410/1286 (31%), Positives = 575/1286 (44%)
		ZP_00660950.1	Hemolysin-type calcium-binding region:Glycosyl hydrolase, BNR repeat; Hemagglutinin [*Prosthecochloris vibrioformis* DSMZ 265]	2e-68	Identities = 394/1396 (28%), Positives = 598/1396 (42%)

*putative symbiosis genes are indicated in bold; Cag_0614 and Cag_1919 were identified by bioinformatic methods. Positives are amino acid matches that are not identical but count as similar in the BLOSUM62 scoring matrix (e.g. isoleucine and valine).

Candidate symbiosis genes are absent in other free-living green sulfur bacteria

Corresponding to the results of our SSH analysis, no close homologues (identities $\geq$ 35%) of Cag _0614, 0616, 1919 and 1920 could be detected in the 8 available genome sequences of free-living green sulfur bacteria. To investigate the distribution of the candidate symbiosis genes in a larger number of free-living green sulfur bacteria, specific probes targeting ORFs Cag_0616, 1919 and 1920 were generated and used in dot blot hybridisations against genomic DNA of 16 phylogenetically diverse green sulfur bacterial strains as well as the epibiont *Chl. chlorochromatii*. The candidate symbiosis genes could not be detected in the free-living strains (Suppl. Fig. 2). Only a faint hybridization signal was detected for *Chlorobium limicola* DSMZ 245^T with the probe against Cag_1919, however, the corresponding sequence was absent from the genome of this strain. For comparison, we also generated and tested a probe targeting the gene for a putative ABC transporter (clone EpiSH6) of *Chl. chlorochromatii*. Weak hybridisation signals were obtained for all other 16 tested strains (Suppl. Fig. 2D), which indicates the presence of similar gene sequences in free-living green sulfur bacteria and corroborates the higher similarity of the EpiSH6 sequence to that of other green sulfur bacteria which had been determined by the bioinformatic approach (Table 1). As a second control, dot blot hybridisation with a 16S rRNA gene probe specific for *Chl. chlorochromatii* was conducted and yielded much stronger hybridisation signals (Suppl. Fig. 2E) which is commensurate with the high similarity of the 16S rRNA sequences of >90% among the green sulfur bacteria (Overmann and Tuschak, 1997). The above results thus substantiate the unique occurrence of candidate symbiosis genes in *Chl. chlorochromatii* CaD.

Transcription of putative symbiosis genes in the symbiotic and nonsymbiotic state

Initially, a series of Northern blot analyses was conducted to determine the lengths of the transcripts of Cag_1919 and 1920 (Cag_0614 and 0616 being too long for this type of analyses). No signals could be detected with the specific probes, indicating a low abundance of the transcripts. Therefore, a RT-PCR approach was employed to assess the transcription of three putative symbiosis genes. Two different primer sets were employed to study the transcripts of Cag_1919 (Suppl. Tab.1). Because of the unusual size of ORF Cag_1920 and, particularly, of the alleged ORF Cag_0616, three different primer sets were constructed which amplified three different regions of each of the two ORFs. The different primer sets targeted regions 695 - 868, 4650 - 4798 and 9881 - 10057 in ORF Cag_1920 and regions 2436 - 2556, 27378 - 27546 and 53556 - 53697 in ORF Cag_0616 (Suppl. (Table 1). Highly specific PCR conditions were established by using RNA extracts of all 16 nonsymbiotic green sulfur bacterial strains as negative controls. As controls for the contamination of RNA extracts with genomic DNA, the RNA extracts were directly subjected to PCR without prior reverse transcription. As a second control, amplifications with a primer set targeting the *sigA*

gene were performed. The latter method permits a detection of traces of DNA at a significantly higher sensitivity than the PCR targeting 16S rRNA genes.

487 bp-long cDNA fragments of Cag_1919 were obtained from the consortium "*C. aggregatum*" as well as from pure cultures of *Chl. chlorochromatii* (Fig. 1A). Employing primers binding to the 5′- and 3′-ends of Cag_1919 in a long-range RT-PCR analysis, transcripts of the entire ORF Cag_1919 could be detected (Fig. 1B) and were verified by direct sequencing. Also constitutively transcribed were the central part of Cag_1920 as well as the 5′end and central parts of Cag_0616 (data not shown). However, transcripts of the 5′and 3′-regions of Cag_1920 and of the 3′-region of Cag 0616 could not be detected.
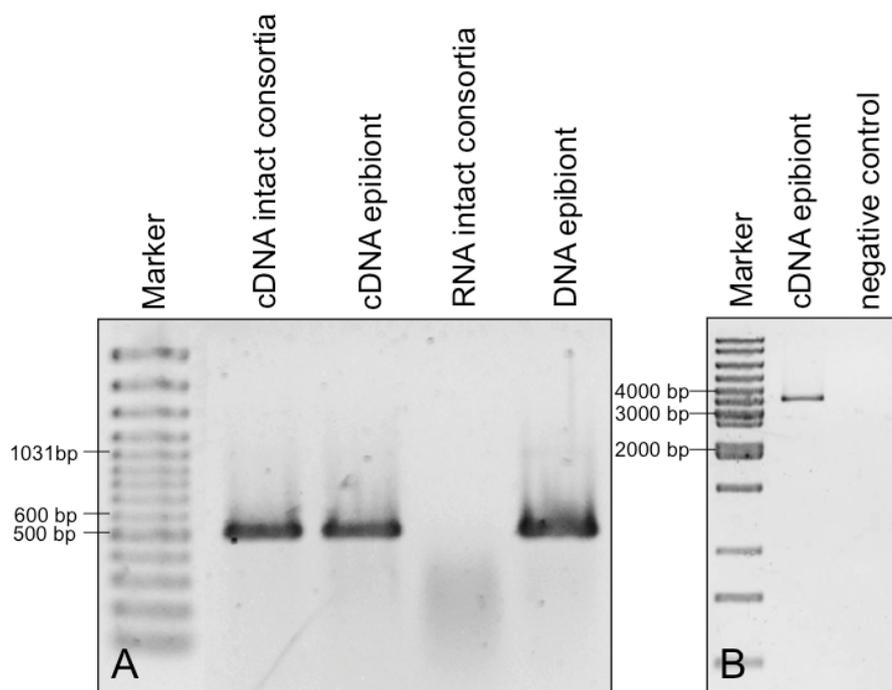
Fig. 1. Detection of transcripts of ORF Cag_1919. **A.** Detection of transcripts of ORF Cag_1919 in pure cultures of *Chlorobium chlorochromatii* and in intact "*Chlorochromatium aggregatum*" consortia. A 487 bp-long PCR product was generated with the primer pair RTX 3797f and RTX 4266r (see Suppl. Table 1). RNA extracts from "*Chlorochromatium aggregatum*" without reverse transcription served as a negative control, and genomic DNA from *Chlorobium chlorochromatii* as a positive control. **B.** Transcript of the almost complete ORF Cag_1919 detected in *Chlorobium chlorochromatii* cultures employing the primer pair RTX502f and RTX 4284r. A PCR without addition of the reverse transcriptase was used as a negative control.

Sequence analysis and modelling of the gene products

Subsequent bioinformatic analyses focussed on the detection of functional modules in Cag_0614, 0616, 1919 and 1920. Motif Scan and Inter Pro Scan detected a hemagglutinin repeat and crystallins beta and gamma Greek key motifs in Cag_0616 (Fig. 2). The two consecutive Greek key motifs contained all established elements of this motif, namely the conserved sequence $(Y/F/W)X_6GX_{28-34}S$ (Ranjini *et al.*, 2001), two putative Ca^{2+} binding sites as well as three or four beta strands (Fig. 2B). Outside of the Greek key motifs, the putative gene product of Cag_0616 harbours three RGD tripeptide motifs known from filamentous hemagglutinin (Relman *et al.*, 1989) plus four different types of repeats (R1´-R4´; Fig. 2A). The first repeat consists of three 711 aa-long sequences, whereas the three downstream repeat regions consist of sequences with lengths of 236, 933 and 877 aa, respectively.

Three sequence regions of Cag_0616 were found to be almost identical to Cag_0614 (Fig. 2A). In contrast to Cag_0616, however, the amino acid sequence of Cag_0614 contains only one RGD tripeptide but harbours seven different repeat regions. Repeat regions 1, 3, 4, 5, and 7 each consisted of two individual repeats with lengths of 1102 aa , 479 aa, 507 aa, 957 aa and 1348 aa. Repeat 2 contains two complete 2072 aa repeats and one incomplete 952 aa fragment. Repeat 6 comprises three 862 aa repeats (Fig. 2A). Amino acid identities between repeats R5 (Cag0614) and R3´ (Cag_0616), and between R6 (Cag_0614) and R4´ (Cag_0616) were 83 and 79%, respectively. A Greek key motif was not detected in the putative gene product of Cag_0614.

Employing the ScanProsite software yielded no hit for ORF Cag_1920. In contrast, the Motif Scan and Inter Pro Scan software identified a bacterial neuraminase repeat (BNR)/Asp box repeat and an inosine-5´-monophosphate (IMP) dehydrogenase/GMP reductase domain.
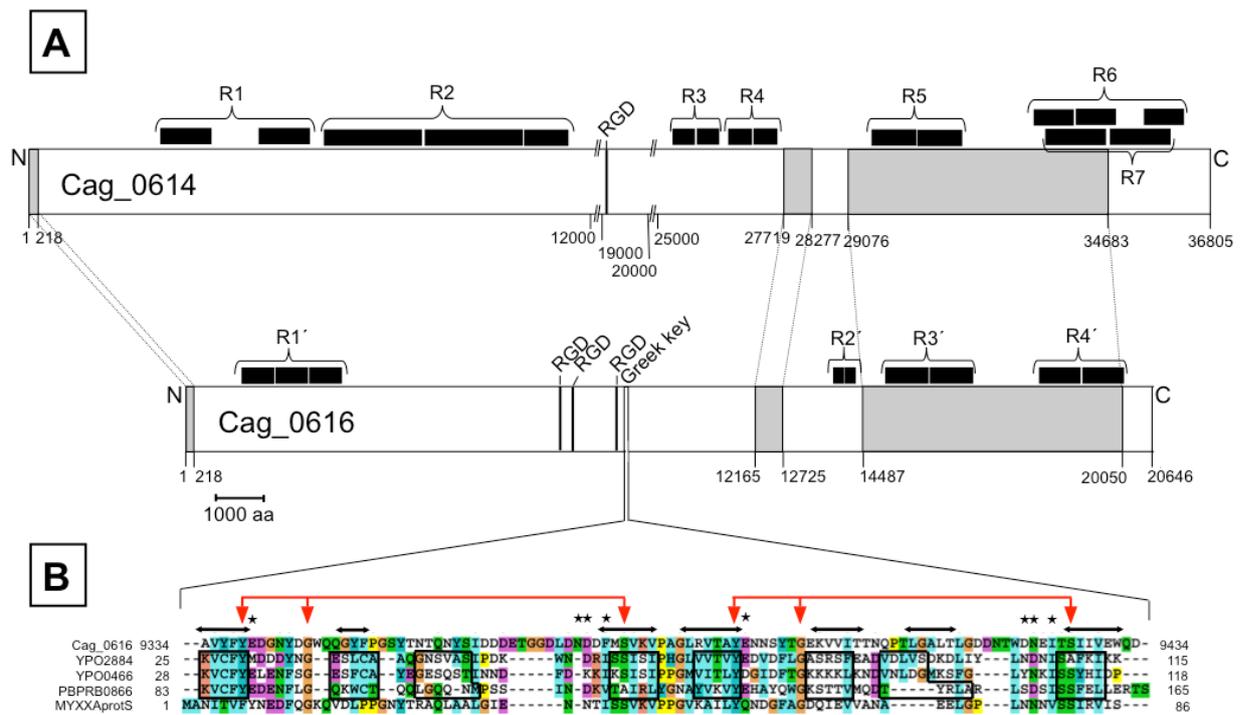
Fig. 2. **A.** Bioinformatic analysis of putative gene products of ORFs Cag_0614 and Cag_0616. Regions of high sequence similarity between the two products are shaded in grey. These regions spanned amino acid (aa) positions 1 - 218, 12165 - 12725 and 14487 - 20050 in Cag_0616 and positions 1 - 218, 27719 – 29076 and 34683 - 36805 in Cag_0614. Black bars denote sequence repeats within each product. In Cag_0614, pairs of repeats were found at positions 2813-3924/4897-5909, 25316-26279, 26549-27579, 29577-31505, and 33134-36000. In Cag_0616, the four repeat regions span nucleotide positions 1194 – 3328, 13913 - 14391, 14974 - 16863 and 18374 – 20129. In addition, the positions of RGD motifs (one in Cag_0614, three in Cag_0616) and of the greek key motifs (in Cag_0616) are indicated. Bar represents 1000 amino acids. **B.** Alignment of the 100 aa-sequence of Cag_0616 containing the two greek key motifs (Y/F/W)X_6GX_{28-34}S (red arrows) with sequences of the hypothetical proteins YPO2884 and YPO0466 of *Yersinia pestis* CO92, PBPRB0866 of *Photobacterium profundum* SS9 and the development-specific protein S of *Myxococcus xanthus* FB. Putative calcium-binding sites (Rajini *et al.*, 2001) are indicated by asterisks. β-strands predicted for Cag_0616 are denoted as double-headed arrows and for comparison are boxed for the *Y. pestis* and *P. profundum* proteins according to Jobby and Sharma (2005).

Within the putative gene product encoded by Cag_1919, a distinct C-terminal hemolysin-type calcium-binding region was identified. Hemolysin-type calcium-binding modules are characterized by multiple repeats of the nonapeptide GGXGXDXLX which include a GGXGXD consensus sequence. The three-dimensional structure of such a Ca^{2+}-binding motif was already reported for the alkaline protease of *Pseudomonas aeruginosa* and features a parallel beta roll structure in which the Ca^{2+} ions are bound at the turns between the two strands (Fig. 3A). Within the turns the Ca^{2+} ions are directly coordinated by the aspartic acid residues of the repeat sequence GGXGXD (Baumann *et*

al., 1993) (Fig. 3A,C). Additional binding sites have been deduced from the crystal structure (Baumann *et al.*, 1993) (Fig. 3A,C). Corresponding to this prediction, modelling of the Cag_1919 gene product yielded a C-terminal beta roll structure (Fig. 3B). Six repeats of the consensus sequence GGXGXD were identified within a 100 amino acid-long region, including an additional motif not found in *Ps. aeruginosa* (Fig. 3C, marked by a brown bar). The entire C-terminal region aligned well with Ca^{2+}-binding domains of different *proteobacterial* proteins, which strongly supports the presence of a parallel beta roll structure in the gene product of Cag_1919 (Fig. 3C).
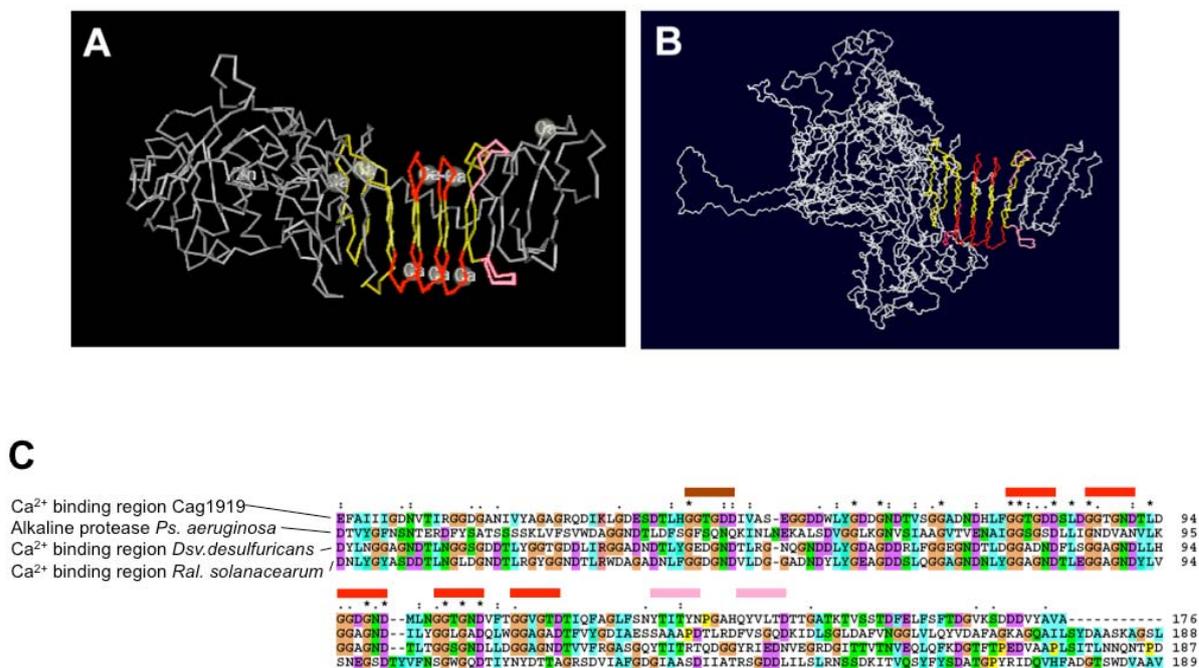
Fig. 3. A. Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa* (Baumann *et al.*, 1993) as available through the structure database of NCBI (Chen *et al.*, 2003) under accession number 1KAP. The structure was displayed with the Cn3D application of the NCBI Entrez retrieval service (http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml). **B.** Three-dimensional model of the gene product of ORF Cag_1919 which at the C-terminus exhibits close similarity to the alkaline protease of *Pseudomonas aeruginosa*. In A. and B., the parallel beta roll is marked in yellow; amino acid residues involved in Ca^{2+} binding are marked in red and pink. **C.** Alignment of the sequences of parallel beta roll motives present in proteins from different proteobacteria. Amino acid residues of the RTX-motif shown to be involved in Ca^{2+} binding in alkaline protease are marked by red bars, Ca^{2+}-binding regions inferred from the crystal structure are denoted by pink bars. An additional Ca^{2+} binding motif which is only found in the putative product of Cag_1919 is indicated by a brown bar. Besides the most closely related sequences of *Dsv. desulfuricans* G20 and *Ral. solanacearum* GMI 1000 (see Fig. 4), the well characterised alkaline protease of *Ps. aeruginosa* is included in the alignment.

No secretion signals known for typical RTX-toxins such as a C-terminal hydrophobic region flanked by glutamic acid or aspartic acid (Economou *et al.*, 1990) or the conserved sequence (E/D)X$_{11}$DX$_{3-5}$(E/D) X$_{14}$E (Sebo and Ladant, 1993) were found in Cag_1919. Analyses using the SignalP (Bendtsen *et al.*, 2004) and Predisi software (Hiller *et al.*, 2004) also did not indicate the presence of other secretion signals. A homolog to the pore-forming region of the adenylate cyclase (CyaA) of *Bordetella pertussis* (Bauche *et al.*, 2006) could not be detected. A hydropathicity plot according to Kyte and Doolittle employing the ProtScale software (Gasteiger *et al.*, 2005) confirmed the absence of transmembrane helices. In contrast to Cag_1919, no matching templates were found during the automated modelling of putative gene products of Cag_0614, 0616 and 1920.

The phylogenetic analysis of the amino acid sequence coded by Cag1919 revealed that the most closely related RTX toxin Ca^{2+}-binding region is present in the deltaproteobacterium *Desulfovibrio desulfuricans* G20 (Fig. 4). Also, most of the other related proteins are RTX toxins and proteins with a hemolysin-type calcium-binding region, and include autotransporter adhesins of *Pseudomonas fluorescens* PfO-1 and *Vibrio vulnificus* YJ016, an iron-regulated protein FrpC from *Vibrio fischeri* ES114, a cable pili-associated 22kDa adhesin AdhA from *Burkholderia cepacia* and a Fat protein of *Synechocystis sp*. PCC6803. A common feature of these proteins is their involvement in adhesion. The Fat protein and the iron-regulated protein FrpC have the ability to bind calcium like RTX toxins and proteins with a hemolysin-type calcium binding region. All related proteins occur in proteobacteria with the exception of the Fat protein of the cyanobacterium *Synechocystis* sp. PCC 6803, which represents the most distantly related amino acid sequence (Fig. 4).
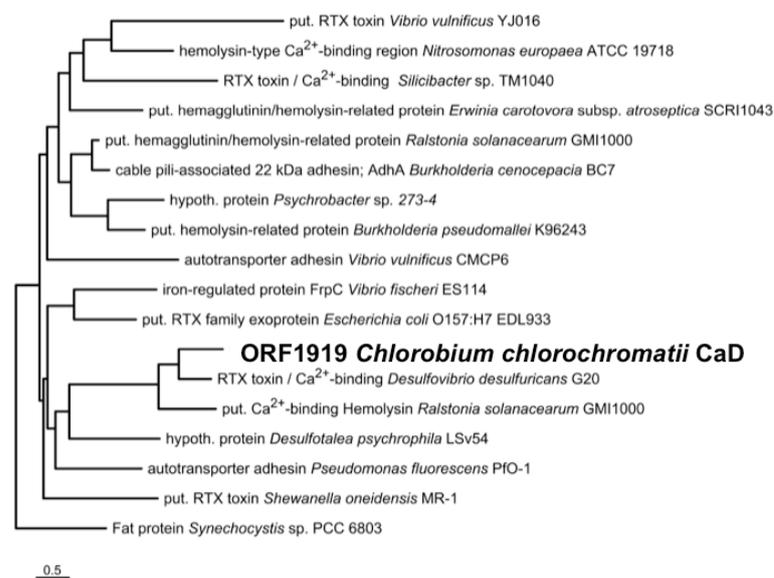
Fig. 4. Phylogenetic analysis of the RTX-module of the C-terminal amino acid sequence of Cag_1919. The analysis is based an alignment of the sequences of other bacteria to the 176 amino acid-long protein sequence of RTX-type motif of *Chlorobium chlorochromatii* (compare Fig. 3C). The bar refers to 0.5 substitutions per 100 amino acid sites.

Detection of Ca^{2+}-binding proteins and role of Ca^{2+} in cell-cell-adhesion

Ca^{2+}-binding proteins in *Chl. chlorochromatii* were studied by ^{45}Ca^{2+}-autoradiography. Within the membrane fraction of the isolated epibiont, three individual protein bands with a molecular mass of about 22, 25 and 29 kD could be detected. In the membrane fraction of the consortium "*C. aggregatum*", four radioactive bands could be detected and matched proteins with an estimated molecular mass of 22, 25, 29 and 34 kD (Fig. 5). No Ca^{2+}-binding proteins were observed in the cytoplasmic fraction of the isolated epibiont *Chl. chlorochromatii* and the consortium "*C. aggregatum*" (Fig 5). Ca^{2+}-binding proteins were also missing in both, the cytoplasmic and membrane fractions of *Cba. tepidum* (Fig 5). As a positive control, the purified adenylate cyclase from *Bordetella pertussis* which contains a beta roll motif could also detected by this method (data not shown). It is highly unlikely that the ^{45}Ca^{2+}-binding bands contained outer membrane lipopolysaccharides since (1) the majority of lipids were extracted in the acetone precipitation step during protein preparation and (2) no signals could be obtained when purified LPS of *Escherichia coli* B or *Serratia marcescens* (Sigma L-6136) were run on the SDS gels (data not shown). Selected bands of Ca2+-binding proteins were subsequently analysed by mass spectroscopy (Fig. 5, arrowheads). The three bands of the epibiont membrane fraction contained 30S ribosomal proteins (S2, S3 and S4), 50S ribosomal proteins (L1, L3 and L5), an ATP-dependent Clp protease and a plsX-type fatty acid/phospholipid synthesis protein. Bands of the consortia membrane fraction also contained 30S ribosomal proteins (S3 and S4) as well as the 50S ribosomal protein L1, a nitrogenase iron protein and a bacteriochlorophyll a protein.
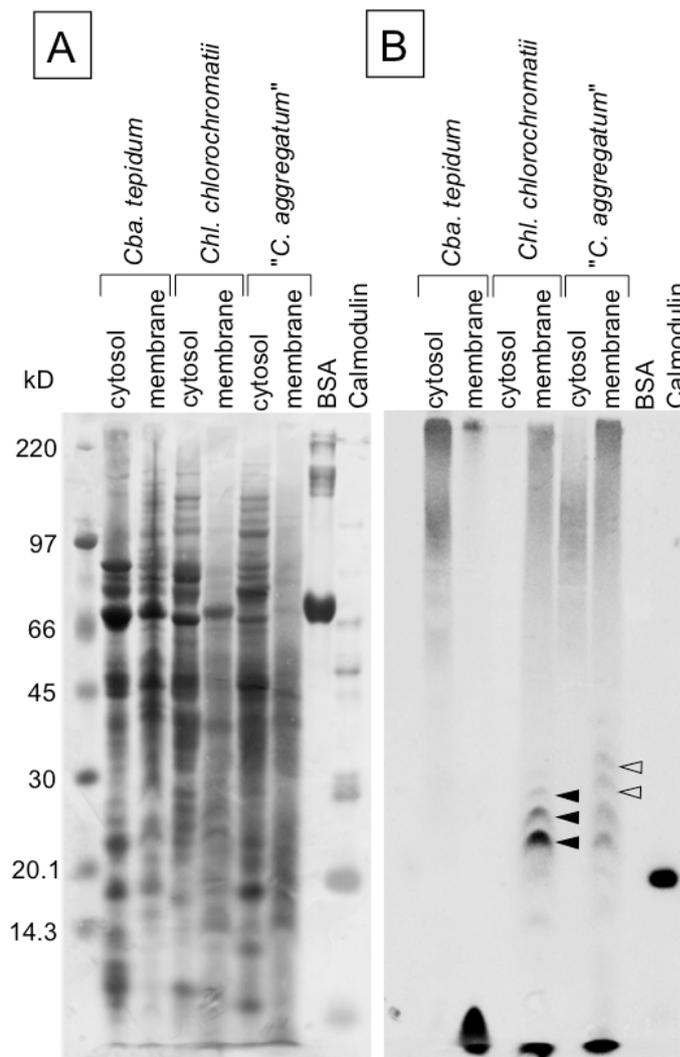
Fig. 5. Ca^{2+} binding assay for cytoplasmic and cell wall proteins from *Cba. tepidum*, *Chl. chlorochromatii* and "*Chlorochromatium aggregatum*". Calmodulin served as positive and BSA as negative control. Proteins were separated by SDS-PAGE, electrotransferred to a polyvinylidene difluoride membrane and incubated with ^{45}Ca. **A.** PVDF membrane stained with amido black. **B.** Autoradiograph. Solid arrowheads denote membrane proteins of *Chl. chlorochromatii*, hollow arrowheads membrane proteins of the consortium "*C. aggregatum*" binding $^{45}Ca^{2+}$ which were sequenced.

The role of Ca^{2+} in the phototrophic consortium "*C. aggregatum*" was investigated further by monitoring the effects of different chemicals on aggregate stability and motility of the consortia. The impact of the redox potential was assessed with reducing agents (ascorbate, L-cysteine, sodium sulfite), and the contribution of lectins to the adhesion by adding a suite of sugars and sugar derivatives (D-(+)-mannose, L-rhamnose, D-galactose, glucose, fucose, arabinose, *N*-acetylglucosamine, glucuronic acid, galacturonic acid, methyl glucoside, methyl mannoside). Also, a variety of ionic and nonionic detergents (SDS, Tween20, Tween80, Triton X-100) and denaturing agents (urea, formamid) was applied. Finally, the testing scheme comprised several hydrolytic

enzymes (hyaluronidase, β-glucuronidase, proteinase K, pepsin, trypsin, chymotrypsin, lysozyme) and complexing agents (EDTA, EGTA, pyrophosphate, citrate, tartrate).

Of all compounds, only Ca^{2+}-complexing agents (EGTA, EDTA, pyrophosphate) and the strong detergents SDS and TritonX-100 exerted a significant effect on cell-cell-binding in "*C. aggregatum*" (Fig. 6). The effects of EGTA, EDTA and pyrophosphate depended on the concentration applied. While more than 70% of the consortia remained intact after 10 min of exposure to 2 mM concentrations of these agents, almost all consortia disaggregated in the presence of 20 mM EGTA. Disaggregation by pyrophosphate was slower compared with that by EDTA and EGTA (Fig. 6). While ascorbate and sulfite at concentrations of 20 mM inhibited the motility of the consortia, they did not affect cell-cell-binding, however. None of the sugar compounds or the denaturing agents exerted any effect (shown for arabinose in Fig. 6). Similarly, no effect was observed for Tween 20 and Tween 80, whereas the stronger detergents SDS and Triton X-100 led to complete immotility and a partial but significant disaggregation of the consortia. Of the hydrolytic enzymes, lysozyme, hyaluronidase, β-glucuronidase, trypsin and chymotrypsin affected motility, but not cell-cell-binding of the consortia (Fig. 6).
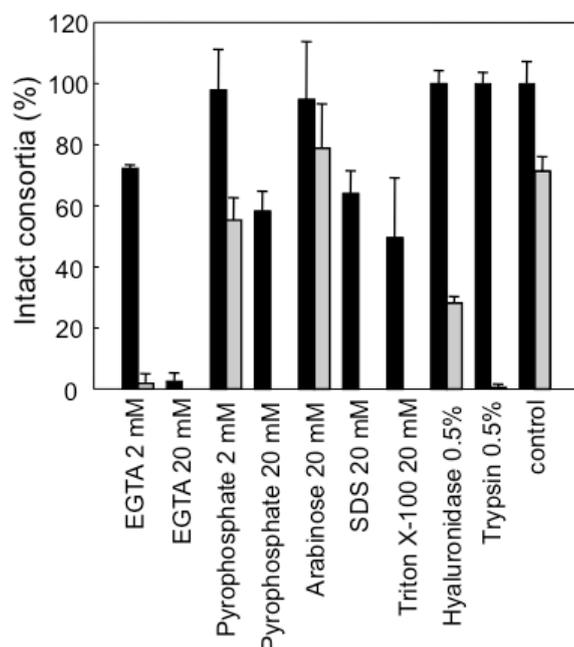
Fig. 6. Percentage of intact (■) and motile (▨) "*Chlorochromatium aggregatum*" consortia remaining after 10 min of incubation in the presence of different chemical agents (final concentrations are given). Vertical bars indicate one standard deviation.

Discussion

Occurrence and transcription of candidate symbiosis genes in Chl. chlorochromatii

The adaptation of bacteria to their particular ecological niches has been traced back to genomic differences in either single genes or gene clusters, or to the differential expression of common genes (Perna *et al.*, 2001; Read *et al.*, 2003; Jones *et al.*, 2006). A well-studied case is the evolution of high- and low-light adapted lineages of the marine genus *Prochlorococcus* leading to hundreds of lineage-specific genes (Rocap *et al.*, 2003). Based on the unique life style and considerable phylogenetic distance of *Chl. chlorochromatii* CaD to other green sulfur bacteria, niche-specific genes were expected to occur in the genome of the epibiont.

So far, identification of niche-specific genes in other bacterial groups has mostly been performed by comparing two or three bacterial genomes (Perna *et al.*, 2001; Jones *et al.*, 2003; Read *et al.*, 2003; van Ham *et al.*, 2003). In order to reliably identify putative symbiosis-specific genes in the epibiont genome, our suppression subtractive hybridisation (SSH) included 16 free-living strains of green sulfur bacteria of different physiology (Overmann, 2001b) which appeared promising since the members of this group are phylogenetically rather closely related (Overmann and Tuschak, 1997). Our results demonstrate that such a mixture of genomic DNA from a larger number of related bacterial strains can be employed as driver to recover genes which are unique to a single tester strain. The unique nature of the putative symbiosis-specific ORFs Cag_0614, 0616, 1919 and 1920 were fully confirmed by subsequent *in silico* analyses of the eight available genome sequences of green sulfur bacteria. Apart from the putative symbiosis genes, SSH retrieved several gene fragments with similarity to functional genes of free-living green sulfur bacteria. This result is explained by the fact that functional genes of green sulfur bacteria exhibit a larger sequence divergence than the 16S rRNA genes (Figueras *et al.*, 2002), as exemplified by the weak dot blot hybridisation of the ABC transporter in Fig. 2D. If sufficiently different, such functional genes will not be depleted during SSH.

It has been speculated that giant genes are expressed during phases of slow growth or in resting cells since synthesis of the large proteins would require rather long time intervals (Reva and Tümmler, 2008). Interestingly, the transcription of the putative symbiosis genes of *Chl. chlorochromatii* is constitutive and does not seem to be regulated by the symbiotic interaction with the *Betaproteobacterium* since our highly specific RT-PCR approach detected transcripts of ORF Cag_0616, Cag_1919 and Cag_1920 in pure (i.e. nonsymbiotic) cultures of the epibiont *Chl. chlorochromatii* as well as in intact "*C. aggregatum*" consortia. Therefore, either the regulation occurs on the posttranscriptional level or expression of the ORFs is not regulated at all. Epibionts seem to be specifically adapted to the life in association with the central bacterium and so far have not been detected in a free-living state in nature (Glaeser and Overmann, 2004). Therefore, a

regulation mechanism for the expression of the three potential symbiosis genes may actually be dispensable.

Cag_0614, Cag_0616 and Cag_1920

ORFs Cag_0614 and 0616 show similarity to a putative hemagglutinin and contained numerous internal repeats. The high sequence similarity and similar structure of Cag_0614 and Cag_0616 suggests that these two ORFs arose through a gene duplication event. Spanning 110,418 and 61,938 bp, respectively, the Cag_0614 and Cag_0616 represent the largest open reading frames known to date and are only surpassed by the human titin gene whose exons together span 114,414 bp. This is confirmed by a recent bioinformatic analysis of 580 available bacterial and archaeal genomes (Reva and Tümmler, 2008). Open reading frames of a size similar to Cag_0614 and Cag_0616 have only been found in the cyanobacterium *Synechococcus* sp. RS9917 (RS9917_01402; 84,534 bp) and *Clostridium botulinum* ATCC 3502 (NC_009495; 57,980 bp). Giant genes of > 20kb in length occur in 62 of the genome sequences and mostly in environmental rather than pathogenic bacteria. They typically encode either polyketide/non-ribosomal peptide synthetases or putative surface proteins.

Contiguous repeats of several hundred amino acids are known for other hemagglutinin-like proteins (Ward *et al.*, 1998). In addition, Cag_0616 codes three arginyl-glycyl-aspartic acid tripeptides, and one Greek key motif. The RGD motif occurs in proteins (e.g., fibronectin) of the extracellular matrix of mammalian cells, in toxins of plant pathogenic fungi or in surface proteins of certain animal viruses, and mediates adhesion of cell surface receptors (Ruoslathi and Pierschbacher, 1986; Isberg and Tran Van Nhieu, 1994; Tan *et al.*, 2001; Senchou *et al.*, 2004). In prokaryotes, the RGD motif is present in the integrin-binding proteins of pathogens like *Bordetella pertussis* that attach to mammalian cells (Isberg and Tran Van Nhieu, 1994; Kajava *et al.*, 2001). Based on their frequent involvement in host-pathogen-interactions, the three RGD tripeptides detected in ORF Cag_0616 may participate in the cell-cell-binding of phototrophic consortia.

The Greek key motif is composed of four antiparallel beta strands and occurs as duplicate motif in vertebrate proteins of the βγ-crystallin superfamily. In bacteria, the motif was found in the spore coat protein S of *Myxococcus xanthus*, a metalloprotease inhibitor of *Streptomyces* and an extracellular protein of *Yersinia pestis* (Wistow, 1990; Rajini *et al.*, 2001; Jobby and Sharma, 2005). These bacterial proteins have been assumed to participate in the response to stress conditions. Among the Green Sulfur Bacteria, Cag_0616 occurs only in the symbiotic *Chl. chlorochromatii*, which indicates that the βγ-crystallin-type gene product is involved in the symbiotic interaction. Duplicate Greek key motifs have been shown to bind two calcium ions

(Rajini *et al.*, 2001). Similarly, the gene product of Cag_0616 may be stabilised by binding of Ca^{2+} ions.

Since the protein sequence with the closest similarity to Cag_0616 was a putative hemagglutinin, we searched for additional properties of the putative gene product of Cag_0616. The filamentous hemagglutinin adhesin (coded by *fhaB*) of *Bordetella pertussis* contains a binding site for sulfated glycolipids and a carbohydrate recognition domain besides its two RGD motifs (Kajava *et al.*, 2001). Adhesion mediated through the lectine-like activity can be partially blocked by galactose (Isberg and Tran Van Nhieu, 1994). On contrast, lectines do not participate in the cell-cell-interaction in "*C. aggregatum*" based on the results of our disaggregation studies.

Known hemagglutins, like the products of the 10,774 bp-long *fhaB* of *B. pertussis* (Domenighini *et al.*, 1990) or of the 12,500 and 14,800 bp-long *lspA1* and *lspA2* of *Haemophilus ducreyi* (Ward *et al.*, 1998) have been shown to be post-translational processed. Therefore, only parts of ORFs Cag_0614 and 0616 may actually be expressed in the epibiont of phototrophic consortia.

The analysis of ORF Cag_1920 revealed the presence of a bacterial neuraminase repeat (BNR)/Asp box repeat which has been found in more than nine non-homologous protein families, including bacterial ribonucleases, sulfite oxidases, reelin, netrins, some lipoprotein receptors and a variety of glycosyl hydrolases. So far, few experimental data are available concerning the general functions of Asp boxes (Copley *et al.*, 2001).

Cag_1919

The C-terminal hemolysin-type Ca^{2+}-binding region coded by Cag_1919 contains several RTX repeats and is highly similar to the RTX-region in alkaline protease of *Ps. aeruginosa*. RTX toxins are typically found in Gram-negative pathogenic bacteria and are characterized by repetitive nonapeptide motifs, which include a GGXGXD consensus (Welch, 1995). The RTX module is necessary for binding to the target cell and without loss of this function can be separated from the other, lytic (as in the *Pasteurella* leukotoxin; Cruz *et al.*, 1990) or catalytic domains (as in adenylate cyclase of *B. pertussis*; El-Azami-El-Idrissi *et al.*, 2003). The binding of RTX toxins of pathogenic bacteria to the target cell involves Ca^{2+} ions (Ludwig *et al.*, 1988; Knapp *et al.*, 2003) which are bound at the GGXGXDXLX repeats with low affinity (Rose *et al.*, 1995; Lilie *et al.*, 2000). Our sequence comparisons and 3D-modeling strongly indicates that the Cag_1919 gene product forms a C-terminal beta roll which represents a *bona fide* Ca^{2+} binding structure. Based on the results of our disaggregation studies, it appears feasible that the RTX domain encoded by Cag_1919 of *Chl. chlorochromatii* is involved in the cell-cell-adhesion between the partner bacteria of phototrophic consortia.

RT-PCR analyses of Cag_1919 revealed that this ORF is transcribed over its entire length. The corresponding protein is expected to have a molecular mass of 155 kD, but was not detected by $^{45}Ca^{2+}$ autoradiography on SDS gels. Posttranslational processing of RTX-type proteins of other Gram-negative bacteria has been demonstrated (Osicka *et al.*, 2004), but fragments of the Cag_1919 gene product were also not detected in bands of smaller Ca^{2+}-binding proteins. Instead, each of these band contained at least one ribosomal protein. Membrane-bound polysomes are known to occur in *Escherichia coli* (Smith *et al.*, 1978) and other eubacteria. However, Ca^{2+}-binding to ribosomal proteins as observed in our study has so far not been demonstrated. The detection limit of the autoradiography method is 2 μg of calcium binding protein, corresponding to 3% of the membrane protein (Maruyama *et al.*, 1984). Since the Cag_1919 protein could not be detected by Ca^{2+}-binding, this protein may thus occur in the epibiont at a rather lower abundance.

Cag_1919 seems not to be secreted by one of the mechanisms known for the RTX-type toxins like for instance the adenylate cyclase of *B. pertussis* (Bauche *et al.*, 2006). Also, pore formation by hydrophobic domains as known from adenylate cyclase is highly unlikely since no transmembrane helices were detected in the Cag_1919 gene product.

Origin of symbiosis genes of the epibiont

Obligately intracellular bacterial symbionts and pathogens are characterised by a reductive evolution of their genome (Cole *et al.*, 2001; van Ham *et al.*, 2003). In contrast, the genome size of epibiont *Chl. chlorochromatii* CaD (2.57 Mb) falls well within the size range of all sequenced green sulfur bacterial genomes (1.97 – 4.44 Mb; http://genome.jgi-psf.org/mic_home.html), which is commensurate with the capacity of the epibiont to grow also independently in the free-living state. An *in silico* subtractive hybridization analysis of the available genome sequences of green sulfur bacteria identified 188 additional ORFs to be unique for *Chl. chlorochromatii* CaD (data not shown), most of them coding for hypothetical proteins without any homology to sequences in free-living green sulfur bacteria. This number is considerably lower than the numbers of niche-specific genes in high-light-adapted (364 ORFs) and low-light-adapted *Prochlorococcus* strains (923 ORFs) (Rocap *et al.*, 2003) and indicates that the adaptation to the symbiosis in phototrophic consortia does not require a large number of additional genes.

Most remarkably, the four putative symbiosis genes identified in the present study showed similarity to different types of virulence factors of typical bacterial pathogens. Genes underlying the adaptation of different evolutionary lineages are either differentially retained from the common ancestor, acquired through gene duplication and divergent evolution, or laterally transferred to an individual lineage from distantly related prokaryotes (Rocap *et al.*, 2003). The four symbiosis-specific ORFs of *Chl. chlorochromatii* lack some of the properties thought to be characteristic for

horizontally transmitted genes (Lawrence and Roth, 1996; Lawrence and Ochman, 1998; Jones *et al.*, 2006), like (i) neighbouring tRNA genes which have been used to localise insertion events, (ii) a different G+C content (symbiosis genes: 42% - 46% GC, *Chl. chlorochromatii* genome average: 44.3 mol% GC), or (iii) a codon usage which differs from that of the entire genome (data not shown).

In contrast, our phylogenetic analysis provided evidence for the origin of the putative symbiosis gene Cag_1919. While the putative gene products of ORFs Cag_0614, 0616 and 1920 were only very distantly related to known proteins and could not be phylogenetically analysed further, the pronounced similarity of the Cag_1919 RTX-domain to other amino acid sequences permitted a detailed phylogenetic analysis. The closest relatives are Ca^{2+}-binding proteins from *Gamma*- and *Deltaproteobacteria*, indicating that *Chl. chlorochromatii* acquired at least the RTX-module from proteobacteria via a horizontal gene transfer event. This is further corroborated by the presence of a transposase (Cag_1918) in close proximity to Cag_1919 and 1920 which also suggests that both ORFs together were laterally transferred during the same event.

In conclusion, the different lines of experimental evidence gathered in the present study provide the first indication that genetic modules known from proteobacterial pathogens of eukaryotes have been laterally transferred to nonrelated bacteria and are employed in symbiotic interactions between different species of prokaryotes.

Experimental Procedures

Bacterial strains

Cultures of green sulfur bacteria were grown in standard SL10 medium (Overmann and Pfennig, 1989) supplemented with 3 mM acetate. For *Chlorobium chlorochromatii* the pH was set to 7.2. It was adjusted to 6.8 in media for *Chlorobium* strains H1D and `Fjord´, *Chl. phaeobacteroides* DSMZ 266[T], *Chl. limicola* E3P1, *Chl. phaeobacteroides* CL 1402, *Chl. phaeobacteroides* brChl, *Chl. phaeobacteroides* DagIII, *Chlorobaculum thiosulfatiphilum* DSMZ 249[T], *Chl. clathratiforme* DSMZ 5477[T], *Chl. limicola* DSMZ 245[T] and *Chl. phaeobacteroides* E2P3. *Chl. phaeovibrioides* DSMZ 269[T], *Chl. luteolum* DSMZ 273[T], *Prosthecochloris vibrioformis* DSMZ 260[T] and *Ptc. aestuarii* DSMZ 271[T] were grown at pH 6.8, 2% (w/v) NaCl and 0.3% (w/v) $MgCl_2 \cdot 6H_2O$. Cultures were incubated at 25°C at 50 µmol quanta·m^{-2}·s^{-1} of continuous illumination (tungsten lamp bulb; Osram 60 W). Light intensity was determined with a Li Cor LI-189 quantum meter equipped with a LI-200 SA pyranometer sensor (Li Cor, Lincoln, Neb., USA). *Chlorobaculum tepidum* ATCC 49652[T] was grown in CL medium (Frigaard and Bryant, 2001), at 46°C and 1000 µmol quanta·m^{-2}·s^{-1}. The consortium "*C. aggregatum*" was grown in K4 medium (Kanzler *et al.*, 2005) in 10 l glass flasks at 15°C and 20 µmol quanta·m^{-2}·s^{-1}. Under these conditions, "*C. aggregatum*" forms an

almost pure biofilm on the inner surface of the vessel which can be harvested separately (Pfannes *et al.*, 2007).

Retrieval of putative symbiosis genes by subtractive hybridisation

Genomic DNA was prepared with the DNeasyTissue Kit (Qiagen, Hilden, Germany) according to a modified protocol. Twice the amount of proteinase K was added to the samples and cell lysis was performed at 55°C for 4 h. After elution, the DNA was diluted in 2 mM Tris (pH 7.0), purified by ultrafiltration in Centricon Ultracel YM-50 Ultrafiltration units (Millipore, Schwalbach, Germany) and quantified by fluorescent dye binding with PicoGreen (MoBiTec, Göttingen, Germany).

Suppression subtractive hybridisation (Diatchenko *et al.*, 1996; Gurskaya *et al.*, 1996) was carried out employing the CLONTECH PCR-Select^{TM} Bacterial Genome Subtraction Kit (BD Biosciences Clontech, Heidelberg, Germany). As tester DNA, 2 μg of *Chl. chlorochromatii* strain CaD was used. As driver, 2 μg of DNA each of the 16 free-living green sulfur bacteria (cf the section *Bacterial strains*) were employed. The tester DNA was split into two samples and the DNA of each sample was ligated to one specific set adaptors. Then, 1 μl of each sample was hybridised separately with 2 μl of each driver DNA at 63°C for 1.5 h. Both hybridisation reactions were combined, 1 μl of driver DNA was added and the mixture was hybridised again at 63°C for 16 h. The resulting second hybridisation mixture was then diluted to a total volume of 200 μl and amplified in two consecutive steps in a GeneAmp PCR System 9700 (Applied Biosystems, Weiterstadt, Germany) employing the BD Advantage^{TM} 2PCR Enzyme System (BD Biosciences Clontech) with primers complementary to the adaptor sequences. The first PCR comprised a 2 min incubation at 72°C to extend the adaptors, then an initial denaturation (95°C, 0.5 min), followed by 30 cycles with denaturation at 94°C for 0.5 min, annealing at 66°C for 0.5 min and primer extension at 72°C for 1.5 min. The second PCR comprised an initial denaturation step at 95°C for 1 min, followed by 15 thermal cycles with denaturation at 94°C for 0.5 min, primer annealing at 68°C for 0.5 min, elongation at 72°C for 1.5 min and a final extension at 72°C for 10 min.

Polymerase chain reaction products were ligated into vector pCR®2.1-TOPO and cloned through chemical transformation using the TOPO TA cloning kit (version R; Invitrogen, Carlsbad, CA). Plasmids were extracted with a QIAprep spin miniprep (Qiagen), and the presence of inserts was verified by digestion with *Eco*RI (MBI Fermentas, St. Leon-Rot, Germany). Nucleotide sequence data were obtained with a ABI Prism 310 genetic analyser (Applied Biosystems), employing the AmpliTaq FS Big Dye Terminator cycle sequencing kit and M13 forward and M13 reverse primers.

Sequence analysis and modelling of three-dimensional structures

The genome sequence of *Chl. chlorochromatii* was recently determined by the Joint-Genome Institute (Department of Energy, USA; http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) and annotation of the data is in progress (D. Bryant and J. Overmann, unpublished data). Based on this genome sequence, the open reading frames (ORFs) corresponding to the gene fragments recovered by subtractive hybridisation could be identified. Similarities to known sequences were assessed by BLAST searches (Altschul *et al.*, 1997) using the BLAST X algorithmus. Subsequent analyses were conducted with the InterProScan (Zdobnow and Apweiler, 2001), ScanProsite (Gattiker *et al.*, 2002) and Motif Scan (Falquet *et al.*, 2002) software packages. Conserved sequence motifs were identified with 3of5 complex pattern search (Seiler *et al.*, 2006), the secondary structure was analysed with PredictProtein (Rost *et al.*, 2004) and repeats within the protein were identified with the REPRO (George and Heringa, 2000) software. Secretion signals were searched with SignalP 3.0 (Bendtsen *et al.*, 2004) and PrediSi (Hiller *et al.*, 2004) software. Hydropathicity plots were obtained with the ProtScale software (Gasteiger *et al.*, 2005).

For sequence comparisons and phylogenetic analyses, related protein sequences were recovered from the Genbank database (Benson *et al.*, 2002) and aligned with CLUSTAL X version 1.8 (Thompson *et al.*, 1997). Phylogenetic trees were calculated with the PROTML program of the Phylogeny Inference Package (PHYLIP Version 3.6.3) (Felsenstein, 2002), employing the maximum likelihood algorithm and the Dayhoff PAM probability model.

In order to recover the appropriate modelling template for ORF Cag_1919, the protein sequence was submitted to the Ex-PDB database (http://swissmodel.expasy.org/) and a SWISS-Model BLAST search was performed. The modelling template and the query sequence were aligned with Clustal X version 1.8. Using the SWISS-MODEL interface in the alignment mode, the alignment was submitted to the SWISS-Model Expert Protein Analysis System (ExPASy) web server (Peitsch, 1995, Schwede *et al.*, 2003). After return, the model was displayed with the DeepView-Swiss-PdbViewer (Guex and Peitsch, 1997).

Dot blot hybridisation

Probes for the detection of the putative symbiosis genes were generated from the plasmids carrying the respective inserts as obtained by subtractive hybridisation. Probes were randomly PCR labelled with digoxigenin (DIG)-11-dUTP (PCR DIG probe synthesis kit; Roche, Mannheim, Germany), employing the nested primers from CLONTECH PCR-Select[TM] Bacterial Genome Subtraction Kit in a step-down PCR. After an initial 5 min denaturing step at 95°C, 10 cycles were conducted which comprised melting at 94°C for 0.5 min, annealing at 60°C for 0.75 min and extension at 72°C for 1 min, and were followed by 25 cycles with the annealing temperature changed to 55°C. Final

extension proceeded at 72°C for 10 min. The probe targeting the ORF Cag1919 was generated with the same PCR program using the primers RTX 3797f and 4266r (Suppl. Table 1) but setting the annealing temperatures to 70°C and 65°C, respectively.

For blotting, 10 ng of genomic DNA of each green sulfur bacterium was denatured for 10 min at 95°C and vacuum blotted onto positively charged nylon membrane (Hybond N+; Amersham, Freiburg, Germany). The membrane was baked at 120°C for 30 min and prehybridised in 10 ml of DIG Easy Hyb buffer (Roche) at 40°C for 1 h. Hybridisation was carried out for 16 h at 40°C in 10 ml Easy Hyb buffer after adding the denatured probe. After hybridisation, the blot was washed twice for 15 min at room temperature in 2x SSC (1x SSC is 150 mM NaCl plus 15 mM sodium citrate [pH 7.0]) containing 0.1 % SDS, followed by three stringent washing steps (twice in 1x SSC plus 0.1% SDS, once in 0.5x SSC plus 0.1 % SDS; each at 40°C). The hybridisation signal was detected with the DIG luminescence detection kit (Roche) and X-ray film (WICO Rex+; Linhardt Röntgenbedarf, Munich, Germany) according to the instructions of the manufacturer.

RNA extraction and RT-PCR

RNA was extracted from *Chl. chlorochromatii* and "*C. aggregatum*" using the RNeasy Mini Kit (Qiagen). After eluting the RNA from the purification columns, the eluates were treated twice with 20 units of RNase-free DNase I (10 U/µl; Roche; overnight at room temperature) in DNase I buffer (10 mM HEPES, 2.5 mM $MgCl_2$, 0.1 mM $CaCl_2$ [pH 7.5] incubated with 0.1 % diethylpyrocarbonate overnight and autoclaved). Eluates were subsequently purified with RNeasy columns (Qiagen) according to the instructions of the manufacturer. RNA concentrations were determined spectrophotometrically at 260 nm. As a highly sensitive test for contamination with genomic DNA, a step-down PCR with a primer set targeting the *sigA* gene (Suppl. Table 1) was performed using 1 µM of the RNA preparation.

Reverse transcription was performed in 20 µl with SuperScript^{TM} III Reverse Transcriptase (Invitrogen) using 800 ng RNA and 183 ng random hexamer primer as recommended by the supplier. The generated cDNA was then amplified with custom designed gene-specific primers at optimized PCR conditions (Suppl. Table 1). Fifty nanograms of genomic DNA of *Chl. chlorochromatii* served as positive control and 800 ng of total RNA without reverse transcription of *Chl. chlorochromatii* as negative PCR control. Full-length transcripts of ORF Cag_1919 were detected by reverse transcription in 20 µl employing LongRange Reverse Transcriptase (Qiagen), 700 ng RNA and 183 ng random hexamer primer. cDNA was then treated with 2U RNAse H (Roche Diagnostics GmbH, Mannheim, Germany) to remove RNA-cDNA-Hybrids. A long range PCR amplification with custom designed gene-specific primers RTX 502f and RTX 4284r

(Supplementary Table 1) followed. A reaction mix without addition of the reverse transcriptase served as a negative control. Finally, the amplification products were sequenced employing primers RTX 502f and RTX 4284r.

Kinetics of disaggregation

The disaggregation of phototrophic consortia upon exposure to different chemical agents was observed in a microscopic chamber. The chamber consisted of a microscopic slide (26 x 76 mm) and a micro cover slip (24 x 60 mm) spaced apart by 0.1 mm. After sealing the chamber with paraffin at three sides, it was filled with an enrichment culture of "*C. aggregatum*" to which different test compounds had been added from anoxic stock solutions. The fourth edge of the chamber was subsequently sealed and the effect of the substances on cell-cell-adhesion in consortia and on their motility was examined by phase contrast microscopy.

Detection of calcium-binding proteins and mass spectrometry

Cells were harvested at 10,000 x g for 30 min, resuspended in 10 mM Tris buffer pH 7.5 containing 1 mM phenylmethanesulfonyl fluoride (PMSF), and broken by three subsequent passages through a French press cell at 16,000 lb in^{-2}. The homogenate was clarified by centrifugation at 20,000 x g for 30 min at 4°C. The membrane fraction was pelleted by centrifugation at 200,000 x g for 1 h and then solubilised with 2% SDS. The extract was centrifuged again at 200,000 x g for 1 h and proteins in the supernatants were precipitated by adding 9 volumes of acetone and incubation at 0°C for 16 h. After collecting membrane proteins by centrifugation at 20,000 x g for 30 min, the pellets were washed once with acetone and resuspended in 10 mM Tris buffer pH 7.5 containing 1 mM PMSF and 1% SDS. Protein concentrations were estimated using the BCA Protein Assay Reagent (Pierce Biotechnology, Rockford, IL, USA) (Smith *et al.*, 1985). 75 µg of each protein fraction, 15 µg of BSA, 15 µg calmodulin (Sigma-Aldrich, Taufkirchen, Germany) and 20 µg of protein marker (high-range rainbow molecular weight marker; Amersham, Buckinghamshire, England) were separated by Tricine-SDS polyacrylamide gel electrophoresis in 8% acrylamide gels (Schägger and von Jagow, 1987). To determine their Ca^{2+}-binding ability, the proteins were blotted onto a Porablot PVDF membrane (Macherey–Nagel, Düren, Germany). The detection of the calcium binding proteins by ^{45}Ca autoradiography was performed according to the method of Maruyama *et al.* (1984) with an extended incubation time of 30 min. Protein spots of interest were excised manually from one-dimensional gels and digested with modified trypsin (Promega, Heidelberg, Germany) using OMX-S (OMX GmbH, Wessling, Germany) (Granvogl *et al.*, 2007). All samples were analysed by nano-LC-ESI-MS/MS on a quadrupole time-of-flight tandem mass spectrometer (Micromass ESI Q-TOF Premier, Waters Ltd. Manchester, United Kingdom). Separation of peptide

mixtures was achieved by nanoAcquity Ultra Performance Liquid Chromatography using a 1.7 µm BEH130 C18, 75 µm x 100 mm reversed phase nano column (Waters). Peptides were eluted by running a gradient using 100% H_2O / 0.1% formic acid as solvent A and 100% acetonitrile / 0.1% formic acid as solvent B, at a flow rate of 0.4 µl·min^{-1}. The MS and MS/MS spectra were used to search the latest version of the SwissProt data base using ProteinLynx Global Server (Waters).

Acknowledgements

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Miller, W., and Lipmann, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402

Bauche, C., Chenal, A., Knapp, O., Bodenreider, C., Benz, R., Chaffotte, A., and Ladant, D. (2006) Stuctural and functional characterization of an essential RTX subdomain of *Bordetella pertussis* adenylate cyclase toxin. *J Biol Chem* 281: 16914-16926

Baumann, U., Wu, S., Flaherty, K.M., and McKay, D.B. (1993) Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J* 12: 3357-3364

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340: 783-795

Benson, D.A., Karsch-Mizrachi, I., Lipmann, D.J., Ostell, J., Rapp, B.A., and Wheeler, D.L. (2002) Gen Bank. *Nucleic Acids Res* 30: 17-20

Chen, J., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I. , Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler-Bauer, A., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Rao, B.S., Panchenko, A.R., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J., and Bryant, S.H. (2003) MMDB: Entrez's 3D-structure database. *Nucleic Acids Res* 31: 474-477

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R. et al. (2001) Massive gene decay in the leprosy bacillus. *Nature* 409: 1007-1011

Copley, R.R., Russell, R.B., and Ponting, C.P. (2001) Sialidase-like Asp-boxes: sequence-similar structures within different protein folds. *Protein Science* 10: 285-292

Cruz, W.T., Young, R., Chang, Y.F., and Struck, D.K. (1990) Deletion analysis resolves cell-binding and lytic domains of the *Pasteurella* leukotoxin. *Mol Microbiol* 4:1933-1939

Diatchenko, L., Lau, Y.F.C., Campell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D., and Siebert, P.D. (1996) Suppression subtractive

hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 93: 6025-6030

Domenighini, M., Relman, D., Capiau, C., Falkow, S., Prugnola, A., Scarlato, V., and Rappuoli, R. (1990) Genetic characterization of *Bordetella pertussis* filamentous haemagglutinin: a protein processed from an unsually large precursor. *Mol Microbiol* 4: 787-800

Economou, A., Hamilton, W.D.O., Johnston, A.W.B., and Dowie J.A. (1990) The *Rhizobium* nodulation gene *nodO* encodes a Ca^{2+}-binding protein that is exported without N-terminal cleavage and is homologous to haemolysin and related proteins. *EMBO J* 9: 349-354

El-Azami-El-Idrissi, M., Bauches, C., Loucka, J., Osicka, R., Sebo, P., Ladant, D., and Leclerc, C. (2003) Interaction of *Bordetella pertussis* adenylate cyclase with CD11b/CD18. *J Biol Chem* 278: 38514-38521

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K., and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res* 30: 235-238

Felsenstein, J. (2002) PHYLIP, Phylogeny Inference Package, Version 3.6(alpha3). Retrieved from http://evolution.geneticswashington.edu/phylip.html, Department of Genome Sciences, University of Washington, Seattle, Washington

Figueras, J.B., Cox, R.P., Højrup, P., Permentier, H.P., and Miller, M. (2002) Phylogeny of the PscB reaction center protein from green sulfur bacteria. *Photosynthesis Res* 71: 155-164

Frigaard, N.U., and Bryant, D.A. (2001) Chromosomal gene inactivation in the green sulfur bacterium *Chlorobium tepidum* by natural transformation. *Appl Environ Microbiol* 67: 2538-2544

Fröstl, J., and Overmann, J. (1998) Physiology and tactic response of "*Chlorochromatium aggregatum*". *Arch Microbiol* 169: 129-135

Fröstl, J., and Overmann, J. (2000) Phylogenetic affiliation of the bacteria that constitute phototrophic consortia. *Arch Microbiol* 174: 50-58

Gasol, J.M., Jürgens, K., Massana, R., Calderón-Paz, J.I., and Pedrós-Alió, C. (1995) Mass development of *Daphnia pulex* in a sulfide-rich pond (Lake Cisó). *Arch Hydrobiol* 132: 279-296

Gasteiger, E., Hoogland ,C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In:Walker, J.M. (ed): The Proteomics Protocols Handbook. Humana Press p. 571-607

Gattiker, A., Gasteiger, E., and Bairoch, A. (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. *Applied Bioinformatics* 1: 107-108

George, R. A., and Heringa , J. (2000) The REPRO server: finding protein internal sequence repeats through the web. *Trends Biochem Sci* 25: 515-517

Glaeser, J., and Overmann, J. (2003) The significance of organic carbon compounds for *in situ* metabolism and chemotaxis of phototrophic consortia. *Environ Microbiol* 5: 1053-1063

Glaeser, J., and Overmann, J. (2004) Biogeography, evolution, and diversity of epibionts in phototrophic consortia. *Appl Environ Microbiol* 70: 4821-4830

Granvogl, B., Gruber, P., and Eichacker, L.A. (2007) Standardisation of rapid in-gel digestion by mass spectrometry. *Proteomics* **7**: 642-654

Guex, N., and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* 18: 2714-2723

Gurskaya, N.G., Diatchenko, L., Chenchik, A., Siebert, P.D., Khaspekov, G.L., Lukyanov, K.A., Vagner, L.L., Ermolaeva, O.D., Lukyanov, S.A., and Sverdlov, E.D. (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristal 13-acetate. *Anal Biochem* 240: 90-97

Hiller, K., Grote, A., Scheer, M., Münch, R., and Jahn, D. (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* 32: W375-W379

Isberg, R.R., and Tran Van Nhieu, G. (1994) Binding and internalization of microorganisms by integrin receptors. *Trends Microbiol* 2: 10-14

Jobby, M.K., and Sharma, Y. (2005) Calcium-binding crystallins from *Yersinia pestis*. *J Biol Chem* 280: 1209-1216

Jones, H., Ostrowski, M., and Scanlan, D.J. (2006) A suppression subtractive hybridization approach reveals niche-specific genes that may be involved in predator avoidance in marine *Synechococcus* isolates. *Appl Environ Microbiol* 72: 2730-2737

Kajava, A.V., Cheng, N., Cleaver, R., Kessel, M., Simon, M.N., Willery, E., Jacob-Dubuisson, F., Locht, C., and Steven, A.C. (2001) Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. *Mol Microbiol* 42: 279-292

Kanzler, B.E.M., Pfannes, K.R., Vogl, K., and Overmann, J. (2005) Molecular characterization of the nonphotosynthetic partner bacterium in the consortium "*Chlorochromatium aggregatum*". *Appl Environ Microbiol* 71: 7434-7441

Knapp, O., Maier, E., Polleichtner, G., Masin, J., Sebo, P., and Benz, R. (2003) Channel formation in model membranes by the adenylate cyclase toxin of *Bordetella pertussis*: Effect of calcium. *Biochemistry* 42: 8077-8084

Lauterborn, R. (1906) Zur Kenntnis der sapropelischen Flora. *Allg Bot Z* 12:196-197

Lawrence, J. G., and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci* USA 95: 9413-9417

Lawrence, J.G., and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143: 1843-1860

Lilie, H., Haehnel, W., Rudolph, R., and Baumann, U. (2000) Folding of a synthetic parallel beta-roll protein. *FEBS Lett* 470: 173-177

Ludwig, A., Jarchau, T., Benz, R., and Goebel, W. (1988) The repeat domain of *Escherichia coli* haemolysin (Hyl A) is responsible for its Ca^{2+}-dependent binding to erythrocytes. *Mol Gen Genet* 214: 553-561

Maruyama, K., Mikawa, T., and Ebashi, S. (1984) Detection of calcium binding proteins by ^{45}Ca autoradiography on nitrocellulose membrane after sodium dodecyl sulfate gel electrophoresis. *J Biochem* 95: 511-519

Osicka, R., Prochazkova, K., Sulc, M., Linhartova, I., Havlicek, V., and Sebo, P. (2004) A novel "clip-and-link" activity of repeat in toxin (RTX) proteins from gram-negative pathogens. *J Biol Chem* 279: 24944-24956

Overmann, J., and Tuschak, C. (1997) Phylogeny and molecular fingerprinting of green sulfur bacteria. *Arch Microbiol* 167: 302-309

Overmann, J., Tuschak, C., Fröstl, J., and Sass, H. (1998) The ecological niche of the consortium "*Pelochromatium roseum*". *Arch Microbiol* 169: 120-128

Overmann, J. (2001a) Phototrophic consortia: A tight cooperation between non-related eubacteria. In: Seckbach, J. (ed.) Symbiosis: Mechanisms and Model systems. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 239-255

Overmann, J. (2001b) Green sulfur bacteria. In: *Bergey´s Manual of Systematic Bacteriology 2nd ed.* Boone, D.R., Castenholz, R.W., and Garrity, G.M. (eds) Springer, New York, pp. 601-605

Overmann, J. (2006) (ed.) *Molecular basis of symbiosis. Progress in molecular and subcellular biology*, Vol. 41. Berlin, Germany: Springer, pp. 21-37

Overmann, J., and Pfennig, N. (1989) *Pelodictyon phaeoclathratiforme* sp. nov., a new brown-colored member of the *Chlorobiaceae* forming net-like colonies. *Arch Microbiol* 152: 401-406

Overmann, J., and Schubert, K. (2002) Phototrophic consortia: model systems for symbiotic interrelations between prokaryotes. *Arch Microbiol* 177: 201-208

Peitsch, M.C. (1995) Protein modeling by E-mail. *Bio/Technology* 13: 658-660

Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D. et al. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409: 529-533

Pfannes, K.R., Vogl, K.,and Overmann, J. (2007) Heterotrophic symbionts of phototrophic consortia: members of a novel diverse cluster of *Betaproteobacteria* characterized by a tandem *rrn* operon structure. *Environ Microbiol* 9: 2782-94.

Rajini, B., Shridas, P, Sundari, C.S., Muralidhar, D., Chandani, S., Thomas, F., and Sharma, Y. (2001) Calcium binding properties of γ-crystallin. *J Biol Chem* 276: 38464-38471

Read, T.D., Peterson, S.N., Tourassse, N., Baillie, L.W., Paulsen, I.T., et al. (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423: 81-86

Relman, D.A., Domenighini, M., Tuomanan, E., Rappuoli, R., and Falkow, S. (1989) Filamentous hemagglutinin of *Bordetella pertussis*: Nucleotide sequence and crucial role in adherence. *Proc Natl Acad Sci USA* 86: 2637-2641

Reva, O., and Tümmler, B. (2008) Think big-giant genes in bacteria. *Environ Microbiol* 10: 768-777

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S. Chain, P. *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424: 1042-1046

Rose, T., Sebo, P., Bellalou, J., and Ladant, D. (1995) Interaction of calcium with *Bordetella pertussis* adenylate cyclase toxin. Characterization of multiple calcium-binding sites and calcium-induced conformational changes. *J Biol Chem* 270: 26370-26376

Rost, B., Yachdav, G., and Liu, L. (2004) The PredictProtein Server. *Nucleic Acids Research* 32: W321-W326

Ruoshlati, E., and Pierschbacher, M.D. (1986) Arg-Gly-Asp: a versatile cell recognition signal. *Cell* 44: 517-518

Schägger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* 166: 368-379

Schink, B. (1998) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev* 61: 262-280

Schink, B. (2002) Synergistic interactions in the microbial world. *Antonie van Leeuwenhoek* 81: 257-261

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31: 3381-3385

Sebo, P., and Ladant, D. (1993) Pepeat sequences in the *Bordetella pertussis* adenylate cyclase toxin can be recognized as alternative carboxy-proximal secretion signals by the *Escherichia coli* -haemolysin translocator. *Mol Microbiol* 9: 999-1009

Seiler, M., Mehrle, A., Poustka, A., and Wiemann, S.(2006) The 3of5 web application for complex and comprehensive pattern matching in protein sequences. *BMC Bioinformatics* 7:144

Senchou, V., Weide, R., Carrasco, A., Bouyssou, H., Pont-Lezica, R., Govers, F., and Canut, H. (2004) High affinity recognition of a *Phytophthora* protein by *Arabidopsis* via an RGD motif. *Cell Mol Life Sci* 61: 502-509

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olsen, B.J., and Klenk, D.C. (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* 150: 76-85

Smith, W.P., Tai, P.C., Davis, B.D. (1978) Nascent peptide as sole attachment of polysomes to membranes in bacteria. *Proc Natl Acad Sci USA* 75: 814–817.

Tan, B.-H., Nason, E., Staeuber, N., Jiang, W., Monastryskaya, K., and Roy, P. (2001) RGD tripeptide of bluetongue virus VP7 protein is responsible for core attachment to *Culicoides* cells. *J Virol* 75: 3937-3947

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24: 4876-4882

van Ham, R.C.H.J., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F. *et al.* (2003) Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci* 100: 581-586

Vogl, K., Glaeser, J., Pfannes, K.R., Wanner, G., and Overmann, J. (2006) *Chlorobium chlorochromatii* sp. nov., a symbiotic green sulfur bacterium isolated from the phototrophic consortium "*Chlorochromatium aggregatum*". *Arch Microbiol* 185: 363-372

Ward, C.K., Lumbley, S.R., Latimer, J.L., Cope, L.D., and Hansen, E.J. (1998) *Haemophilus ducreyi* secretes a filamentous hemagglutinin-like protein. *J Bacteriol* 180: 6013-6022

Welch, R.A. (1995) Phyogenetic analyses of the RTX toxin family. In *Virulence mechanisms of bacterial pathogens 2nd ed.* Roth, J.A., Bolin, C.A., Brogden, K.A., Minion, F.C., and Wannemuehler, M.J. (eds) ASM press, Washington DC, pp. 195-206

Wistow, G. (1990) Evolution of a protein superfamily: relationship between vertebrate lens crystallins and microorganism dormancy proteins. *J Mol Evol* 30: 140-145

Zdobnov, E.M., and Apweiler, R. (2001) "InterProScan – an integration platform for the signature-regognition methods in InterPro". *Bioinformatics* 17: 847-848

Supplementary Material
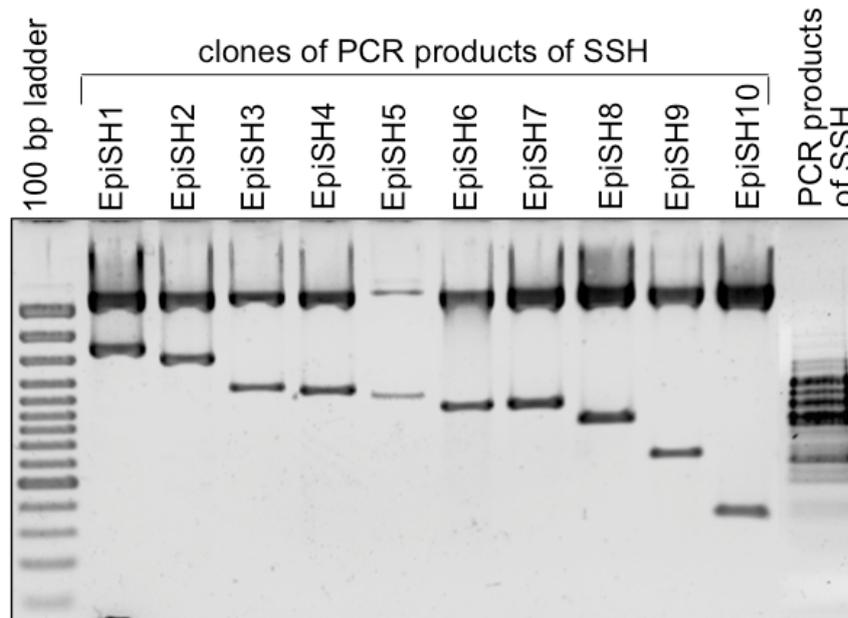
Suppl. Fig. 1. Fragments cloned from the PCR products of the suppression subtractive hybridisation. The ten different size classes are depicted. The vector is visible as a 3.9 kb band at the top of the ten lanes. PCR products are shown on the right for comparison. A negative image of an ethidium bromide-stained gel is shown. The size of the inserts surpasses that of the respective PCR products due to the presence of short vector fragments at both ends.
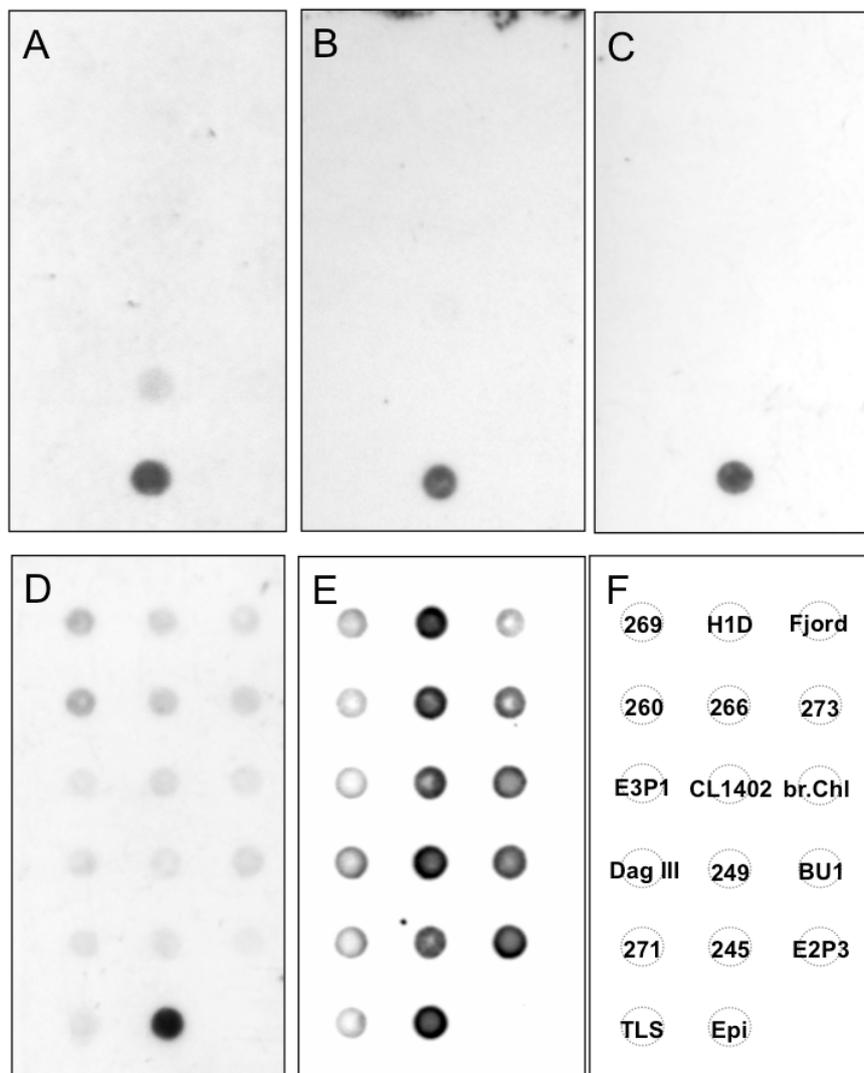
Suppl. Fig. 2. Dot Blot hybridisations of genomic DNA from 17 different green sulfur bacteria with probes targeting putative symbiosis genes of *Chlorobium chlorochromatii* CaD. **A.** Hybridization with a probe against ORF Cag1919. **B.** Probe targeting clone EpiSH4 (ORF Cag0616). **C.** Probe targeting clone EpiSH8 (ORF Cag1920). **D.** Probe targeting clone EpiSH6 (ABC transporter, compare Table 1). **E.** Probe targeting the 16S rRNA gene of *Chl. chlorochromatii*. **F.** Blotting scheme of the samples. Strain designations: 269, *Chl. phaeovibrioides* DSMZ 269[T]; H1D, *Chlorobium* strain H1D; Fjord, *Chlorobium* strain Fjord; 260, *Ptc. vibrioformis* DSMZ 260[T]; 266, *Chl. phaeobacteroides* DSMZ 266[T]; 273, *Chl. luteolum* DSMZ 273[T]; E3P1, *Chl. limicola* E3P1; CL 1402, *Chl. phaeobacteroides* CL 1402; br. Chl, *Chl. phaeobacteroides* br.Chl.; DagIII, *Chl. phaeobacteroides* DagIII; 249, *Cba. thiosulfatiphilum* DSMZ 249[T]; BU1, *Chl. clathratiforme* DSMZ 5477[T] (BU1); 271, *Ptc. aestuarii* DSMZ 271[T]; 245, *Chl. limicola* DSMZ 245[T]; E2P3, *Chl. phaeobacteroides* E2P3; TLS, *Cba. tepidum* ATCC 49652[T] (TLS); Epi, *Chl. chlorochromatii* CaD (epibiont)

Suppl. Table 1. Gene specific oligonucleotides used for PCR amplification; *all primers were used at a final concentration of 1µM

Target gene	Oligonucleotide sequence (5' to 3')*	Annealing temperature (°C)	Components of PCR reaction
RTX toxin-like protein ORF Cag_1919			
RTX502f RTX4284r	ACGTTACCGTTGACCTGC CACATCGTTACCCGTACC	40x63°C	1x PCR buffer (Qiagen) 1xQ-solution 2 mM MgCl$_2$ 0.3 mM of each dNTP 5 U of Taq DNA polymerase (Qiagen)
RTX3797f RTX4266r	ATCAACGCCAGACCAAGC CACATCGTTACCCGTACC	10x70°C, 30x65°C	
ORF Cag_1920			
RTX695f RTX868r	CGTTGGCGTATCCTTCAGT GGAGTAGGGGCATAATCAAA	10x72°C, 30x67°C	
RTX4650f RTX4798r	AATGGTATGCCGGCGGTATG AGTTACGGTACCGGTTGGCTTATC	10x72°C, 30x67°C 5 % acetamide	1x PCR buffer (GeneAmp) 2 mM MgCl$_2$ 0.2 mM of each dNTP 1.25U of AmpliTaq Gold
RTX9881f RTX10057r	GAAGTGGCGATTAAAACAGG GCATAAGCATCCGGTACAAT	10x74°C, 30x69°C	
Putative adhesion protein ORF Cag_0616			
Adh2436f Adh2556r	ACACACATGGCAGTTCCCTTCA TAGCAATTGCCCGCCGTATCT	10x72°C, 30x67°C	
Adh27378f Adh27546r	TGGTAGTGGCACGGGTGAG TCCAGCGGTCATTTTTCTCA	10x68°C, 30x63°C	
Adh53556f Adh53697r	AGGCGATGCAGAGATTAAA TGGCTTCAAGTCTCAGG	10x70°C, 30x65°C	
Sigma factor A			
GSB-SigA-F4 GSB-SigA-R1	ATTGTGCG(AC)(CT)T(GT)CC AT(AT)GG(CT)ATGGA(CT)AATCCGCT	10x61°C, 30x56°C	1x PCR buffer (GeneAmp) 3.5 mM MgCl$_2$ 0.2 mM of each dNTP

Chapter 4

Expression-based identification of genetic determinants of the bacterial symbiosis "*Chlorochromatium aggregatum*"

Summary

The phototrophic consortium "*Chlorochromatium aggregatum*" is a highly structured association of green sulfur bacterial epibionts surrounding a central, motile bacterium and is the most specific symbiosis currently known between two phylogenetically distinct bacterial species. Genes and gene products potentially involved in the symbiotic interaction were identified on the genomic, transcriptomic and proteomic level. As compared to the 11 available genomes of free-living relatives, only 186 ORFs were found to be unique to the epibiont genome. 2-D differential gel electrophoresis (2-D DIGE) of the soluble proteomes recovered 1612 protein spots of which 54 were detected exclusively in consortia but not in pure epibiont cultures. Using mass spectrometry analyses, the 13 most intense of the 54 spots could be attributed to the epibiont. Analyses of the membrane proteins of consortia, of consortia treated with cross-linkers, and of pure cultures indicated that a branched chain amino acid ABC-transporter binding protein is only expressed in the symbiotic state of the epibiont. Furthermore, analyses of chlorosomes revealed that an uncharacterized 11 kDa epibiont protein is only expressed during symbiosis. This protein may be involved in the intracellular sorting of chlorosomes. Application of a novel prokaryotic cDNA suppression subtractive hybridisation technique led to identification of 14 differentially regulated genes, and comparison of the transcriptomes of symbiotic and free-living epibionts indicated that 328 genes were differentially transcribed. The three approaches were mostly complementary and thereby yielded a first inventory of 352 genes that are likely to be involved in the bacterial interaction in "*C. aggregatum*". Notably, most of the regulated genes encoded components of central metabolic pathways whereas only very few (7.5%) of the unique `symbiosis genes´ turned out to be regulated under the experimental conditions tested. This pronounced regulation of central metabolic pathways may serve to fine-tune the symbiotic interaction in `*C. aggregatum*´ in response to environmental conditions.

Introduction

Research on symbiotic interactions involving prokaryotes so far has focused on their associations with higher eukaryotes. However, numerous cases of highly structured, purely prokaryotic associations between phylogenetically distinct bacterial species have been documented (Overmann, 2002). These so-called consortia include aggregates of deltaproteobacteria and archaea that mediate anaerobic methane oxidation (Boetius *et al.*, 2000), or associations consisting of two different archaeal species (Huber *et al.*, 2002). Monospecific cell-cell interactions have also been documented between cells of a single deltaproteobacterium species constituting the highly structured magnetotactic multicellular prokaryotes (Wenter *et al.*, 2009). Microbial consortia are not only relevant for maintaining biogeochemical cycles in different environments, but also are of medical (e.g. in dental plaque; Whittaker *et al.*, 1996) and technological significance (e.g. in wastewater treatment; de Bok *et al.*, 2004). Finally, the molecular mechanisms of bacteria-bacteria interactions in consortia have implications for the evolution of bacteria interacting with human or plant hosts.

Phototrophic consortia represent the most highly developed type of bacterial interactions between different bacteria (Schink, 2002). Nineteen different types of phototrophic consortia have been described to date (Glaeser and Overmann, 2004). Depending on the type, up to 69 green sulfur bacterial epibionts surround a central colourless, rod-shaped betaproteobacterium in a highly ordered fashion. Although already known for more than a century (Lauterborn, 1906), a stable laboratory culture of the phototrophic consortium "*Chlorochromatium aggregatum*" was established only a decade ago (Fröstl and Overmann, 1998). The epibiont of "*C. aggregatum*" could subsequently be isolated in pure culture and was described as *Chlorobium chlorochromatii* CaD (Vogl *et al.*, 2006). It is an obligately anaerobic, immotile and photolithoautotrophic green sulfur bacterium which utilises sulfide as electron donor and thus resembles free-living green sulfur bacterial species. The epibiont cells contain chlorosomes which represent the principal photosynthetic light-harvesting structures of green sulfur bacteria. Each chlorosome consists of up to 250,000 self-assembled bacteriochlorophylls surrounded by a monogalactosyl-diglyceride monolayer and chlorosomal (Csm) proteins (Frigaard and Bryant, 2006). When in association, specific morphological adaptations are observed in the epibiont. These include an additional layered structure at the inner face of the cytoplasmatic membrane, as well as an intracellular sorting process that leaves the adhesion site void of chlorosomes (Vogl *et al.*, 2006; Wanner *et al.*, 2008). Intact phototrophic consortia show a chemotactic behaviour toward sulfide and 2-oxoglutarate (Fröstl and Overmann, 1998; Glaeser and Overmann, 2003). They accumulate in the light by means of a scotophobic response, whereby the central bacterium confers motility to the consortium (Fröstl and Overmann, 1998; Glaeser and Overmann, 2003). Cell division of the partner bacteria in

phototrophic consortia is highly coordinated producing two complete daughter consortia (Overmann *et al.*, 1998). These observations indicate that rapid exchange of multiple signals and highly specific reciprocal regulation mechanisms must exist between the epibionts and the central bacterium.

Recently, four putative symbiosis genes unique to the epibiont amongst sequenced green sulfur bacterial genomes were described. Two of the genes code for unusually large haemagglutinin-like proteins, one for a putative haemolysin and another one for a RTX toxin-like protein predicted to form a C-terminal calcium-binding beta roll structure. These four genes exhibit similarity to virulence factors of typical proteobacterial pathogens and may have been transferred laterally into the epibiont genome (Vogl *et al.*, 2008).

"*C. aggregatum*" represents the first cultivable model system available to dissect the molecular basis of the symbiotic interaction between different prokaryotes and therefore provides the unique opportunity to elucidate general principles of such highly specific interactions between different bacterial cells. In the present study we extended the analysis of the symbiosis in "*C. aggregatum*" beyond the four known putative symbiosis genes. We combined genomic approaches with comparative transcriptomic and proteomic analyses of consortia and pure epibiont cultures in order to elucidate the genes and gene products of the green sulfur bacterial epibiont that are involved in the interaction with its nonphotosynthetic partner.

Results

Identification and classification of genes unique to the epibiont genome

The 12 available genome sequences of green sulfur bacteria cover the known phylogenetic diversity of this group (Eisen *et al.*, 2002; Z. Liu, T. Li, F. Zhao, J. Overmann, and D. A. Bryant, unpublished results). Therefore, a comparison of the epibiont genome to the 11 genomes of the free living relatives was used in order to identify open reading frames (ORFs) which are unique to the epibiont *Chl. chlorochromatii* CaD. Such genes could potentially be involved in the interaction with the central bacterium in the "*C. aggregatum*". Pairwise comparisons by *in silico* subtractive hybridisation analysis revealed that the epibiont genome contained 339 (compared to *Chlorobium clathratiforme* DSM 5477[T]) to 812 (compared to *Chloroherpeton thalassium* ATCC 35110[T]) unique genes. When compared to the 11 other green sulfur bacterial genomes as a whole, a total of 186 ORFs were identified to be unique for the epibiont (Suppl. Table S1). Of these, 99 ORFs code for hypothetical proteins with unknown functions, 25 for proteins of DNA or protein modification (acetyl-, nucleotidyl-, glycosyl- or methyltransferases), 18 for proteins involved in cell membrane, cell wall and capsule formation, 12 for resistance against foreign DNA and phages, 6 for transcriptional regulators, another 6 for conjugative transfer, 5 for ABC-transporters, 4 for virulence factors, another 4 for signalling and regulation, and 7 for proteins with other functions. This

analysis extended the inventory of known epibiont ORFs with similarities to known proteobacterial virulence factors (Vogl *et al.*, 2008). Cag_0615 codes for an outer membrane efflux protein that contains a conserved TolC-like domain and typically is part of bacterial type I secretion systems. The gene product of Cag_1408 is related to the *Escherichia coli* membrane fusion protein HlyD which mediates the transport of haemolysin across the periplasm as part of the type I secretion system (Balakrishnan *et al.*, 2001), whereas the product of Cag_1570 is related to VapD, a putative toxin of a toxin-antitoxin pair found in many pathogenic bacteria (Daines *et al.*, 2004).

Based on the information provided by the IMG database (compare Experimental Procedures) and additional analyses of signal peptides and membrane protein topology using SPOCTOPUS (Viklund *et al*., 2008), 36 of the ORFs had a signal peptide and 28 possessed predicted transmembrane helices (Suppl. Table S1). 15 ORFs showed both characteristics and thus most likely encode membrane proteins targeted to the cytoplasmatic membrane using the Sec machinery (Driessen and Nouwen, 2008). The unique ORFs of the epibiont genome formed 21 gene clusters with lengths between 0.98 and 22.7 kb (average length, 11 kb) corresponding to clusters of 2 – 23 ORFs (average, 10 ORFs).

Changes in the soluble proteome of symbiotic and non-symbiotic epibiont cells

In order to detect differentially expressed soluble proteins in the symbiotic and free-living state, a comparative proteome analysis of "*C. aggregatum*" and axenic epibiont cultures was conducted using 2-D DIGE technology (Fig. 1, Suppl. Fig. S1). In order to differentiate growth phase-related from symbiosis-related effects, the protein expression patterns in the exponential and stationary phases of pure epibiont cultures were also analysed (data not shown). Three spots were detected to be present or absent depending on the growth phase and thus excluded from further analysis.

2-D DIGE of the soluble proteomes recovered a total of 1612 different protein spots over all gels. Extracts of consortia on average yielded 1574 spots of which 54±7 (3.6%) were detected exclusively in extracts from consortia but not in pure epibiont cultures. The 13 most intense of the 54 spots could be attributed to the epibiont based on ESI-MS/MS analysis and comparison with the available genome sequence (spots labeled C1-C13; Fig. 1; Table 1), whereas the protein content of the other spots was too low to be analyzed. A specific function could be assigned to ten of the proteins. Most notably, the proteins identified as nitrogen regulatory protein P-II, 2-isopropylmalate synthase and glutamate synthase (Table 1) are likely to be involved in the central amino acid metabolism. Two proteins involved in sugar metabolism (glyceraldehyde-3-phosphate dehydrogenase, phosphotransferase protein IIA), as well as porphobilinogen deaminase and UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase were identified (Table 1). In addition to the 13 epibiont proteins, three of the proteins which were exclusively present in intact consortia were

identified as phosphoenolpyruvate carboxykinase, UDP-glucose-4-epimerase and enolase (C14-C16; Fig. 1, Table 1). The respective amino acid sequences were subjected to BLASTX analysis (http://blast.ncbi.nlm.nih.gov) and revealed a phylogenetic affiliation with the corresponding proteins of betaproteobacteria, thereby indicating a possible origin of these proteins from the central rod. In contrast to symbiotic epibionts, only two (0.1%) of all protein spots were exclusively present in the soluble proteome of the free-living epibiont (E1, E2; Fig. 1). These two proteins were identified as leucine aminopeptidase and an oxidoreductase of the short chain dehydrogenase family (Table 1). Further 2-D DIGE analyses were performed in order to study the effects of peptone or consortia culture supernatant on the expression of soluble proteins of the epibiont. Addition of peptone stimulated the expression of the phosphotransferase system protein IIA (Cag_1468) in pure epibiont cultures which was otherwise only detectable in symbiotic epibionts (data not shown). However, peptone did not induce the expression of any of the 53 other differentially expressed soluble proteins. Addition of supernatant of consortia cultures caused the expression of elongation factor Tu (Cag_1853) and of phenylalanyl-tRNA synthetase (Cag_1544) (data not shown) neither of which had been detected in the symbiotic nor in the free-living state of the epibiont.
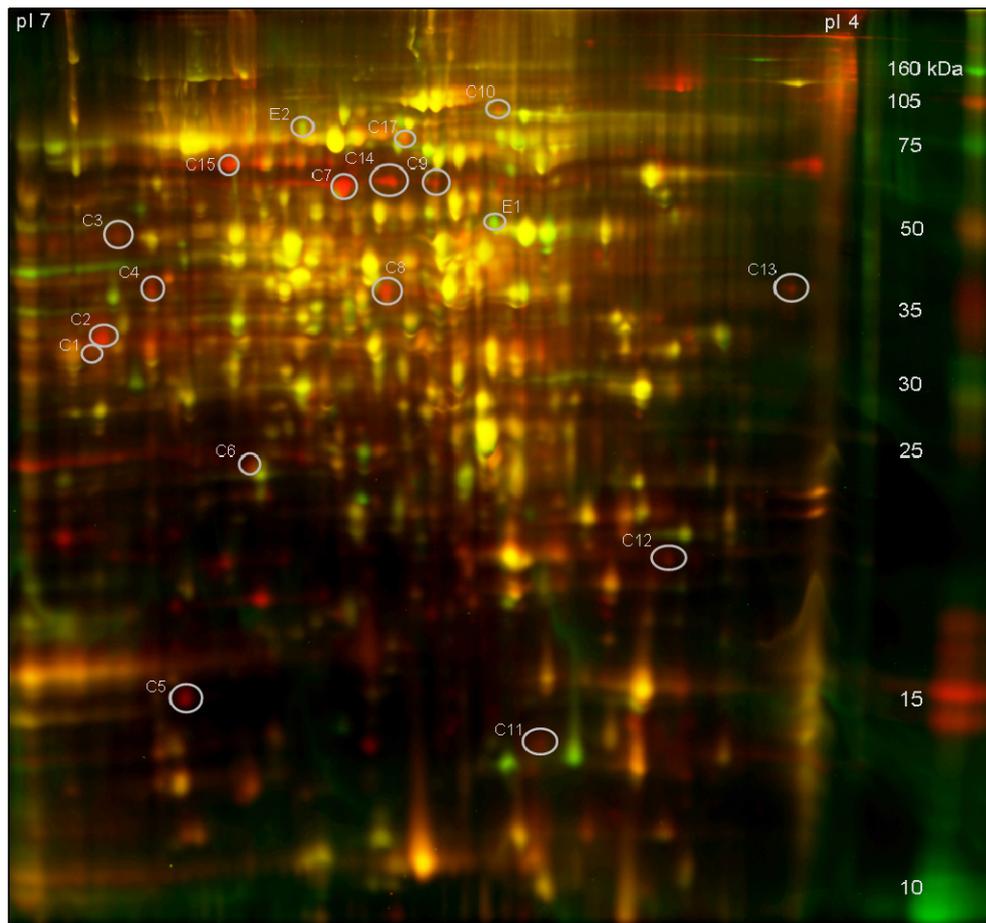
Fig. 1. Comparative 2-D DIGE analysis of the epibiont cytoplasmatic proteom in the symbiotic (red) and non-symbiotic state (green). Proteins expressed equivalently in both conditions register as yellow in the image overlay.

Table 1: Differentially regulated proteins identified by a combination of 2D-DIGE, ESI-MS/MS and database analysis of consortia and epibiont pure cultures

Spot name/ Locus tag	Protein description			Protein expression (+/- present)	
	Protein name	UniProtKB/TrEMBL/EC-numbers	Predicted function	C	E
C1/Cag_1515	Porphobilinogen deaminase	Q3AQF5/2.5.1.61	Porphyrin metabolism, heme biosynthesis	+	-
C2/Cag_1420	Glyceraldehyde-3-phosphate dehydrogenase	Q3AQP7/1.2.1.12	Glycolysis/glyconeogenesis	+	-
C3/Cag_1150	Protein disulfide-isomerase	Q3ARG3/-	Protein disulfide oxidoreductase activity; thioredoxin domain	+	-
C4/Cag_1154	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	Q3ARF9/2.3.1.-	LpxD-like; Lipid A biosynthesis	+	-
C5/Cag_1468	PTS IIA-domain protein	Q3AQK1/-	Phosphotransferase system protein IIA with fructose-specific domain	+	-
C6/Cag_0538	Flavoprotein cofactor	Q3AT64/1.18.1.2	Cofactor of glutamate synthase (Cag_0537); NAD-binding	+	-
C7/Cag_1673	2-Isopropylmalate synthase, yeast type	Q3AQ00/2.3.3.13	Leucine biosynthesis	+	-
C8/Cag_0188	Putative uncharacterized protein	Q3AU59/-	2-Nitropropane dioxygenase; electron carrier/oxidoreductase	+	-
C9/Cag_0537	Glutamate synthase	Q3AT65/1.4.1.13	Glutamate biosynthesis; glutamine degradation; nitrogen metabolism	+	-
C10/Cag_1737	Putative uncharacterized protein	Q3APT7/-	Ferritin-like AB metal binding domain	+	-
C11/Cag_1245	Nitrogen regulatory protein P-II	Q3AR69/-	Amino acid metabolism; enzyme regulator of glutamine synthetase; regulation of nitrogen utilization	+	-
C12/Cag_0444	Inorganic pyrophosphatase	Q3ATF8/3.6.1.1	Pyrophosphate hydrolysis	+	-
C13Cag_1572	Putative uncharacterized protein	Q3AQ98/-	Nucleoside-diphosphate-sugar epimerase; nucleotide-sugars	+	-

ID	Name	Accession/EC	Function		
E1/Cag_0823	Leucine aminopeptidase	Q3ASD6/3.4.11.1	Release of an N-terminal amino acids, preferably leucine; protein processing and degradation; transcription factor	-	+
E2/Cag_1642	Oxidoreductase, short chain dehydrogenase family	Q3AQ31/1.1.1.184	Reduction of short chained ketones	-	+
C14/BB2434 *Bordetella bronchiseptica* RB50	Phosphoenolpyruvate carboxykinase	Q7WJQ9/4.1.1.32	Pyruvate metabolism via TCA cycle	+	-
C15/NMB0064 *Neisseria meningitidis* MC58	UDP-glucose-4-epimerase	P56985/5.1.3.2	Galactose and nucleotide sugars metabolism	+	-
C17/bpro_3184 *Polaromonas* sp. JS666	Enolase	Q128E4/4.2.1.11	Glycolysis/gluconeogenesis	+	-

Comparison of consortia and epibiont membrane proteins

Two distinct proteins bands were found in extracts of membrane proteins of consortia when compared with those of pure epibiont cultures (Fig. 2; lanes C and E; bands CM1, CM2). Both protein bands were absent in protein extracts of the epibiont (Fig. 2A; lane E) and were missing in consortia cross-linked with 3,3′-dithio*bis*[sulfosuccinimidylpropionate] (DTSSP) or *bis*[sulfosuccinimidyl] suberate (BS^3) (Fig. 2B; lanes DTSSP, BS^3), indicating that the respective proteins were localised at the cell surface or in the periplasm.

In the protein band running at 40 kDa (CM2), all ten different peptides detected by ESI-MS/MS were related to the amino acid binding protein of the branched-chain amino acid ABC-transporter of *Chl. chlorochromatii* CaD (Cag_0853). To analyse the induction of this protein in more detail, epibiont pure cultures were grown in K3 medium supplemented with either peptone (0.05 %, w/v), branched chain amino acids (leucine, valine or isoleucine in non-inhibitory concentrations of 0.1 mM or 1 mM) or consortia culture supernatant. Whereas supplementation with the former two did not affect the pattern of membrane proteins, cultivation of the epibiont with consortia culture supernatant stimulated the expression of the ABC-transporter binding protein (Fig. 2A, lane Es).

The distinct high molecular mass protein band at 100 kDa (CM1) contained different peptides related to filamentous haemagglutinin-like proteins of *Burkholderia pseudomallei* as well as to type IV pilus biogenesis outer membrane proteins of *Delftia acidovorans* and *Variovorax paradoxus.* Since the latter proteins have a very similar molecular weight of about 100 kDa and are phylogenetically affiliated with betaproteobacteria, the detected peptides most likely originate from the central rod. Within this band, an additional single peptide was detected which matched a haemagglutinin-like protein of the epibiont (Cag_1053). In the epibiont genome, this gene clusters with another haemagglutinin-like protein (Cag_1055) and two haemolysin activation/secretion proteins (Cag_1054/Cag_1056). Based on the 53-65% amino acid sequence similarity to corresponding sequences of *Chlorobium ferrooxidans* DSM 13031[T], these ORFs are not unique to the "*C. aggregatum*" epibiont, however. Subsequent searches of the epibiont genome identified yet another filamentous haemagglutinin-like protein (Cag_1512) which shows 65% similarity to an ORF of *Chlorobium clathratiforme* DSM 5477[T].

Additional experiments demonstrated that cell-cell attachment of the partner bacteria in "*C. aggregatum*" becomes permanent if BS^3 or DTSSP are added to consortia cultures. After cross-linking, consortia could not be disintegrated any more by EGTA or other Ca^{2+}-chelating agents. These results suggest that proteins located at the cell surface are involved in the specific cell-cell-binding within phototrophic consortia.
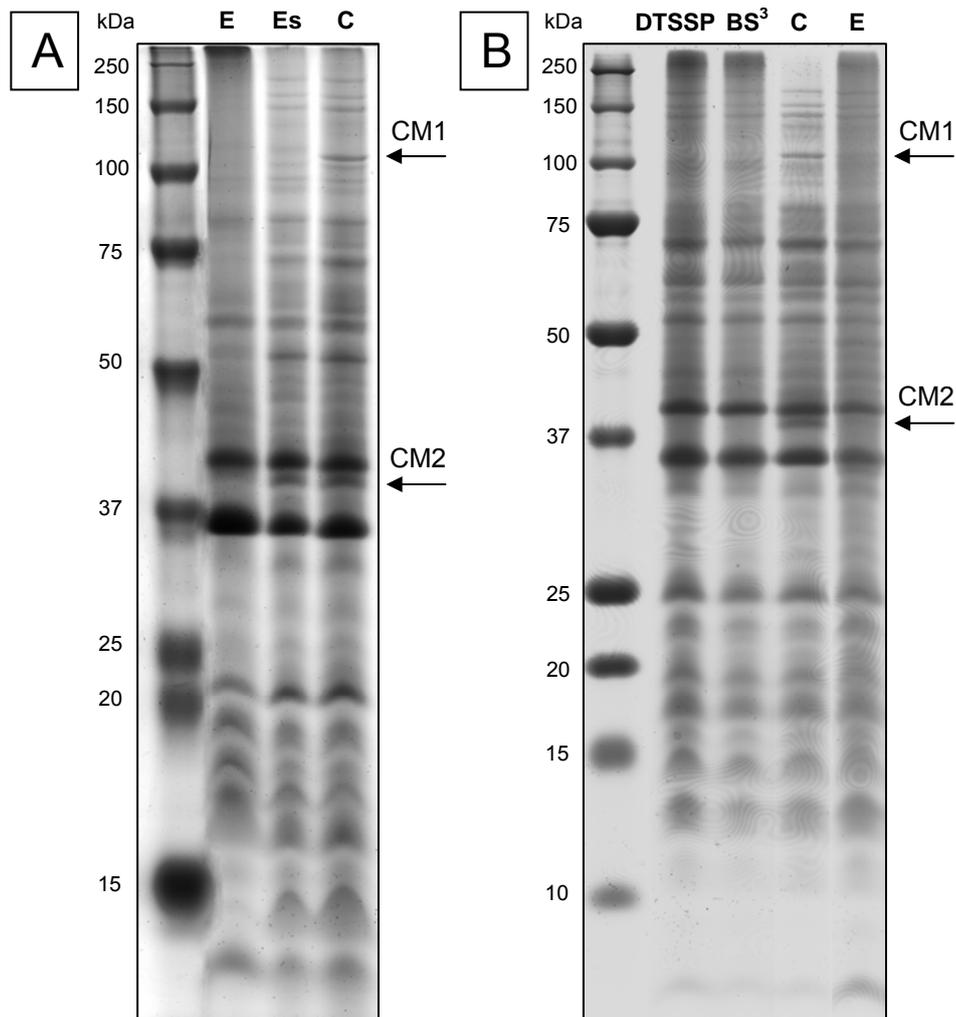
Fig. 2. A. SDS-PAGE of membrane proteins of epibiont pure culture (E), of a pure culture stimulated with consortia culture supernatant (Es) and membrane proteins of consortia (C). **B.** SDS-PAGE of BS[3]- and DTTSP-crosslinked consortia membrane proteins compared to a non-crosslinked control of consortia (C) and membrane proteins of epibiont cultures (E). Arrows indicate the presence of a filamentous haemagglutinin/type IV pilus-related protein of the central bacterium (CM1) and a most likely periplasmic binding protein of the epibiont (CM2).

Comparison of consortia and epibiont chlorosomes

A comparison of chlorosome membrane proteins of free-living and symbiotic epibionts by SDS-PAGE revealed that a small, uncharacterized protein of 97 amino acids (aa) with a molecular mass of about 11 kDa (Cag_1285; Suppl. Fig. S2, arrow; Table 1) was exclusively present in chlorosome extracts of symbiotic epibionts. Amino acid sequence analysis of the Cag_1285 gene product revealed an N-terminal signal peptide, a high proportion of C-terminal repetitive elements and the absence of membrane helices. In the genome of the epibiont *Chl. chlorochromatii* CaD, ORF Cag_1285 is not located near genes for any other chlorosome protein. Based on the presence of the signal peptide, this protein most likely is localised in the cell envelope and copurifies with the chlorosomes.

Phylogenetic analysis of Cag_1285 revealed a distant relationship (similarity values < 30%; Fig. 3) to putative uncharacterized proteins of the four green sulfur bacteria *Chlorobium phaeovibrioides* DSM 265, *Chl. luteolum* DSM273[T], *Chl. phaeobacteroides* BS1 and *Prosthecochloris aestuarii* DSM271[T] as well as to proteins of human- and phytopathogenic betaproteobacteria belonging to the genera *Ralstonia*, *Burkholderia* and *Neisseria* (Fig. 3). No genes similar to Cag_1285 are present in the other green sulfur bacterial genomes. Most of the betaproteobacterial proteins are unidentified, but the amino acid sequence of *Neisseria gonorrhoeae* Q53Z52 is annotated as an outer membrane lipoprotein with unassigned function.
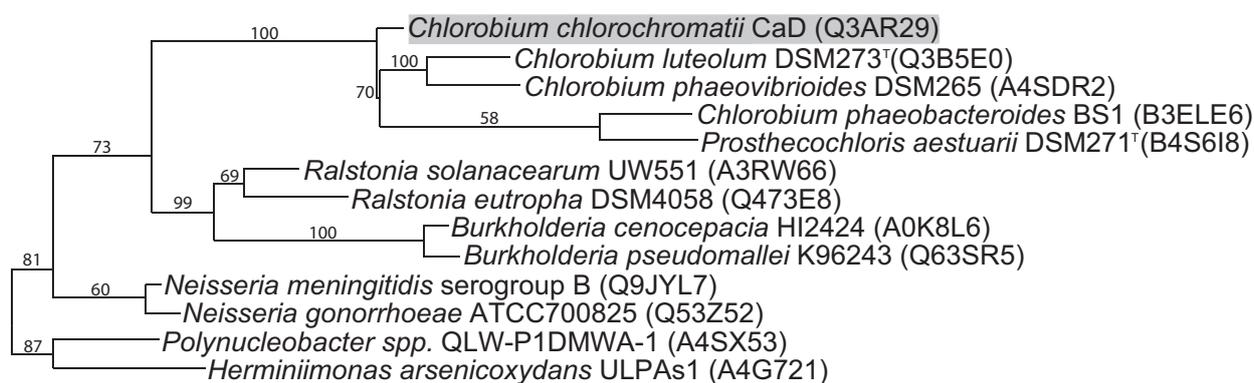
0.10

Fig. 3. Phylogenetic affiliation of the gene product of *Chlorobium chlorochromatii* Cag_1285. The phylogenetic tree was constructed using the ProteinML maximum likelihood algorithm as implemented in the ARB program package. Bootstrap values ≥ 50% are indicated at nodes and represent percentages of 100 replicates. UniProtKB/TrEMBL accession numbers for the amino acid sequences are given in parentheses. The sequences of *Polynucleobacter* sp. and *Herminiimomas arsenicoxidans* were used as outgroup. Scale bar indicates 10% substitutions per amino acid position.

Differential transcription of selected epibiont genes

Of the 15 genes which are exclusively expressed in symbiotic epibiont cells and which could be identified in the present study (13 by 2-D DIGE and 2 by crosslinking experiments), the 8 genes involved in central amino acid and sugar metabolism and in the synthesis of outer membrane, as well as the chlorosome proteins were chosen to study transcriptional regulation. RT-qPCR analyses standardised against *rpoD* transcripts showed that, when compared to the free-living state, the transcription of five of the genes was up-regulated in consortia (Fig. 4). The most conspicuous change was observed for the gene encoding nitrogen regulatory protein P-II for which a (189 ± 10)-fold increase in transcript abundance was determined. The second largest increase of (4.4 ± 1.1)-fold was detected for protein Cag_1285, corroborating the results of our membrane protein analysis. Transcription of the genes for glyceraldehyde-3-phosphate dehydrogenase, PTS IIA-domain protein

and the yeast-type isopropylmalate synthase was increased slightly but significantly [between (1.5 ± 0.2)-fold and (1.5 ± 0.3)-fold; $p < 0.025$, t-test]. Contrary to the results of the proteome analyses, however, no increase in transcription was observed for glutamate synthase, UDP-3-hydroxymyristoyl glucosamine N-acyltransferase and the binding protein of the branched-chain amino acid ABC-transporter (Fig. 4).
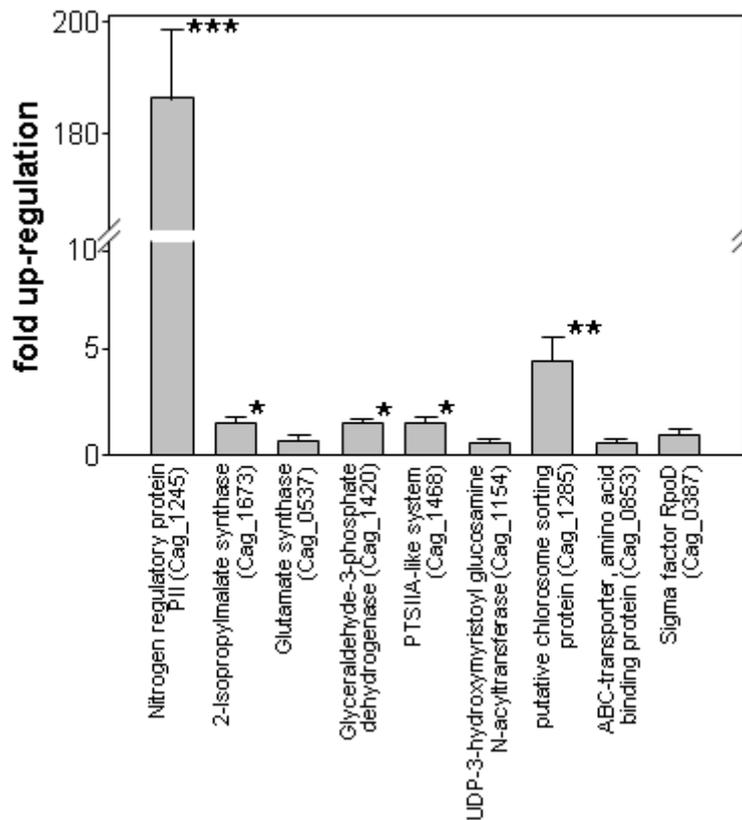
Fig. 4. RT-qPCR analysis of transcription up-regulation of selected ORFs coding for proteins only expressed in symbiotic epibionts. *, $p < 0.025$; **, $p < 0.005$; ***, $p < 0.001$.

Comparison of the transcriptomes of the symbiotic and the free-living state

Complementing the proteome-based detection of gene regulation in the epibiont, the differences in the transcriptomes of symbiotic and free-living epibiont cells were investigated using (*a*) a novel prokaryotic cDNA suppression subtractive hybridisation (cDNA-SSH) technique and (*b*) global transcriptome analyses based on Illumina sequencing.

In the analysis by cDNA-SSH, 94 clones generated by the subtraction reactions using consortia cDNA as tester were sequenced. Ten of the clones were affiliated with epibiont genes, whereas all others contained 16S or 23S rRNA gene fragments of the central bacterium. The ten transcripts of the epibiont originated from four different non-ribosomal genes (Table 2; Cag_1239, Cag_1244, Cag_1245 and 1246). ORF Cag_1239 is absent in the genome sequences of free-living green sulfur bacteria, encodes a 2461 aa long peptide and contains nine

Vibrio/*Colwellia*/*Bradyrhizobium*/*Shewanella* (VCBS) domain repeats. The 100 residue-long VCBS domain occurs in up to 35 copies in VCBS species but in lower numbers in several other bacteria. Cag_1239 resembles its known counterparts (Yousef and Espinosa-Urgel, 2007) in that it contains three of the cadherin domains thought to participate in cell-cell-adhesion. Subsequent bioinformatic searches for VCBS-domains revealed that the epibiont genome encodes the three additional VCBS-domain proteins Cag_0738 (8871 aa), Cag_1242 (16311 aa) and Cag_1560 (1838 aa). In contrast to Cag_1239, homologues of the latter three proteins are present in *Chl. luteolum* DSM 273[T] and *Chl. ferrooxidans* DSM 13031[T].

The product of Cag_1244 was identified as nitrogenase subunit NifH. The downstream ORFs Cag_1245 and Cag_1246 both encode the carbon-nitrogen regulatory protein P-II. This result is commensurate with the pronounced upregulation of transcription observed for Cag_1245 (Fig. 4) and with the detection of protein P-II by 2-D DIGE. Forty-five clones were generated from subtraction reactions using pure epibiont cultures as tester. In this experiment, 10 of the clones analysed represented epibiont mRNA transcripts, whereas the remainder contained 16S or 23S cDNA fragments of the epibiont. Sequence analysis revealed that each of the 10 mRNAs originated from genes involved in various central metabolic functions (Table 2).

A first comparative analysis of global transcriptomes of the free-living and symbiotic epibiont was achieved by Illumina sequencing of cDNA libraries prepared from a pure culture of *Chl. chlorochromatii* CaD and a "*C. aggregatum*" consortium culture. After removing sequence tags that matched to rRNA, a total of 14,117 cDNA sequence tags from the epibiont grown in the consortium were mapped to the genome of *Chl. chlorochromatii* CaD while 33,901 sequence tags originating from the free-living cells were mapped to the genome. The decrease in the consortium likely reflects the contribution of mRNAs from the central bacterium to the cDNA library. On average, sequence tags mapped to 47% of 2,002 annotated *Chl. chlorochromatii* CaD protein coding genes and to 76% of the 44 tRNAs. However, the coverage of individual protein coding genes was low; 51 genes had more than 10% of their length covered by reads in both samples and 102 had more than 5% of their length covered. 106 genes in the symbiotic association library and 153 of the genes in the free living library were covered at >10%.

A total of 328 epibiont protein coding genes were determined to be significantly differentially expressed between the consortium and free living state (Suppl. Table S2). This is 16% of the 2,002 annotated protein coding genes of the *Chl. chlorochromatii* CaD genome. Of these, 107 protein coding genes were only found in the library from "*C. aggregatum*" consortium while 84 genes were only detected in the library from the free living epibiont. Of the 137 genes represented in both libraries, Cag_1725, annotated as a potassium uptake protein, was most highly induced in the consortium (36-fold) while Cag_1430, annotated as a UDP-N-acetylenolpyruvoylglucosamine

reductase, was most highly (10-fold) induced in the free-living epibiont. In the symbiotic state, the expression of six additional epibiont genes involved in cellular build-up and maintenance was increased by 15- to 20-fold. These genes coded for ClpX, the ATP-binding subunit of Clp protease (Cag_0183), TPR repeat-containing protein (Cag_0341), DNA helicase II (Cag_1415), RecN-like DNA repair protein (Cag_1749), the 30S ribosomal protein S19 (Cag_1847) and the translation elongation factor G (Cag_1854) (Suppl. Table S2). The tetratricopeptide repeat (TPR) protein mediates protein-protein interactions and the assembly of multiprotein complexes (D'Andrea and Regan, 2003). Proteins containing TPRs are involved in a variety of biological processes, such as cell cycle regulation, transcriptional control and protein folding.

Sequence tags were detected in both free-living *Chl. chlorochromatii* CaD and the `C. aggregatum´ consortium for four putative symbiosis genes identified by genomic SSH: Cag_0614, 0616, 1919, and 1920 (Vogl *et al.*, 2008). The expression ratio indicated that mRNA for Cag_0614 ($p \leq 0.003$) and Cag_1920 ($p \leq 0.009$) was only 1.2 to 1.3-fold more abundant in consortia, while Cag_1919 decreased 1.4-fold ($p \leq 0.029$), and the 1.2-fold increase for Cag_0616 was not supported as statistically significant ($p \leq 0.30$). These data were consistent with end-point RT-PCR data that indicated that mRNAs of the four putative symbiosis genes were present both in the `C. aggregatum´ consortium and free-living *Chl. chlorochromatii* CaD (Vogl *et al.*, 2008). Notably, the expression ratio for Cag_0615, located between Cag_0614 and 0616, increased 4.3-fold ($p \leq 0.002$) in symbiotically associated epibiont cells relative to the free living state. This was consistent with the prediction that an independent promoter exists for this open reading frame (Vogl *et al.*, 2008)

Table 2: Differentially transcribed ORFs identified by a combination of cDNA subtractive hybridisation and database analysis; C: consortia as tester; E: epibiont pure culture as tester; +/- refers to transcripts detected/not detected

Locus tag	ORF name	Function prediction	C	E
Cag_1239	VCBS	VCBS and cadherin domain protein; cell-cell adhesion	+	-
Cag_1244	Nitrogenase iron protein	Nitrogenase subunit NifH; nitrogen metabolism	+	-
Cag_1245	Nitrogen regulatory protein P-II	Amino acid metabolism; enzyme regulator of glutamine synthetase; regulation of nitrogen utilization	+	-
Cag_1246	Nitrogen regulatory protein P-II	Amino acid metabolism; enzyme regulator of glutamine synthetase; regulation of nitrogen utilization	+	-
Cag_0140	ATP synthase subunit A/H+-transporting two-sector	ATP synthesis	-	+
Cag_0141	H+-transporting two-sector ATPase, gamma subunit	ATP synthesis	-	+
Cag_0227	Magnesium-chelatase, subunit H	Cobalamin biosynthesis protein CobN; Porphyrin and chlorophyll metabolism	-	+
Cag_0443	Pyrophosphate-energized vacuolar membrane proton pump	Inorganic H^+ pyrophosphatase, oxidative phosphorylation	-	+
Cag_0474	Phosphoglucomutase/phosphomannomutase	Phosphomannomutase; GDP-mannose synthesis	-	+
Cag_0475	Lipid A disaccaride synthase LpxB	Lipid A biosynthesis	-	+
Cag_0624	Thiolperoxidase	Peroxiredoxin/thioredoxin-like; posttranslational modification/protein turnover	-	+
Cag_1143	Cation efflux membrane protein	Co/Zn/Cd cation transporter	-	+
Cag_1144	Putative uncharacterized protein	Unknown function	-	+
Cag_1163	FAD-binding Oxidoreductase	D-lactate dehydrogenase; Pyruvate metabolism	-	+

Discussion

To date, the molecular determinants of consortia formation have remained unknown. Employing the phototrophic consortium "*C. aggregatum*" as a model system, the present comparative genomic, transcriptomic and proteomic study yielded a first inventory of genes with potential relevance for such interactions, and generates novel hypotheses regarding the molecular mechanisms that underlie the formation of morphologically and physiologically united higher entities by two or more different species, i.e. bacterial `heterologous multicellularity´.

Unique genes of the epibiont with functional implications for the symbiosis

Based on our *in silico* analysis, the genome of the `C. aggregatum´ epibiont contains 186 unique ORFs. While the major fraction of these ORFs code for hypothetical proteins with unknown functions and hence await future indentification, the eight unique epibiont genes Cag_0614, Cag_0615, Cag_0616, Cag_1239, Cag_1408, Cag_1570, Cag_1919, Cag_1920 match known bacterial virulence factors and therefore provide promising targets for future functional studies of the molecular coupling across the interface of the partner cells in phototrophic consortia. Of particular interest are the genes suggested to encode haemagglutinins, since the latter are typically associated with the cell membrane, are exposed at the cell surface and mediate the attachment of pathogenic bacteria to their host cells (Relman *et al.*, 1989; Kajava *et al.*, 2001); other homologues are involved in the adherence of non-pathogenic bacteria to surfaces and other microorganisms (Dalisay *et al.*, 2006).

A second conspicuous group of unique genes in the epibiont genome are the seven ORFs Cag_0648 to Cag_0650, Cag_0665, Cag_0668, Cag_0673 and Cag_0675, which are closely located to each other on the genome (Suppl. Table S1). Their gene products are likely to participate in capsular exopolysaccharide biosynthesis (Cag_0649, Cag_0650) or in transmembrane polysaccharide export (Cag_0648, Cag_0665). ORF Cag_0668 was annotated as *capA* gene which is part of a minimal set of four genes (*capB*, *C*, *A* and *E*) required for poly-γ-glutamate synthesis by Gram-positive members of the Bacillales (Candela and Fouet, 2006) but is present in only 19 (2.2%) genomes of Gram-negative bacteria. Since the epibiont genome does not contain homologues of the other three *cap* genes, Cag_0668 and the neighbouring genes maybe involved in polysaccharide formation rather than poly-γ-glutamate synthesis. Based on electron microscopic studies, cell attachment in phototrophic consortia is mediated by a dense interconnecting network of up to 150 nm long and 3 nm wide hair-like filaments which cover the epibiont cells and in turn form an elastic capsule enclosing the central rod (Wanner *et al.*, 2008). Furthermore, thick capsules have been observed in phototrophic consortia particularly from natural populations (Overmann *et al.*, 1998) when compared to laboratory cultures. The synthesis of extracellular capsular material by the

epibiont not only may contribute to the formation of cell aggregates but may also confer an additional protective mechanism to the bacterial cells in `*C. aggregatum*´ under suboptimal growth conditions in nature.

Preadaptation of free-living green sulfur bacteria to the symbiotic interaction

Pairwise comparisons with the genomes of 11 free-living green sulfur bacteria identified 339 to 812 (mean ± S.D., 557±122) genes to be unique to the epibiont genome. Considering the limited 16S rRNA gene sequence similarity between the 12 green sulfur bacteria analysed (90.2 - 97.0%; Overmann and Tuschak, 1997) these results were unexpectedly low since similar fractions of unique genes had previously been detected in much more closely (16S rRNA gene sequence similarity, 96.4% - 100%) related bacterial lineages of thermophilic *Synechococcus* spp. (393 and 503 lineage-specific ORFs; Bhaya *et al.*, 2007) and of two different *Prochlorococcus* ecotypes (364 and 923 unique genes; Rocap *et al.*, 2003). An even higher fraction of the bacterial genome encoding niche-specific functions has been documented for enterohaemorrhagic *Escherichia coli* O157:H7 that harbors 1,387 genes not found in the non-pathogenic *E. coli* K-12 (Perna *et al.*, 2001).

Even more conspicuous was the limited number of 186 ORFs that were found to be unique for the epibiont when compared to the 11 other green sulfur bacterial genomes as a whole. Low numbers of niche-specific genes have so far only been reported for pathogenic bacteria like *Salmonella enterica* (200 genes; Bowe *et al.*, 1998), or *Bacillus anthracis* (141 unique proteins compared to its close relative *Bacillus cereus*; Read *et al.*, 2003) and have been interpreted as indication for preadaptation of the nonpathogenic ancestor (Groisman and Ochman, 1997). We hypothesize that preadaptation of the ancestor and access to the existing gene pool of free-living green sulfur bacteria by lateral gene transfer represent two key features of the evolution of the mutualistic interaction in phototrophic consortia, and resulted in the limited number of novel symbiosis genes found in the epibiont.

The currently most promising candidates for `symbiosis genes´ which were already present in the free-living ancestor of the "*C. aggregatum*" epibiont and which predisposed the ancestor to a symbiotic lifestyle are the eight non-unique epibiont genes Cag_0738, Cag_1053, Cag_1054, Cag_1055, Cag_1056, Cag_1242, Cag_1512, Cag_1560 that matched known bacterial virulence factors and at the same time had homologues in free-living green sulfur bacteria. The product of one of these genes (the haemagglutinin-like protein of Cag_1053) crosslinked with a filamentous haemagglutinin/Type IV pilus protein of the central bacterium protein, indicating a direct role in the specific cell-cell-binding to the betaproteobacterial partner. These findings also provide the first indication for virulence factor-like genes mediating multicellular adhesion in non-pathogenic

bacteria.

The present study also identified the first candidate gene for the conspicuous cytological changes which occur in the epibiont in association with the central bacterium. The protein encoded by Cag_1285 copurified with the chlorosome fraction and is most likely localised in the cytoplasmic membrane or periplasmic space. It is only distantly related to its homologues in other green sulfur bacteria and was detected only in the symbiotic state. In the symbiotic state, chlorosomes are absent from the inner face of the epibiont cytoplasmic membrane at the site of attachment to the central bacterium, whereas chlorosomes are evenly distributed along the cytoplasmic membrane in epibiont cells from pure cultures (Wanner *et al.*, 2008) as in all other green sulfur bacteria. Cag_1285 could therefore be involved in the intercellular sorting of chlorosomes in the epibiont by modifying its cytoplasmic membrane so as to exclude chlorosomes from the cell-cell adhesion site. Cag_1285 thus represents a first target for the elucidation of this peculiar cell biological process in a bacterial cell.

Future studies of the genes mentioned above and their products should include comparative functional analyses of their homologues in non-symbiotic green sulfur bacteria in order to elucidate how an existing gene inventory of free-living bacteria is modified to establish beneficial interactions with another prokaryotic species. Such analyses will also provide a guideline for the study of other types of mutualistic prokaryotic associations.

Comparison of methods to assess differential gene expression in the epibiont

The three approaches to analyse differential gene expression covered different fractions of the epibiont genome. Compared to the 2002 protein coding genes annotated in the epibiont genome (http://genome.jgi-psf.org/mic_home.html), 1520 soluble proteins were detected by 2-D DIGE in axenic epibiont cultures. Most of the additional proteins identified in the consortium proteome originate from the epibiont, because only 19% of the identified protein spots could be assigned to the central bacterium.

The transcriptome analysis covered 47% of the annotated protein genes. This limitation of the transcriptome analysis is reflected by the fact that the P-II protein (Cag_1245) was only detected by a single unique sequence tag in the consortium sample despite the 189-fold increase in transcript abundance of this ORF seen by RT-qPCR. Low coverage will disproportionately affect small ORFs like Cag_1245 (351 bp), and the chlorosome-fraction associated protein Cag_1285 (291 bp) that are among the smallest proteins annotated in the "*C. aggregatum*" epibiont genome. Deeper sequencing and improved rRNA subtraction methods will improve on this in the future. On the other hand, the transcriptome analysis proved superior to the proteomic approach in identifying regulated genes (328 versus 17), because the quantitative comparison of protein spots in the case of the consortia

proteome is inherently restricted by the interference of proteins from the central bacterium.

Combining the complementary results from analyses of the proteome (Table 1), cDNA-SSH (Table 2) and trancriptome (Suppl. Table S2) the total number of genes which were differentially expressed in the symbiotic and free-living state was 352 (Fig. 5).
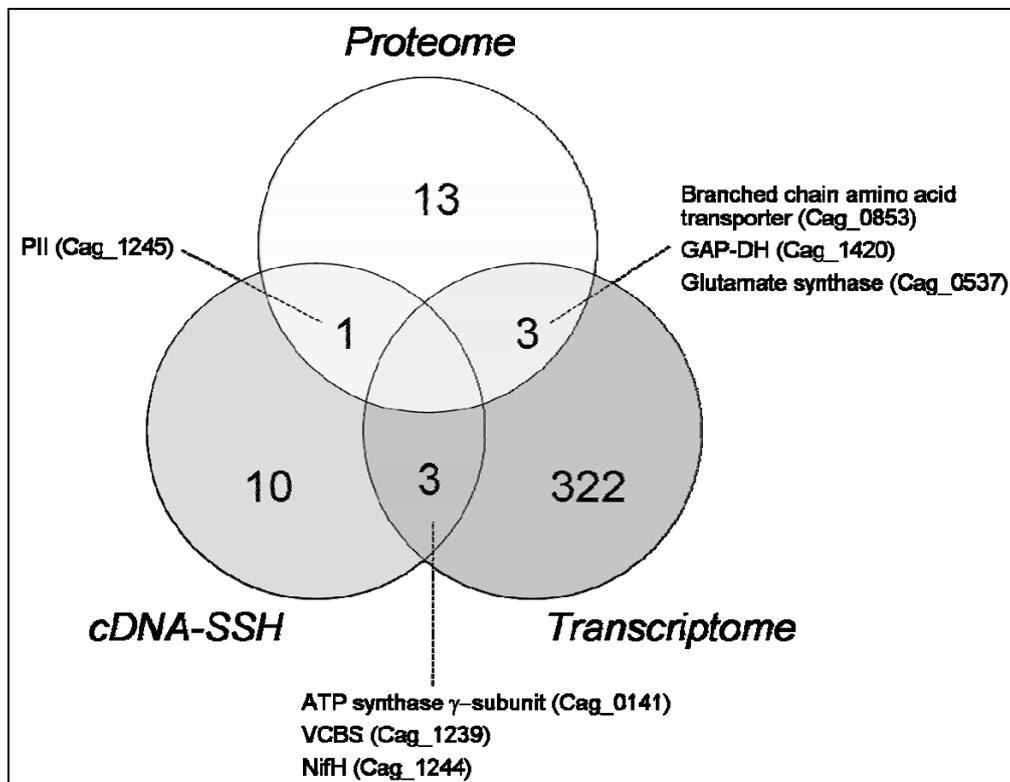
Fig. 5. Combined results of the proteome, cDNA-SSH and transcriptome analyses. The three approaches were largely complementary and identified 352 genes which were differentially expressed in the symbiotic and the free-living state of the epibiont. Differentially regulated genes detected by two of the approaches are stated explicitly.

The pronounced changes in epibiont gene expression involve central metabolic pathways rather than `symbiosis genes´

Based on our combined results, 23.3% of the genes detected in the cDNA library (328 of 1403) are differentially expressed in the symbiotic and free-living state of the epibiont. This high fraction of regulated genes is comparable to that observed in *Salmonella enterica* during the switch between the extracellular and intravacuolar environment (20.6%; Eriksson *et al.*, 2003) and to *Listeria monocytogenes* during temperature shock (maximum 25%; van der Veen *et al.*, 2007), but higher than the fractions determined for many other bacteria, like for the sulfate-reducing *Desulfovibrio vulgaris* during a heat shock response (14%; Chhabra *et al.*, 2006), for *Escherichia coli* under 100 different randomly simulated environments (13.6%; Gianchandani *et al.*, 2009), for *Thiobacillus denitrificans* switching from aerobic to denitrifying growth (10%; Beller *et al.*, 2006) or for

Thermotoga maritima switching between biofilm and planktonic populations (6%; Pysz *et al.*, 2004).

The large fraction of differentially expressed genes notwithstanding, the genome of the epibiont encodes only a few (56) proteins for environmental sensing and regulatory responses, similar to the genomes of other green sulfur bacteria (Eisen *et al.*, 2002). Extended analyses of role categories (http://cmr.jcvi.org/cgi-bin/CMR/shared/Menu.cgi?menu=genome) confirmed that all green sulfur bacterial genomes encompass much less (39 - 75) genes involved in regulation than members of other bacterial phyla (up to 764 regulatory proteins in proteobacteria). At present it remains unclear whether the pronounced changes in gene expression observed between the symbiotic and free-living state of the epibiont are controlled by the few regulatory proteins identified to date or whether the substantial fraction of regulatory proteins has been overlooked so far.

Most notably, the observed changes in gene expression of the epibiont of "*C. aggregatum*" that were detected by the transcriptome approach largely pertain to non-unique protein coding genes (314 of the 1816 non-unique genes) rather than to the unique symbiosis genes identified by in silico subtractive hybridisation (14 of 186 ORFs, marked in bold type in Suppl. Table 1). The majority of the regulated non-unique genes encoded components of central metabolic pathways as well as genes for replication, recombination and repair or for translation and ribosomal structure. In contrast, 15 of the 16 virulence factor-like genes were constitutively expressed based on the present and our previous (Vogl *et al.*, 2008) study, the VCBS-domain containing Cag_1239 representing the only exception. Thus, the symbiotic interaction in phototrophic consortia is comparable to the antagonistic interaction of human pathogens (Bowe *et al.*, 1998) since it involves the differential regulation of a large number of genes of central metabolic pathways and housekeeping functions. Similar to the emergence of pathogens like *Bacillus anthracis* (Read *et al.*, 2003), an altered gene expression of such basic cellular functions may actually represent one decisive step towards the formation of the symbiosis in phototrophic consortia.

Implications for the physiological coupling between the partner bacteria in phototrophic consortia

The experimental data previously available suggested a role of 2-oxoglutarate for the physiological coupling of the partner bacteria in phototrophic consortia. Free-living green sulfur bacteria are known to excrete significant amounts of 2-oxoglutarate (Sirevag and Ormerod, 1970) which represents a potential carbon substrate of the central bacterium. Whereas intact consortia have been shown to incorporate 2-oxoglutarate (Glaeser and Overmann, 2003), *Chl. chlorochromatii* CaD itself does not use this compound (Vogl *et al.*, 2006), suggesting that 2-oxoglutarate is in fact taken up by the central bacterium. Yet, the 2-oxoglutarate uptake in intact consortia appears to be

controlled by the physiological state of the epibionts and does not occur in the absence of light or sulfide (Glaeser and Overmann, 2003).

Based on the results of the present study, the metabolic coupling between the partner bacteria in "*C. aggregatum*" may also involve amino acids. Several of the differentially regulated and non-unique epibiont genes, among them 2-isopropylmalate synthase, glutamate synthase, glutamine synthetase, leucine aminopeptidase, various Nif proteins, nitrogen regulatory protein P-II and the amino acid binding protein of a branched chain amino acid ABC transporter (Cag_0853) are involved in nitrogen metabolism. Cag_0853 exhibits high sequence similarities of 70 to 80% to genes present in *Chl. ferrooxidans* DSM 13031[T] and *Chl. clathratiforme* DSM 5477[T]. In other bacteria, this ABC-transporter is typically upregulated under nitrogen limiting conditions (Nikodinovic-Runic *et al.*, 2009). Together with the significantly increased expression of the nitrogenase genes *nifH*, *nifE*, *nifB* in the symbiotic epibiont cells, the expression pattern of Cag_0853 hence suggests that epibionts experience nitrogen limitation in the symbiotic state which maybe caused by an increased synthesis and transfer of branched chain amino acids. A transfer of amino acids or unidentified nitrogen compounds to prokaryotic epibionts has also been suggested for other interspecific bacterial interactions like the associations of *Rhizobium* sp. WH2K with the nitrogen fixing filamentous cyanobacterium *Anabaena* sp. SSM-00 (Behrens *et al.*, 2008) and for archaeal consortia of *Ignicoccus hospitalis* and *Nanoarchaeum equitans* (Jahn *et al.*, 2008). In the case of archaeal consortia, analyses of membrane proteins suggest that the binding protein of an ABC transporter of *I. hospitalis* is involved in metabolite exchange between both partners (Burghardt *et al.*, 2009).

Interestingly, branched chain amino acids have recently been shown to play a decisive role for the root nodule symbiosis (Prell *et al.*, 2009). In this symbiotic system, the plant host provides branched chain amino acid and thereby causes a down regulation of their synthesis of the *Rhizobium* bacteroids. This in turn leads to conditional (`symbiotic´) auxotrophy of the bacterial symbionts, a shut down of its own ammonia assimilation and the excretion of ammonia by the bacteria (Lodwig *et al.*, 2003). The interaction in root nodules involves branched chain amino acid ABC transporters. Based on our observations of a differential regulation of the amino acid binding protein of a branched chain amino acid ABC transporter (Cag_0853), and of the nitrogen fixing capability of the epibiont (Vogl *et al.*, 2006), future studies of the exchange and reciprocal control in phototrophic consortia by branched chain amino acids appear highly rewarding. Providing a first indication of such reciprocal control between the partner bacteria, the expression of the binding protein in pure epibiont cultures could also be provoked by the addition of culture supernatant from consortia, indicating that the expression of the binding protein is controlled by some sort of signal exchange with the central bacterium.

The by far most pronounced change in expression occurred for the P-II protein (Cag_1245) which was 189-fold up-regulated in symbiotic epibionts. The nitrogen regulatory protein P-II interacts with various target proteins including transcription factors, key regulatory and metabolic enzymes, most of which are involved in crucial reactions in nitrogen assimilatory pathways. It responds to changes of cellular energy status and the cellular carbon-nitrogen balance in cyanobacteria (Forchhammer, 2008). Our results thus provide further evidence for nitrogen limitation of the epibiont cells in the symbiotic state. In the amino acid metabolism, P-II acts as enzyme and transcriptional regulator of the glutamine synthetase during nitrogen limitation. The fact that the transcription of glutamine synthetase (Cag_1588) was detected exclusively in consortia and parallels the upregulation of P-II confirms that P-II exerts a regulatory function also in the green sulfur bacterial epibiont.

Epibionts of phototrophic consortia seem to be specifically adapted to a symbiosis with the central rod and have never been detected free-living in nature (Glaeser and Overmann, 2004). While a comparison between a symbiotic and a free-living state was chosen in the present study in order to elicit the most pronounced response, such a switch under natural conditions is unlikely. Yet, even this most pronounced change in environmental conditions did not elicit a regulation of most of the unique symbiosis genes. Such a regulation hence seems to be dispensable for the symbiotic interaction, or the symbiotic association has been stable in nature for a sufficient time that regulation of the potential symbiosis genes no longer confers any selective advantage. Rather, the regulatory response detected in the present study pertains to the central metabolic pathways. Since a switch between non-symbiotic and symbiotic lifestyle is unlikely, regulation may serve the purpose of fine-tuning the physiological coupling between the epibiont and the central bacterium if the consortium as a whole experiences changes in environmental conditions.

Experimental Procedures

Bacterial cultures and growth conditions

Cultures of the phototrophic consortium "*C. aggregatum*" were grown in 10 l glass flasks containing anoxic K3 medium (pH 7.2) with 1 mM sulfide as electron donor and reductant (Kanzler *et al.*, 2005). The flasks were incubated at 20 °C and under continuous illumination of 20 μmol quanta·m^{-2}·s^{-1} incident light intensity. Under these conditions, the consortia form a dense monolayer biofilm on the inner surface of the culture vessel, which can be harvested separately from the accompanying bacteria (Pfannes *et al.*, 2007). The pure culture of the epibiont *Chl. chlorochromatii* CaD was grown under the same conditions in 500 ml bottles. For induction experiments, the epibiont was grown (i) in medium consisting of two-thirds sterile filtered supernatant of the `C. aggregatum´ enrichment culture plus one-third 3fold concentrated K3 medium, (ii) in K3 medium containing 0.05 % (w/v) peptone or (iii) in K3 medium supplemented with 0.1 mM or 1 mM each of the branched chain amino acids leucine, valine or isoleucine.

Genome sequencing

The draft genome sequence of the epibiont was determined by the Joint Genome Institute (U. S. Department of Energy, Walnut Creek, CA) by Sanger sequencing of the ends of 12,718 small insert (~3 kb), 10,437 medium insert (~6 kb) and 1909 fosmid (~36 kb) clones. The resulting sequence data were assembled using "phred, phrap, consed" software package (Ewing and Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998), which produced an initial assembly of ~40 contigs. The contig arrangement was deduced from the information contained in paired-end reads and split genes (Yu *et al.,* 2002), and this led to a predicted assembly containing only four scaffolds. All remaining gaps were closed by sequencing of PCR amplicons, the primers for which were designed on the basis of the predicted contig arrangement. Combinatorial PCR was performed to close the four gaps for which no scaffolding clones existed. Regions of low sequence quality were amplified by PCR and resequenced to improve and verify the sequence. The final, completed genome consisted of a single, circular DNA molecule containing 2,572,079 bp (GenBank Accession NC_007514). Annotation of the resulting genomic information was performed as described (Larimer *et al.*, 2003).

In silico subtractive hybridisation analysis

In silico subtractive hybridization was conducted with the Phylogenetic Profiler available at the DOE Joint genome Institute website (http://img.jgi.doe.gov). The *Chl. chlorochromatii* CaD genome (http://genome.jgi-psf.org/finished_microbes/chlag/chlag.home.html) was screened for single genes which had no homologs based on BLASTP alignments against the other 11 publicly available genome sequences of the green sulfur bacteria *Chlorobaculum parvum* NCIB 8327, *Cba.*

tepidum ATCC 49652[T], *Chl. ferrooxidans* DSM 13031[T], *Chl. limicola* DSM 245[T], *Chl. luteolum* DSM 273[T], *Chl. phaeobacteroides* BS1, *Chl. phaeobacteroides* DSM 266[T], *Chl. phaeovibrioides* DSM 265, *Chl. clathratiforme* DSM 5477[T], *Chloroherpeton thalassium* ATCC 35110[T] and *Prosthecochloris aestuarii* DSM 271[T] (http://img.jgi.doe.gov/cgi-bin/geba/main.cgi?section=TaxonList&page=taxonListPhylo&pidt=14955.1250667420). A maximum e-value of 10^{-5} and a minimum identity of 30% were applied for identification of homologs.

Two-dimensional difference gel electrophoresis (2-D DIGE) of the soluble proteome

For the extraction of soluble proteins cells were harvested by centrifugation (at 10,000 x g, 30 min and at 4°C), resuspended in 10 mM Tris-HCl buffer containing 1 mM phenylmethanesulfonyl fluoride (PMSF) and broken by five subsequent passages through a French press cell at 16,000 psi. After clarifying the crude extract by centrifugation (20,000 x g, 30 min and 4°C), the supernatant was centrifuged at 200,000 x g for 1 h at 4 °C. Concentrations of the soluble proteins in the supernatant were measured in triplicates using the BCA Protein Assay Reagent (Thermo Scientific Pierce, Rockford, USA). Since intact consortia consist of 16±3 epibiont cells per one central bacterium (biomass ratio ~21:1) in the cultures used for proteome analyses (Wanner *et al.*, 2008), the majority of soluble proteins in the gels were expected to originate from the epibiont. However, the potential presence of proteins of the central bacterium precluded the exact quantification of individual protein spots. Therefore we focused our analysis on only those proteins which were present in the symbiotic state of the epibiont but not detectable in axenic cultures, or vice versa. The two different fluorescently labelled protein extracts were compared directly on a single gel to avoid gel-to-gel variations.

For comparative analyses, 200 µg of soluble protein in 10 mM Tris/HCl (pH 8.5) each derived from the epibiont pure culture or the consortia culture were stained separately for 20 min at 4°C in the dark with 320 pmol of either CyDye™ Cy5 or Cy3 (DIGE Flours for Ettan DIGE , minimal dyes, GE Healthcare, Amersham, United Kingdom). The reaction was stopped by adding 1 µl 20 mM lysine solution and incubating the resulting solution in the dark at 4 °C for 10 min. After labelling, both samples were pooled, mixed with 350 µl rehydration buffer (9 M urea, 2 M thiourea, 4 % CHAPS, 1 % DTT, 12.5 µM EDTA) and incubated for 30 min at room temperature. To remove insoluble protein, the samples were centrifuged briefly for 1 min at 14,000 × *g* . Then 3.5 µl of carrier ampholytes (Merck, Darmstadt, Germany) and 1.8 µl of a 0.2% (w/v) bromophenol blue solution as tracing solution were added. A total amount of 400 µg protein in 400 µl was immediately subjected to isoelectric focusing (IEF) on 18 cm long IPG strips spanning a linear pH gradient of 4 - 7 (Immobiline™ DryStrip; GE Healthcare). The IPG strips were covered with 1.5 ml

mineral oil. The voltage settings for IEF as performed with the IPphor™ II isoelectric focusing system (GE Healthcare) were as follows: (1) 30 V for 10 h, (2) a linear increase to 200 V over 20 min, (3) 200 V for 1 h, (4) a linear increase to 500 V over 30 min, (5) 500V for 1 h, (6) a linear increase to 2000 V for 2 h, (7) 2000 V for 2 h, a linear increase to 8000 V for 2 h, followed by a hold until a total of 80 kVh was reached. After IEF the IPG strips were incubated for 15 min with gentle shaking in 10 ml equilibration buffer (50 mM Tris-HCl pH 6.8, 6 M urea, 2 % w/v SDS and 30 % v/v glycerol) containing 1 % (w/v) DTT in the first, and 2.5 % (w/v) iodacetamide in the second step. For second dimension protein separation an Ettan Dalt II electrophoresis system (GE Healthcare) was used. IPG strips were placed on top of polyacrylamide gels (180 x 250 x 1 mm) composed of a 4.8 % stacking gel containing 125 mM Tris-HCl (pH 6.8) and a 12.5 % separating gel containing 375 mM Tris-HCl (pH 8.8) and 4 M urea. 10 µl of Cy3/Cy5 labelled wide range protein standard (ECL Plex Fluorescent Rainbow marker, GE Healthcare) was provided by a piece of filter paper (3 x 5 x 30 mm) next to the IPG strip. Gels were overlain with 0.5 % (w/v) low melting agarose solution. Gels were run at 15°C with a running buffer composed of 25 mM Tris, 192 mM glycine and 0.1% (w/v) SDS. Electrophoresis was conducted at 600 V and 18 mA per gel over night and stopped when the bromophenol blue marker reached the bottom of the gel.

For each experimental condition tested, three biological replicas (independent cultures) were extracted and each culture analysed in duplicate gels (technical replicates). Overall, 24 parallel gels were thus analyzed. Afterwards, the CyDye™ labeled protein gels were scanned directly between the low-fluorescence glass plates using a Typhoon Trio scanner (GE Healthcare) with a resolution of 100 µm and the photomultiplier tube set at 600 V. After scanning, the 2D gels were stained with colloidal coomassie (Neuhoff *et al.*, 1988) and dried on 3 mm Whatman paper (GE Healthcare) in a gel dryer. DeCyder software package (version 4.0; GE Healthcare) was used for determining the total spot numbers and for a refined control of spot boundaries. Comparative, qualitative image analysis was performed manually using ImageQuant v5.2 (Molecular Dynamics, Sunnyvale, USA). Only protein spots which were detected exclusively either in consortia extracts, in extracts of pure epibionts, or in extracts of supplemented pure epibiont cultures were selected for subsequent protein identification. Spots were excised manually from dried, colloidal coomassie stained gels.

Membrane proteins

Membrane fractions of free-living epibionts were compared with those of untreated consortia in order to identify membrane proteins potentially involved in the cell-cell interactions. Parallel samples were treated with the cross-linkers bis[sulfosuccinimidyl] suberate (BS[3]) and 3,3´-dithiobis[sulfosuccinimidylpropionate] (DTSSP) (Thermo Scientific Pierce) which both comprise a 12 Å spacer. Since both cross-linking agents are water-soluble and cannot pass the cytoplasmic

membrane, they only react with extracellular or periplasmic proteins (Luethy *et al.*, 1995). BS[3] is non-cleavable and DTSSP is thiol-cleavable. A fresh consortia biofilm was pelleted and dissolved in 5 ml anoxic HEPES buffer (5 mM; pH 7.5) containing 5 mM BS[3] or 2 mM DTSSP. After incubation for 30 min at room temperature, Tris-HCl (pH 7.5) was added to a final concentration of 50 mM and incubated for 15 min to quench the cross-linking reaction. To test the success of cross-linking epibionts and its central rod, EGTA was added to cross-linked consortia in a final concentration of 20 mM, since the cells in phototrophic consortia disaggregate in the presence of EGTA (Vogl *et al.*, 2008).

For extraction of membrane proteins, cells were harvested and broken as described above. The homogenates were clarified by centrifugation at 20,000 × *g* for 30 min at 4°C and the membrane fractions were pelleted by ultra-centrifugation at 200,000 × *g* for 1 h at 4 °C. After solubilization of proteins with 2% (w/v) SDS at 45 °C for 30 min, the extracts were centrifuged again at 200,000 × *g* for 1 h and solubilized proteins in the supernatants were precipitated by adding 9 volumes of acetone, followed by incubation at 0 °C for 16 h. Membrane proteins were collected by centrifugation at 20,000 × *g* for 30 min, the pellets were washed once with acetone and then re-suspended in 10 mM Tris-HCl buffer containing 1 mM PMSF and 1 % SDS. Protein concentrations were measured in triplicates using the BCA Protein Assay Reagent (Thermo Scientific Pierce). To break the cross-links in control assays, DTSSP-cross-linked consortia proteins were treated with 5% β-mercaptoethanol for 10 min at 56°C prior to gel electrophoresis. 45 μg of each membrane protein extract and 3 μl of protein standard (ECL Plex Fluorescent Rainbow marker, GE Healthcare) were separated by Glycine-SDS-PAGE (Laemmli, 1970) in 10 % (w/v) polyacrylamide gels. Electrophoresis was conducted at room temperature and 20 mA for 16 h. Gels were stained with colloidal Coomassie (Neuhoff *et al.*, 1988).

Isolation of chlorosomes and analysis of chlorosome proteins

Cultures of "*C. aggregatum*", *Chl. chlorochromatii* CaD and *Cba. tepidum* ATCC 49652[T] were harvested in the exponential growth phase by centrifugation at 10,000 × *g* and 4°C. Cells were resuspended in isolation buffer (10 mM Tris-HCl, 0.5 M betainhydrochloride, 0.5 mM PMSF and 1mM dithiothreitol; pH 7.4), and then treated with lysozyme (3 mg·ml[-1]) for 20 min. Cells were disrupted by two subsequent sonication steps (each for 5 min, 50% pulse and 45% power; Cell Disruptor B15, Branson, Danburry, USA) with intermittent cooling on ice (5 min). After a centrifugation at 12,000 × *g* for 20 min, the supernatant was loaded onto a 10 − 60 % sucrose gradient. The sucrose gradients were centrifuged at 26,000 × *g* for 18 h. This procedure yielded a visible green band containing the chlorosomes at about 30% sucrose. The entire isolation procedure was performed in dim light and at 4°C.

Fluorescence spectroscopy was employed to assess whether chlorosomes had remained intact during the isolation procedure. The fluorescence of free bacteriochlorophyll *c* (BChl *c*) shows a maximum at 666 nm and is much more intense than the fluorescence of aggregated BChl *c* in intact chlorosomes which fluoresce at an emission maximum of 766 nm (Bryant *et al.*, 2002). Therefore damages to the chlorosomes causing release of BChl *c* can be detected by fluorescence analysis. The chlorosome fraction was diluted 20-fold with isolation buffer and fluorescence was measured at 450 nm (FluoroMax-3, Horiba, Kyoto, Japan). The emission spectrum showed the distinct maximum at 766 nm typical for chlorosome BChl *c* aggregates, indicating that chlorosomes had remained intact during the isolation process. Electron microscopy was performed as an additional control for the integrity of chlorosomes. For negative contrasting, the chlorosome-containing fractions were diluted with water in a ratio of 1:25 and dropped on a collodium coated, carbon steamed grid. After 1 min, the fraction was removed and the grid was dried at room temperature. Subsequently, one drop of uranyl acetate (1% w/v) was applied, the grid incubated for 1 min, dried again at room temperature and viewed in a Zeiss EM 912 electron microscope (Zeiss, Oberkochen, Germany) with an OMEGA filter in zero-loss mode. Electron microscopic inspection demonstrated that chlorosomes derived from free-living as well as from symbiotic epibionts exhibited a smooth surface and the ellipsoid shape typical for an intact chlorosome membrane.

20 to 100 μg of each chlorosome protein extract and 8 μl of protein marker (Polypeptide SDS-Page Standards, Bio-Rad Laboratories, Hercules, USA) were separated by Tricine-SDS-PAGE (Schägger and von Jagow, 1987) in 12 % (w/v) polyacrylamide gels. Electrophoresis was conducted at 4°C and 40 mA for 20 h and afterwards the gels were stained either with colloidal Coomassie (Neuhoff *et al.*, 1988) or silver (Blum *et al.*, 1987).

After identification of epibiont chlorosome proteins by mass spectrometry (see below), the amino acid sequence *Chl. chlorochromatii* CaD Cag_1285 was subjected to Expasy "Fasta3" (www.expasy.ch) and BLASTP (http://blast.ncbi.nlm.nih.gov) similarity searches as well as an IMG othologous cluster analysis (http://img.jgi.doe.gov). The sequence of Cag_1285 and the 12 most closely related amino acid sequences recovered from the databases were imported into ARB (Ludwig *et al.*, 2004) and aligned with the ClustalW Protein alignment tool. Maximum-likelihood trees were calculated based on 60 amino acid positions using the ProteinML program implemented in ARB and applying the Dayhoff amino acid substitution model matrix. The robustness of trees was inferred by bootstrap analysis after 100 resamplings using PhyML.

Mass spectrometry

Protein bands and spots were excised manually from one- or two-dimensional gels and digested with modified trypsin (Promega, Heidelberg, Germany) using OMX-S according to the protocol of the manufacturer (OMX GmbH, Wessling, Germany) (Granvogl *et al*., 2007). All samples were analyzed by nano-LC-ESI-MS/MS on a quadrupole time-of-flight tandem mass spectrometer (Micromass ESI Q-TOF Premier, Waters, Manchester, United Kingdom). For trapping, samples were loaded onto a 5 μm Symmetry C18 180 μm x 20 mm column (Waters). Loading and washing of the sample on the trapping column was achieved at a flow rate of 5 μl·min^{-1} for 3 min. Separation of peptide mixtures was achieved by nanoAcquity Ultra Performance Liquid Chromatography using a 1.7 μm BEH130 C18, 75 μm x 100 mm reversed phase nano column (Waters). Peptides were eluted by running a linear gradient using 100% H_2O, 0.1% formic acid as solvent A and 100% acetonitrile, 0.1% formic acid as solvent B, at a flow rate of 0.4 μl/min within 60 min. For protein identification, sequence tags obtained from the fragment spectra were used in protein database searches. Proteins were identified by similarity searches (BLAST) of the amino acid sequence tags in the frame "fasts3" from the European Bioinformatics Institute (EBI, www.ebi.ac.uk/fasta33). The MS and MS/MS spectra were used to search the latest version of the SwissProt database using ProteinLynx Global Server (Waters). For manual sequence analysis of the sequence tags, the *Chl. chlorochromatii* CaD database was used, which was not linked to the ProteinLynx Global Server. Only proteins exactly matching predicted proteins of *Chl. chlorochromatii* CaD were counted as positives. False positives were excluded by checking the MS and MS/MS spectra manually.

Reverse transcription, quantitative real-time PCR (RT-qPCR)

Chl. chlorochromatii CaD and `*C. aggregatum*´ cultures were mixed with 12.5 % ice-cold EtOH/phenol stop solution (5 % phenol pH 4.5 - 5.5 in ethanol) to avoid RNA-degradation. Total RNA was isolated using phenol-chloroform (Chromczynski and Sacci, 1987) and subsequently purified with the RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany) according to the protocol of the manufacturer. The RNA was treated with Turbo DNA free (Applied Biosystems, Forster City, USA) to remove all remaining DNA contamination. RNA concentrations were determined spectrophotometrically at 260 nm in a NanoDrop 1000 (Thermo Fisher Scientific NanoDrop, Wilmington, USA) and quality was assessed on a formaldehyde gel (3.1 %, w/v). As a highly sensitive test for the absence of genomic DNA, a step-down PCR with a primer set targeting the *rpoD*-gene (Cag_0387) which codes for the RNA polymerase sigma factor A of *Chl. chlorochromatii* CaD (Vogl *et al*., 2008) was conducted using 200 ng of the total RNA preparation.

Reverse transcription was performed with Superscript III Reverse Transscriptase (Invitrogen) using 400 ng RNA and 200 ng random hexamer primer (Eurofins MWG, Ebersberg, Germany) as

recommended by the supplier. Negative controls to verify absence of contaminating genomic DNA were prepared by omitting the reverse transcriptase. RT-qPCR reactions with custom designed gene specific primers were run at optimized PCR conditions (Suppl. Table S3) in an iQ5 real-time PCR detection system (Bio-Rad) using 12.5 µl iQ SYBR Green Supermix (Bio-Rad), 10 ng cDNA and 160 nM of each primer in a final volume of 25 µl. Tenfold serial dilutions of the cDNA were used to construct standard curves for PCR efficiency determination. Transcript quantities of each target gene were normalized to the transcript quantities of the housekeeping gene *rpoD* (primers rpoDf and rpoDr; Suppl. Table S3). Changes in the relative abundance of transcripts between the symbiotic and free-living state of the epibiont were determined using the Pfaffl method (Pfaffl, 2001).

Prokaryotic cDNA suppression subtractive hybridisation (cDNA-SSH)

Total RNA was isolated and DNase-treated as described above. The separation of the mRNA from 40 µg of total RNA was performed with the MICROBExpress bacterial mRNA enrichment kit (Applied Biosystems) according to the instructions of the manufacturer. Multiple preparations were pooled and concentrated with the RNeasy MinElute Cleanup Kit (Qiagen). The efficiency of rRNA depletion in the mRNA extract was analyzed using an Agilent 2100 bioanalyzer with a RNA LabChip (Agilent Technologies, Santa Clara, USA).

First strand cDNA synthesis was done with Superscript III Reverse Transscriptase (Invitrogen) as described by the manufacturer, using 2 µg of mRNA 0.5 µl of 10 µM PCS primer (De Long *et al.*, 2008) and 400 U reverse transcriptase. After 90 min of incubation, additional 400 U of the enzyme were added and the incubation continued for another 90 min. Two first strand synthesis reactions were pooled and the second strand cDNA synthesis performed with the PCR-Select™ cDNA substraction kit (Clontech Laboratories, Mountain View, USA) according to the manufacturers instructions. One microliter of DNase-free RNase (500 µg·ml^{-1}) (AppliChem, Darmstadt, Germany) was added and the samples incubated for 30 min at 37°C. The cDNA was purified with a QIAquick PCR purification kit (Qiagen) and yields were quantified by absorbance at 260 nm. Tester and driver were digested with RsaI and purified using the MinElute reaction cleanup kit (Qiagen). Successful digestion of the cDNA was verified by gel electrophoresis on a 1% agarose gel stained with ethidium bromide. Adaptor ligation and the ligation efficency test using the primer pair rpoDf/rpoDr (Suppl. Table S3) were performed according to the manual for the PCR-Select™ cDNA substraction kit (Clontech). The first and second hybridisations were done following the instructions of the kit with the exception that 3 µl denaturated driver and 1 µl 4x hybridisation buffer were used during the second hybridisation. Primary and secondary nested suppression PCRs were run in a GeneAmp 9700 Thermal Cycler (Applied Biosystems) with Advantage 2 cDNA

polymerase mix (Clontech) according to the PCR-Select™ cDNA substraction kit (Clontech) protocol using 50 nM of nested-PCR primer 2R in the secondary PCR. Amplification products of this secondary PCR were cloned using the TOPO TA cloning kit (Invitrogen). Plasmids were isolated from selected clones and inserts were sequenced with M13 forward and reverse primers (Invitrogen). Sequence analysis was performed on basis of the annotated genome sequence of *Chl. chlorochromatii*. The complete open reading frames (ORFs) of the gene fragments of the DNA sequences obtained by cDNA subtractive hybridization were obtained from the DOE Joint genome Institute website (http://img.jgi.doe.gov).

Illumina cDNA sequencing

Aliquots of the cDNA prepared for cDNA-SSH (see above) were used for a comparison of the transcriptome of symbiotic with free-living epibionts. Prior to library construction, cDNA was fragmented using the Applied Biosystems RNA fragmentation reagents according to the instructions of the manufacturer. Libraries for sequencing were then prepared using the Illumina mRNA-Seq kit beginning with end repair of the cDNA fragments prior to adapter ligation. Sequencing was performed at the University of Delaware using the Illumina Genome Analyzer II. The loading volume for cluster generation was optimized using a qPCR assay to determine the concentration of properly constructed fragments (B. Kingham; pers. comm.).

Raw sequence data were trimmed to 30 bp and mapped to the *Chl. chlorochromatii* CaD genome (NC_007514) in the Eland software package with a mismatch tolerance of $\leq$ 2bp. Coverage of annotated protein coding genes (reference database NC_007514.ptt), RNA coding genes (reference database NC_007514.rnt), and intergenic regions > 50 bp was determined using custom Perl scripts (available from T. Hanson on request) to calculate the number of sequence tags that mapped to each genomic region. The library size for symbiotic epibiont cells was 14,117 tags and that for free-living epibionts 33,901 tags. Accordingly, the expression ratio of genes was calculated as: (tags per gene in consortium library/14,117)/(tags in free living library/33,901). Differentially expressed genomic regions were defined as those with an expression ratio of $\leq$ 0.67 or $\geq$ 1.5 and resulting in a *p*-value of < 0.01 when calculated as described by Audic and Claverie (1997) using the number of non-ribosomal matches to the genome as the library size for each sample. For sequences present in only one library, the *p*-value criterion ensured that at least 4 sequence tags matching that gene were observed to include it in the differentially expressed list.

Acknowledgements

References

Audic, S., Claverie, J.M. (1997) The significance of digital gene expression profiles. *Genome Res* 7: 986-995

Balakrishnan, L., Hughes, C., Koronakis, V. (2001) Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by *Escherichia coli*. *J Mol Biol* 313: 501-510

Behrens, S., Lösekann, T., Pett-Ridge, J., Weber, P.K., Wing-On, N., Stevenson, B.S., Hutcheon, I.D., Relman, D.A., Spormann, A.M. (2008) Linking microbial phylogeny to metabolic activity at the single-cell level by using enhanced element labeling-catalyzed reporter deposition fluorescence in situ hybridization (EL-FISH) and NanoSIMS. *Appl Environ Microbiol* 74: 3143-3150

Beller, H.R., Letain, T.E., Chakicherla, A., Kane, S.R., Legler, T.C., Coleman, M.A. (2006) Whole-genome transcriptional analysis of chemolithoautotrophic thiosulfate oxidation by *Thiobacillus denitrificans* under aerobic versus denitrifying conditions. *J Bacteriol* 188: 7005-7015

Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., Melendrez, M.C., Bateson, M.M., Ward, D.M., Heidelberg, J.F. (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J* 1: 703-713

Blum, H., Beier, H., Gross, H.J. (1987) Improved silver staining for plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8: 93-99

Boetius, A., Ravenschlag, K., Schubert, C.J., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jørgensen, B.B., Witte, U., Pfannkuche, O. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* 407: 623-626

Bowe, F., Lipps, C.J., Tsolis, R.M., Groisman, E., Heffron, F., Kusters, J.G. (1998) At least four percent of the *Salmonella typhimurium* genome is required for fatal infection of mice. *Infect Immun* 66: 3372-3377

Bryant, D.A., Vassilieva, E.V., Frigaard, N.U., Li, H. (2002) Selective protein extraction from *Chlorobium tepidum* chlorosomes using detergents. Evidence that CsmA forms multimers and binds bacteriochlorophyll a. *Biochemistry* 48: 14403-14411

Burghardt, T., Näther, D.J., Junglas, B., Huber, H., Rachel, R. (2007) The dominating outer membrane protein of the hyperthermophilic Archaeum *Ignicoccus hospitalis*: a novel pore-forming complex. *Mol Microbiol* 63: 166-176

Burghardt, T., Junglas, B., Siedler, F., Wirth, R., Huber, H., Rachel, R. (2009) The interaction of *Nanoarchaeum equitans* with *Ignicoccus hospitalis*: proteins in the contact site between two cells. *Biochem Soc Trans* 37: 127-132

Candela, T., Fouet, A. (2006) Poly-gamma-glutamate in bacteria. *Mol Microbiol* 60: 1091-1098

Chhabra, S.R., He, Q., Huang, K.H., Gaucher, S.P., Alm, E.J., He, Z., Hadi, M.Z., Hazen, T.C., Wall, J.D., Zhou, J., Arkin, A.P., Singh, A.K. (2006) Global analysis of heat shock response in *Desulfovibrio vulgaris* Hildenborough. *J Bacteriol* 188: 1817-1828

Chromczynski, P., Sacci, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroformextraction. *Anal Biochem* 162: 156-159

Daines, D.A., Jarisch, J., Smith, A.L. (2004) Identification and characterization of a nontypeable *Haemophilus influencae* putative toxin-antotoxin locus. *BMC Microbiol* 4: 30

Dalisay, D.S., Webb, J.S., Scheffel, A., Svenson, C., James, S., Holmström, C., Egan, S., Kjelleberg, S. (2006) A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga *Ulva australis*. *Microbiology* 152: 2875-2883

D'Andrea, L.D., Regan, L. (2003) TPR proteins: the versatile helix. *Trends Biochem Sci* 28: 655-662

De Bok, F.A., Plugge, C.M., Stams, A.J. (2004) Interspecies electron transfer in methanogenic propionate degrading consortia. *Water Res* 38: 1368-1375

De Long, S.K., Kinney, K.A., Kirisits, M.J. (2008) Prokaryotic suppression subtractive hybridisation PCR cDNA substraction, a targeted method to identify differentially expressed genes. *Appl Environ Microbiol* 74: 225-232

Driessen, A.J.M., Nouwen, N. (2008) Protein translocation across the bacterial cytoplasmatic membrane. *Annu Rev Biochem* 77: 643-667

Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M. et al. (2002) The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc Natl Acad Sci USA* 99: 9509-9514

Eriksson, S., Lucchini, S., Thompson, A., Rhen, M., Hinton, J.C. (2003) Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol* 47: 103-118

Ewing, B., Green, P. (1998) Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8: 186-194

Ewing, B., Hillier, L., Wendl, M., Green, P. (1998) Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175-185

Forchhammer, K. (2008) PII signal transducers: novel functional and structural insights. *Trends Microbiol* 16: 65-72

Frigaard, N.-U., Bryant, D.A. (2006) Chlorosomes: antenna organelles in green photosynthetic bacteria, p. 79-144. *In* J. M. Shively (ed.) Complex Intracellular Structures in Prokaryotes, Microbiology Monographs, Vol. 2. Springer, Berlin

Fröstl, J.M., Overmann, J. (1998) Physiology and tactic response of `Chlorochromatium aggregatum´. *Arch Microbiol* 169: 129-135

Gianchandani, E.P., Joyce, A.R., Palsson, B.Ø., Papin, J.A. (2009) Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system *PLoS Comput Biol* 5: e1000403

Glaeser, J. Overmann J. (2003) The significance of organic carbon compounds for *in situ* metabolism and chemotaxis of phototrophic consortia. *Environ Microbiol* 5: 1053-1063

Glaeser, J., Overmann, J. (2004) Biogeography, evolution and diversity of the epibionts in phototrophic consortia. *Appl Environ Microbiol* 70: 4821-4830

Gordon, D., Abajian, C., Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8: 195-202

Granvogl, B., Gruber, P., Eichacker, LA. (2007) Standardization of rapid in-gel digestion by mass spectrometry. *Proteomics* 7: 642-654

Groisman, E.A, Ochman, H. (1997) How *Salmonella* became a pathogen. *Trends Microbiol* 5: 343-349

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417: 63-67

Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K.O., Rachel, R., Huber, H. (2008) *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. *J Bacteriol* 190: 1743-1750

Kajava, A.V., Cheng, N., Cleaver, R., Kessel, M., Simon, M.N., Willery, E., Jacob-Dubuisson, F., Locht, C., Steven, A.C. (2001) Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. *Mol Microbiol* 42: 279-292

Kanzler, B.E.M., Pfannes, K.R., Vogl, K., Overmann, J. (2005) Molecular characterization of the non-photosynthetic partner bacterium in the consortium `*Chlorochromatium aggregatum*´. *Appl Environ Microbiol* 71: 7434-7441

Krämer, R. (1994) Systems and mechanisms of amino acid uptake and excretion in prokaryotes. *Arch Microbiol* 162: 1-13

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685

Larimer, F.W., Chain, P., Hauser, L., Lamderdin, J., Malfatti, S. *et al.* (2003) Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nature Biotechnol* 22: 55-61

Lauterborn, R. (1906) Zur Kenntnis der sapropelischen Flora. *Allg Bot Z* 12: 196-197

Lodwig, E.M., Hosie, A.H., Bourdès, A., Findlay, K., Allaway, D., Karunakaran, R., Downie, J.A., Poole, P.S. (2003) Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* 422: 722-726

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

Luethy, M.H., Thelen, J.J., Knudten, A.F., Elthon, T.E. (1995) Purification, characterization, and submitochondrial localization of a 58-kilodalton NAD(P)H dehydrogenase. *Plant Physiol* 107: 443-450

Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G250 and R250. *Electrophoresis* 9: 255-262

Nikodinovic-Runic, J., Martin, L.B., Babu, R., Blau, W., O'Connor, K.E. (2009) Characterization of melanin-overproducing transposon mutants of *Pseudomonas putida* F6. *FEMS Microbiol Lett* 298: 174-183

Overmann, J., Tuschak, C., Sass, H., Fröstl, J.M. (1998) The ecological niche of the consortium "*Pelochromatium roseum*" *Arch Microbiol* 169: 120-128

Overmann, J., Schubert, K. (2002) Phototrophic consortia, model systems for symbiotic interrelations between prokaryotes. *Arch Microbiol* 177: 201-208

Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J. *et al.* (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409: 529-533

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real time RT-PCR. *Nucleic Acids Res* 29: e45

Pfannes, K.R., Vogl, K., Overmann J. (2007) Heterotrophic symbionts of phototrophic consortia: members of a novel diverse cluster of betaproteobacteria characterized by tandem rrn operon structure. *Environ Microbiol* 9: 2782-2794

Prell, J., White, J.P., Bourdes, A., Bunnewell, S., Bongaerts, R.J., Poole, P.S. (2009) Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc Natl Acad Sci USA* 106: 12477-12482

Pysz, M.A., Conners, S.B., Montero, C.I., Shockley, K.R., Johnson, M.R., Ward, D.E., Kelly, R.M. (2004) Transcriptional analysis of biofilm formation processes in the anaerobic, hyperthermophilic bacterium *Thermotoga maritime*. *Appl Environ Microbiol* 70: 6098-6112

Read, T.D., Peterson, S.N., Tourassse, N., Baillie, L.W., Paulsen, I.T., *et al.* (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423: 81-86

Relman, D.A., Domenighini, M., Tuomanan, E., Rappuoli, R., and Falkow, S. (1989) Filamentous hemagglutinin of *Bordetella pertussis*: Nucleotide sequence and crucial role in adherence. *Proc Natl Acad Sci USA* 86: 2637-2641

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S. Chain, P. *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424: 1042-1046

Schägger, H., von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range of 1 to 100 kDa. *Anal Biochem* 166: 368-379

Schink, B. (2002) Synergistic interactions in the microbial world. *Antonie Leeuwenhoek* 81: 257-261

Sirevag, R., Ormerod, J.G. (1970) Carbon dioxide-fixation in photosynthetic green sulfur bacteria. *Science* 169: 186-188Van der Veen, S., Hain, T., Wouters, J.A., Hossain, H., de Vos, W.M., Abee, T., Chakraborty, T., Wells-Bennik, M.H. (2007) The heat-shock response of *Listeria monocytogenes* comprises genes involved in heat shock, cell division, cell wall synthesis, and the SOS response. *Microbiology* 153: 3593-3607

Viklund, H., Bernsel, A., Skwark, M., Elofsson, A. (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* 24: 2928-2929

Vogl, K., Glaeser, J., Pfannes, K.R., Wanner, G., Overmann, J. (2006) *Chlorobium chlorochromatii* sp. Nov., a symbiotic green sulfur bacterium isolated from the phototrophic consortium `*Chlorochromatium aggregatum*´. *Arch Microbiol* 185: 363-372

Vogl, K., Wenter, R., Dressen, M., Schlickenrieder, M., Plöscher, M., Eichacker, L.A., Overmann, J. (2008) Identification and analysis of four candidate symbiosis genes from

`*Chlorochromatium aggregatum´*, a highly developed bacterial symbiosis. *Environ Microbiol* 10: 2842-2856

Wanner, G., Vogl, K., Overmann, J. (2008) Ultrastructural characterization of the prokaryotic symbiosis in `*Chlorochromatium aggregatum´. J Bacteriol* 190: 3721-3730

Wenter, R., Wanner, G., Schüler, D., Overmann, J. (2009) Ultrastructure, tactic behavior and potential for sulfate reduction of a novel multicellular magnetotactic prokaryote from North Sea sediments. *Environ Microbiol* 11: 1493-1505

Whittaker, C.J., Klier, C.M., Kolenbrander, P.E. (1996) Mechanisms of adhesion by oral bacteria. *Annu Rev Microbiol* 50: 513-552

Yousef F., Espinosa-Urgel, M. (2007) *In silico* analysis of large microbial surface proteins. *Res Microbiol* 158: 545-550

Yu, Z., Li, T., Zhao, J., Luo, J. (2002) PGAAS: a prokaryotic genome assembly assistant system. *Bioinformatics* 18: 661-665

Supplementary Material
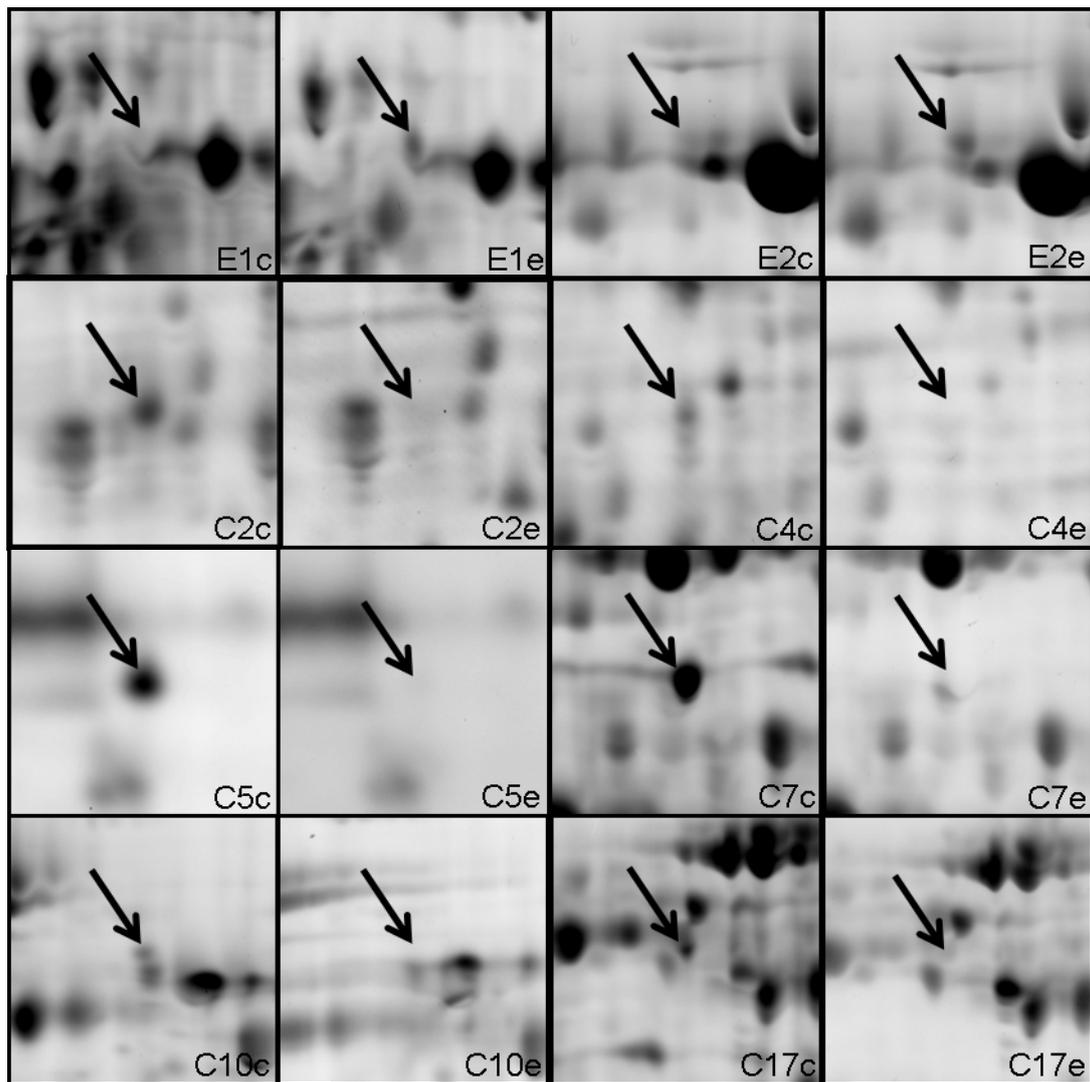
Suppl. Fig. S1. Examples of qualitative 2-D DIGE spot analysis of pure culture (e) and consortia (c) epibionts only present in non-symbiotic (E) and symbiotic (C) state.
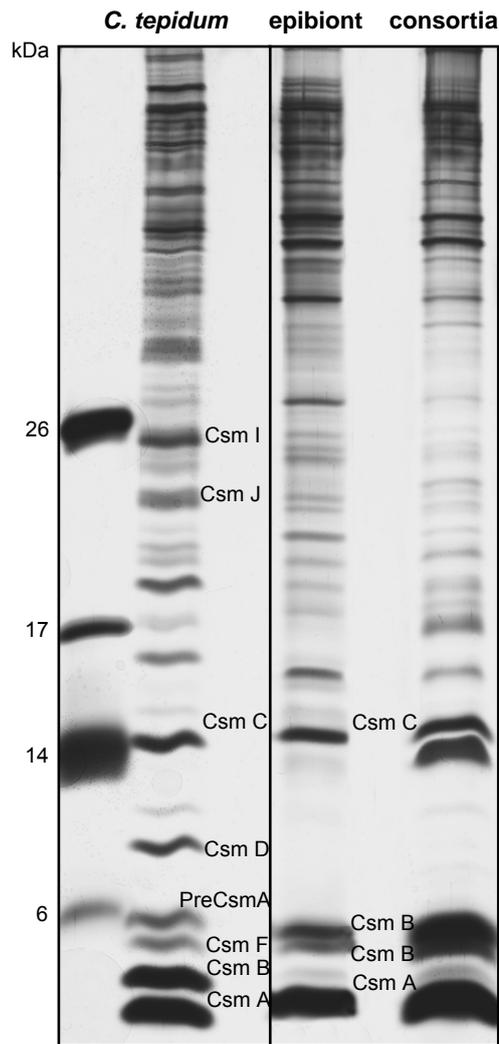
Suppl. Fig. S2. SDS-PAGE of chlorosome membrane proteins of the epibiont *Chl. chlorochromatii* CaD in free-living and consortia state. Chlorosome extracts of *Cba. tepidum* ATCC 49652[T] served as a reference. The arrow indicates an uncharacterized protein of the consortia epibiont representing the only qualitative change in the protein pattern compared to the pure culture.

Suppl. Table S1: 186 unique ORFs of the epibiont *Chlorobium chlorochromatii* CaD genome identified by *in silico* subtractive hybridisation based on BLASTP alignments against 11 other green sulfur bacteria genomes. ORFs which were found to be differentially regulated are marked in bold type.

Locus Tag	ORF name	Amino acids	Function prediction	Signal Peptide	Transmembrane Helices
Cag_0016	Hypothetical protein	124	Nucleotidyltransferase	No	No
Cag_0033	Hypothetical protein	422	ATP binding	No	No
Cag_0042	Hypothetical protein	634	ATP-dependent endonuclease; OLD family	No	No
Cag_0134	Hypothetical protein	238		No	Yes
Cag_0135	Hypothetical protein	54		No	No
Cag_0136	Hypothetical protein	31		No	No
Cag_0144	ATPase	565	ATPase of ABC-transporter	No	No
Cag_0152	Hypothetical protein	158	ATPase; involved in DNA repair	No	No
Cag_0153	Hypothetical protein	418		No	No
Cag_0256	Hypothetical protein	167	ATP binding; involved in virulence	No	No
Cag_0266	Protein of unknown function DUF132	141	Nucleic acid binding; PilT-like PIN domain protein	No	No
Cag_0269	Hypothetical protein	179	Extracellular solute binding protein of ABC-transporter	No	No
Cag_0278	Hypothetical protein	230		No	No
Cag_0279	Hypothetical protein	67		No	No
Cag_0281	Hypothetical protein DUF1255	103	TonB box dependent vitamin B12 binding receptor; transport	No	No
Cag_0300	Hypothetical protein	136		Yes	No
Cag_0308	Hypothetical protein	639	Glutamate-cysteine ligase; glutamate metabolism	No	No

Cag_ID	Protein	Size	Function		
Cag_0309	Hypothetical protein	392	Amino acid transport and metabolism; DNA binding metalloprotease	No	No
Cag_0311	Hypothetical protein	628	Abortive infection phage resistance (AIPR)	No	No
Cag_0312	Hypothetical protein	86		No	No
Cag_0325	Hypothetical protein DUF 262/1542	569		No	No
Cag_0368	Hypothetical protein	287		No	No
Cag_0376	Hypothetical protein	178		No	No
Cag_0378	Helicase domain protein	475	RNA metabolism	No	No
Cag_0380	DEAD/DEAH box helicase-like	1301	RNA metabolism	No	No
Cag_0396	Hypothetical protein	180		Yes	Yes
Cag_0424	Drug-proton antiporter	407	Small solutes/small peptide gradient transport	Yes	Yes
Cag_0427	Hypothetical protein	205	WD40 repeat, protein-protein interaction and complexes	Yes	No
Cag_0434	Hypothetical protein	259		No	No
Cag_0445	Hypothetical protein	431		No	No
Cag_0507	Hypothetical protein	614	Von Willenbrand factor; protein bindung and signalling	No	No
Cag_0519	Hypothetical protein	394		No	No
Cag_0532	Hypothetical protein	287		No	Yes
Cag_0544	Hypothetical protein	188		Yes	No
Cag_0546	Hypothetical protein	188	Outer membrane porin	Yes	No
Cag_0553	Hypothetical protein (Heliobacterium)	218	FeS protein	No	No
Cag_0561	Hypothetical protein	83		No	No
Cag_0576	Hypothetical protein	73		No	No

Cag_0577	**Hypothetical protein**	**312**		**No**	**No**
Cag_0582	Hypothetical protein	134		No	No
Cag_0584	Transcriptional regulator, CopG family	91	Transscription	No	No
Cag_0599	Hypothetical protein	277		No	No
Cag_0600	Hypothetical protein	97		No	No
Cag_0615	**Outer membrane protein-like**	**577**	**Proper expression of outer membrane proteins**	**Yes**	**No**
Cag_0616	Parallel beta-helix repeat protein	20646	Haemagglutinin-like giant protein involved in cell-cell adhesion	No	Yes
Cag_0646	**Hypothetical protein**	**157**		**Yes**	**No**
Cag_0648	Periplasmic protein involved in polysaccharide export	270	TonB box ; transmembrane export	Yes	Yes
Cag_0649	Exopolysaccharide biosynthesis protein	806	Capsular exopolysaccharide biosynthesis	No	Yes
Cag_0650	Exopolysaccharide biosynthesis protein, CapC family	239	Capsular exopolysaccharide biosynthesis	No	No
Cag_0655	Hypothetical protein	79		No	No
Cag_0663	Hypothetical protein	104		No	No
Cag_0665	Exopolysaccharide biosynthesis protein	426	O-antigen/teichonic acid export	Yes	Yes
Cag_0666	Hypothetical protein	233	Methylase in ubi-/menaquinone biosynthesis	No	No
Cag_0667	Hypothetical protein	584		No	No
Cag_0668	Poly-gamma-glutamate biosynthesis enzyme	377	Poly-gamma-glutamate capsule formation; virulence/adhesion	No	No
Cag_0670	Hypothetical protein	273		No	No
Cag_0673	Glycosyltransferase	327		Yes	No

Cag ID	Annotation	Length	Description		
Cag_0674	Hypothetical protein	321		Yes	Yes
Cag_0675	Glycosyl transferase	245	Cell wall biogenesis	No	No
Cag_0680	Hypothetical protein	119		No	No
Cag_0682	Transposase	382	IS66 family	No	No
Cag_0683	Hypothetical protein	165		No	No
Cag_0690	**Hypothetical protein**	**565**		**No**	**No**
Cag_0705	Hypothetical protein	126		No	No
Cag_0706	Hypothetical protein	77		No	No
Cag_0709	Hypothetical protein	54		No	No
Cag_0714	Hypothetical protein	68		No	No
Cag_0718	Plasmid segregation, centromere-binding protein	83	ParG-like	No	No
Cag_0720	Hypothetical protein DUF155	259		No	No
Cag_0733	Hypothetical protein	496	N-acetyltransferase	No	No
Cag_0734	Hypothetical protein	159	N-acetyltransferase	No	No
Cag_0735	Hypothetical protein	180		Yes	No
Cag_0736	Hypothetical protein	132		No	No
Cag_0737	Nucleotidyltransferase	107		No	No
Cag_0739	Hypothetical protein	235		Yes	No
Cag_0745	Glycosyl transferase, family 25	244	LPS biosynthesis	No	No
Cag_0746	Hypothetical protein	252	Glycosyltransferase	No	Yes
Cag_0747	Hypothetical protein	230	Methylase in ubiquinone biosynthesis	No	No

Cag_0749	Hypothetical protein	173	Methyltransferase	No	No
Cag_0750	Hypothetical protein	183		No	No
Cag_0751	Hypothetical protein	368		No	No
Cag_0759	Hypothetical protein	912	DNA Methylase/Restriction enzyme type III	No	No
Cag_0760	DEAD/DEAH box helicase-like	646	RNA/DNA helicase	No	No
Cag_0761	Hypothetical protein	398	ATP-dependend DNA ligase	No	No
Cag_0766	Hypothetical protein	85		Yes	No
Cag_0767	Abortive infection phage resistance protein	567	Restriction modification operon	No	No
Cag_0771	Hypothetical protein)	163		Yes	No
Cag_0787	**Alkaline phosphatase**	**444**	**Inorganic ion transport; Folate biosynthesis; 2-component system**	**Yes**	**No**
Cag_0790	Hypothetical protein	512	Carbohydrate kinase; conjugative transfer	Yes	No
Cag_0818	Hypothetical protein	130	PilT domain	No	No
Cag_0819	Hypothetical protein	79		No	No
Cag_0834	Hypothetical protein	267		Yes	Yes
Cag_0835	Hypothetical protein	137		Yes	Yes
Cag_0840	TPR-repeat protein	1371	N-acetylglucosamylcysteinetransferase; hydrolase activity	No	No
Cag_0854	Hypothetical protein	73		No	Yes
Cag_0866	Hypothetical protein	87		No	No
Cag_0877	Hypothetical protein	245	FimC membrane bound pili assembly protein	No	No
Cag_0878	Hypothetical protein	360	M14 peptidase; ABC-transporter related	No	No
Cag_0894	Hypothetical protein DUF262	578		No	No

Cag_0911	Hypothetical protein	561	ATPase	No	No
Cag_0958	Hypothetical protein	111		No	No
Cag_0984	Hypothetical protein	413		Yes	No
Cag_0992	Hypothetical protein	60		No	No
Cag_0993	**Hypothetical protein**	**99**		**No**	**No**
Cag_0994	Hypothetical protein	60	CRISPR array protein; bacteriophage resistance	No	No
Cag_0995	Hypothetical protein	92		No	No
Cag_1003	Hypothetical protein	125		No	No
Cag_1008	**CRISPR-associated protein, Csd1 family**	**642**	**Glutamate-cysteine ligase**	**No**	**No**
Cag_1051	Hypothetical protein	100	Plasmid stabilization system	No	No
Cag_1061	Conserved hypothetical phage AbiD protein	304	Abortive infection phage related (AIPR)	No	No
Cag_1062	Hypothetical protein	488		No	No
Cag_1097	Hypothetical protein	346	Fic-like, filamentation/cell division regulation via folate biosynthesis	No	No
Cag_1098	Hypothetical protein	465		No	No
Cag_1104	Hypothetical protein	79	CopG-like; helix-turn-helix motive; transcription regulation	No	No
Cag_1122	Restriction endonuclease S subunits-like	428	Methylase; against invasion of foreign DNA	No	No
Cag_1131	Hypothetical protein	265	Uracil-DNA glycosylase	No	No
Cag_1147	Hypothetical protein	114		No	Yes
Cag_1178	Hypothetical protein	79		No	No
Cag_1191	Hypothetical protein	177		No	No

Cag ID	Annotation		Function		
Cag_1196	Adenine-specific DNA methyltransferase	1059	Helicase function	No	No
Cag_1197	Hypothetical protein	84		No	No
Cag_1201	Hypothetical protein	120		No	No
Cag_1202	Hypothetical protein	207		No	Yes
Cag_1236	Hypothetical protein	235		Yes	No
Cag_1238	Hypothetical protein	170		Yes	Yes
Cag_1254	Hypothetical protein	221		Yes	No
Cag_1269	Hypothetical protein	128	Aspartyl-protease; posttranslational modification/protein turnover	No	No
Cag_1272	Hypothetical protein DUF 1778	79		No	No
Cag_1277	Hypothetical protein	62		No	No
Cag_1278	Hypothetical protein	198		No	No
Cag_1284	Hypothetical protein	110		No	No
Cag_1287	Hypothetical protein	87		No	No
Cag_1303	Hypothetical protein	168		Yes	No
Cag_1309	**Methylase**	**933**	**Adenine specific DNA-methylase; involved in defence**	**No**	**No**
Cag_1310	HNH nuclease	224		No	No
Cag_1318	**Hypothetical protein**	**77**		**No**	**No**
Cag_1321	Hypothetical protein	182		Yes	No
Cag_1352	Hypothetical protein	167	OmpA-like ; outer membrane porine channel formation	No	No
Cag_1353	Hypothetical protein	101		No	No
Cag_1354	Nucleic acid-binding protein; PIN domain-like	148	PilT-like; polymerization of pilus fibre	No	No

Gene	Annotation	Size	Description		
Cag_1362	Hypothetical protein	311		No	Yes
Cag_1363	Serine/threonine protein kinase	478	Signal transduction mechanisms	No	No
Cag_1364	Hypothetical protein	73	Protein binding	Yes	No
Cag_1374	Hypothetical protein	81		No	No
Cag_1387	ATP-dependent endonuclease, OLD family-like	550		No	No
Cag_1391	Putative type II restriction enzyme	260	Protection against invading DNA	No	No
Cag_1406	Hypothetical protein	359		No	No
Cag_1407	**Peptidase, M50 family**	**720**	**Multidrug resistance efflux pump**	**No**	**Yes**
Cag_1408	Membrane-fusion protein-like	631	HlyD-like; RTX, MDR, ABC/ GAF motives; protein interactions	No	No
Cag_1409	Membrane-fusion protein-like	312		Yes	Yes
Cag_1427	Hypothetical protein	118		Yes	Yes
Cag_1494	Hypothetical protein	76	YcfA-like lipoprotein	No	No
Cag_1505	Hypothetical protein	72		No	No
Cag_1506	Hypothetical protein	112		No	No
Cag_1511	**Hypothetical protein**	**93**	**Receptor tyrosine kinase**	**Yes**	**No**
Cag_1536	Hypothetical protein	122		Yes	Yes
Cag_1559	Transposase for IS1663	351		No	No
Cag_1570	Virulence-associated protein D	92	VapD-like; virulence	No	No
Cag_1571	Hypothetical protein	79	DNA-directed RNA polymerase	No	No
Cag_1599	Hypothetical protein	244	HNH nuclease	No	No
Cag_1600	ATPase	427	RecF_N-like; structural maintenance of chromosomes	No	No

Gene	Annotation	Size	Function		
Cag_1606	Hypothetical protein	457		Yes	No
Cag_1609	Hypothetical protein	521	Methyltransferase	No	No
Cag_1610	Hypothetical protein	315	Type II desoxyribonuclease; restriction enzyme/hydrolase	No	No
Cag_1611	ATPase involved in DNA repair	966	Chromosome segregation or ABC-transport	No	No
Cag_1617	Hypothetical protein	373	Metallophosphoesterase; regulation	Yes	Yes
Cag_1629	Hypothetical protein DUF 262	365		No	No
Cag_1630	Hypothetical protein	229		No	No
Cag_1690	Hypothetical protein	81		No	No
Cag_1697	Hypothetical protein	81		Yes	Yes
Cag_1703	**Hypothetical protein**	**819**	**N6 adenine specific DNA-methylase**	**No**	**No**
Cag_1723	Hypothetical protein	296		Yes	Yes
Cag_1769	Protein of unknown function DUF132	141	Nucleotide binding PIN-domain (PilT-like)	No	No
Cag_1770	hypothetical protein	73		No	No
Cag_1772	hypothetical protein	573	ABC-transporter related	Yes	No
Cag_1776	hypothetical protein	66		No	No
Cag_1795	hypothetical protein	104		No	Yes
Cag_1796	**Ankyrin**	**1047**	**DNA/RNA helicase; horizontal gene transfer; transcription**	**No**	**No**
Cag_1860	**Hypothetical protein**	**415**		**No**	**No**
Cag_1861	Hypothetical protein	80		No	No
	Hypothetical protein				
Cag_1880	Hypothetical protein	407		No	Yes

Cag_1881	Hypothetical protein	135		No	No
Cag_1904	Hypothetical protein DUF 86	69	Succinylbenzoic acid synthetase/L-alanine DL glutamate epimerase	No	No
Cag_1917	Hypothetical protein	89		Yes	Yes
Cag_1976	Hypothetical protein	116		Yes	Yes
Cag_1985	Cold shock protein	310	DNA-binding motif	No	No

Suppl. Table S2. *Chl. chlorochromatii* protein coding genes that display significant differential expression between free living epibionts and the symbiotic stage in the "*C. aggregatum*" consortium.

Name	# Tags		% of library		Cons/Free	p-value
	Cons	Free	Cons	Free		
Cag_0026 - secD preprotein translocase subunit SecD	10	10	0.0708	0.0295	2.40	0.0031
Cag_0029 - DNA gyrase, B subunit	2	22	0.0142	0.0649	0.22	0.0009
Cag_0036 - glycyl-tRNA synthetase	17	18	0.1204	0.0531	2.27	0.0002
Cag_0050 - UDP-N-acetylmuramyl-tripeptide synthetase	14	11	0.0992	0.0324	3.06	0.0003
Cag_0076 - cell wall hydrolase/autolysin	11	9	0.0779	0.0265	2.94	0.0012
Cag_0082 - bioD dithiobiotin synthetase	9	5	0.0638	0.0147	4.32	0.0011
Cag_0088 - DNA gyrase, subunit A	1	16	0.0071	0.0472	0.15	0.0028
Cag_0090 - pyrG CTP synthetase	5	1	0.0354	0.0029	12.01	0.0036
Cag_0183 - clpX ATP-dependent protease ATP-binding subunit ClpX	7	1	0.0496	0.0029	16.81	0.0003
Cag_0185 - gltD glutamate synthase subunit beta	2	33	0.0142	0.0973	0.15	0.0000
Cag_0189 - hypothetical protein	12	8	0.085	0.0236	3.60	0.0004
Cag_0194 - hypothetical protein	8	4	0.0567	0.0118	4.80	0.0016
Cag_0214 - cold-shock DNA-binding domain-containing protein	10	3	0.0708	0.0088	8.00	0.0001
Cag_0226 - 8-amino-7-oxononanoate synthase	1	19	0.0071	0.056	0.13	0.0009
Cag_0229 - magnesium-protoporphyrin IX monomethyl ester anaerobic oxidative cyclase	11	11	0.0779	0.0324	2.40	0.0021
Cag_0247 - hypothetical protein	6	4	0.0425	0.0118	3.60	0.0091
Cag_0263 - alpha amylase domain-containing protein	13	17	0.0921	0.0501	1.84	0.0019
Cag_0282 - aspartate-semialdehyde dehydrogenase, USG-1 related	5	1	0.0354	0.0029	12.01	0.0036
Cag_0290 - hypothetical protein	13	6	0.0921	0.0177	5.20	0.0001
Cag_0341 - TPR repeat-containing protein	7	1	0.0496	0.0029	16.81	0.0003
Cag_0343 - photosystem P840 reaction center, large subunit	10	6	0.0708	0.0177	4.00	0.0008
Cag_0358 - DNA-directed RNA polymerase subunit beta'	9	14	0.0638	0.0413	1.54	0.0090
Cag_0365 - 2-oxoglutarate ferredoxin oxidoreductase subunit beta	6	4	0.0425	0.0118	3.60	0.0091
Cag_0391 - peptidase M41, FtsH	14	12	0.0992	0.0354	2.80	0.0004
Cag_0454 - phosphoglucomutase/phosphomannomutase family protein	2	25	0.0142	0.0737	0.19	0.0003
Cag_0458 - folylpolyglutamate synthetase	7	5	0.0496	0.0147	3.36	0.0060
Cag_0465 - gatA aspartyl/glutamyl-tRNA amidotransferase subunit A	2	20	0.0142	0.059	0.24	0.0018
Cag_0485 - hypothetical protein	10	3	0.0708	0.0088	8.00	0.0001
Cag_0527 - cell division protein, putative	8	2	0.0567	0.0059	9.61	0.0003
Cag_0529 - hypothetical protein	2	19	0.0142	0.056	0.25	0.0025
Cag_0530 - hydrogenase formation HypD protein	17	5	0.1204	0.0147	8.16	0.0000
Cag_0543 - hypothetical protein	12	10	0.085	0.0295	2.88	0.0008

Gene						
Cag_0548 - DNA primase	9	9	0.0638	0.0265	2.40	0.0047
Cag_0560 - ubiquinone/menaquinone biosynthesis methyltransferase	10	11	0.0708	0.0324	2.18	0.0038
Cag_0615 - Outer membrane protein-like	9	5	0.0638	0.0147	4.32	0.0011
Cag_0625 - riboflavin synthase subunit alpha	2	17	0.0142	0.0501	0.28	0.0049
Cag_0626 - recombination factor protein RarA	13	8	0.0921	0.0236	3.90	0.0002
Cag_0646 - hypothetical protein	9	2	0.0638	0.0059	10.81	0.0001
Cag_0669 - sugar transferase	5	1	0.0354	0.0029	12.01	0.0036
Cag_0689 - PucC protein	5	2	0.0354	0.0059	6.00	0.0081
Cag_0690 - hypothetical protein	11	3	0.0779	0.0088	8.81	0.0000
Cag_0703 - ribonucleotide-diphosphate reductase subunit alpha	8	32	0.0567	0.0944	0.60	0.0012
Cag_0727 - NUDIX/MutT family protein	13	3	0.0921	0.0088	10.41	0.0000
Cag_0738 - VCBS	28	43	0.1983	0.1268	1.56	0.0000
Cag_0769 - exodeoxyribonuclease V, beta subunit	7	39	0.0496	0.115	0.43	0.0002
Cag_0770 - exodeoxyribonuclease V, alpha subunit	4	24	0.0283	0.0708	0.40	0.0020
Cag_0777 - argininosuccinate synthase	9	12	0.0638	0.0354	1.80	0.0077
Cag_0798 - aspartate aminotransferase	5	1	0.0354	0.0029	12.01	0.0036
Cag_0799 - oxidoreductase, FAD-binding	18	11	0.1275	0.0324	3.93	0.0000
Cag_0801 - rplS 50S ribosomal protein L19	3	19	0.0213	0.056	0.38	0.0050
Cag_0817 - phosphomethylpyrimidine kinase	5	1	0.0354	0.0029	12.01	0.0036
Cag_0829 - transcriptional modulator of MazE/toxin, MazF	8	8	0.0567	0.0236	2.40	0.0071
Cag_0843 - hypothetical protein	8	5	0.0567	0.0147	3.84	0.0026
Cag_0847 - helicase RecD/TraA	6	4	0.0425	0.0118	3.60	0.0091
Cag_0861 - hypothetical protein	1	13	0.0071	0.0383	0.18	0.0085
Cag_0882 - sulfide-quinone reductase, putative	1	17	0.0071	0.0501	0.14	0.0019
Cag_0893 - DNA polymerase III, alpha subunit	8	9	0.0567	0.0265	2.13	0.0087
Cag_0971 - NolG efflux transporter	2	16	0.0142	0.0472	0.30	0.0068
Cag_0998 - hypothetical protein	15	21	0.1063	0.0619	1.72	0.0011
Cag_1008 - CRISPR-associated Csd1 family protein	7	3	0.0496	0.0088	5.60	0.0022
Cag_1048 - hypothetical protein	9	14	0.0638	0.0413	1.54	0.0090
Cag_1065 - moaC bifunctional molybdenum cofactor biosynthesis protein C/ molybdopterin-binding protein	5	2	0.0354	0.0059	6.00	0.0081
Cag_1092 - hypothetical protein	1	19	0.0071	0.056	0.13	0.0009
Cag_1106 - glycosyl transferase	5	2	0.0354	0.0059	6.00	0.0081
Cag_1124 - 5,10-methylenetetrahydrofolate reductase	9	5	0.0638	0.0147	4.32	0.0011
Cag_1124 - Type I site-specific deoxyribonuclease HsdR	6	3	0.0425	0.0088	4.80	0.0056
Cag_1125 - propionate--CoA ligase	11	14	0.0779	0.0413	1.89	0.0036
Cag_1138 - putative metal-dependent hydrolase	2	17	0.0142	0.0501	0.28	0.0049

Gene						
Cag_1205 - Na+/H+ anti-porter	6	3	0.0425	0.0088	4.80	0.0056
Cag_1212 - cell division transporter substrate-binding protein FtsY	3	19	0.0213	0.056	0.38	0.0050
Cag_1223 - methylmalonyl-CoA mutase	9	6	0.0638	0.0177	3.60	0.0018
Cag_1239 - VCBS	8	7	0.0567	0.0206	2.74	0.0055
Cag_1247 - nitrogenase molybdenum-iron protein alpha chain	12	9	0.085	0.0265	3.20	0.0006
Cag_1252 - ferredoxin, 2Fe-2S	10	8	0.0708	0.0236	3.00	0.0018
Cag_1257 - O-acetylhomoserine/O-acetylserine sulfhydrylase	7	6	0.0496	0.0177	2.80	0.0084
Cag_1286 - bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase	8	7	0.0567	0.0206	2.74	0.0055
Cag_1301 - eximuclease ABC subunit A	14	10	0.0992	0.0295	3.36	0.0002
Cag_1305 - groES co-chaperonin GroES	39	55	0.2763	0.1622	1.70	0.0000
Cag_1306 - groEL chaperonin GroEL	52	77	0.3684	0.2271	1.62	0.0000
Cag_1309 - methylase	7	6	0.0496	0.0177	2.80	0.0084
Cag_1318 - hypothetical protein	8	3	0.0567	0.0088	6.40	0.0008
Cag_1319 - transcriptional modulator of MazE/toxin, MazF	8	6	0.0567	0.0177	3.20	0.0040
Cag_1377 - D-3-phosphoglycerate dehydrogenase	8	7	0.0567	0.0206	2.74	0.0055
Cag_1407 - M50 family peptidase	7	4	0.0496	0.0118	4.20	0.0039
Cag_1415 - DNA helicase II	7	1	0.0496	0.0029	16.81	0.0003
Cag_1420 - glyceraldehyde-3-phosphate dehydrogenase, type I	3	38	0.0213	0.1121	0.19	0.0000
Cag_1430 - UDP-N-acetylenolpyruvoylglucosamine reductase	1	24	0.0071	0.0708	0.10	0.0001
Cag_1435 - phosphoribosylaminoimidazole-succinocarboxamide synthase	7	5	0.0496	0.0147	3.36	0.0060
Cag_1439 - coaD phosphopantetheine adenylyltransferase	1	15	0.0071	0.0442	0.16	0.0041
Cag_1511 - hypothetical protein	5	2	0.0354	0.0059	6.00	0.0081
Cag_1544 - phenylalanyl-tRNA synthetase subunit beta	1	16	0.0071	0.0472	0.15	0.0028
Cag_1548 - S-adenosyl-L-homocysteine hydrolase	5	22	0.0354	0.0649	0.55	0.0052
Cag_1560 - VCBS	10	16	0.0708	0.0472	1.50	0.0066
Cag_1583 - heterodisulfide reductase, subunit A/hydrogenase, delta subunit, putative	12	6	0.085	0.0177	4.80	0.0001
Cag_1585 - adenylylsulfate reductase subunit alpha	11	9	0.0779	0.0265	2.94	0.0012
Cag_1586 - adenylylsulfate reductase subunit beta	7	4	0.0496	0.0118	4.20	0.0039
Cag_1587 - sat sulfate adenylyltransferase	1	17	0.0071	0.0501	0.14	0.0019
Cag_1597 - hypothetical protein	3	20	0.0213	0.059	0.36	0.0037
Cag_1615 - hypothetical protein	2	20	0.0142	0.059	0.24	0.0018
Cag_1620 - magnesium chelatase ATPase subunit D	9	9	0.0638	0.0265	2.40	0.0047
Cag_1621 - magnesium chelatase ATPase subunit I	8	8	0.0567	0.0236	2.40	0.0071
Cag_1656 - gcvT glycine cleavage system aminomethyltransferase T	2	17	0.0142	0.0501	0.28	0.0049
Cag_1688 - leuS leucyl-tRNA synthetase	1	20	0.0071	0.059	0.12	0.0006
Cag_1699 - hypothetical protein	8	8	0.0567	0.0236	2.40	0.0071
Cag_1701 - isochorismate synthase	2	18	0.0142	0.0531	0.27	0.0035

Gene						
Cag_1711 - phenylalanyl-tRNA synthetase, alpha subunit	9	3	0.0638	0.0088	7.20	0.0003
Cag_1725 - Kup system potassium uptake protein	15	1	0.1063	0.0029	36.02	0.0000
Cag_1749 - DNA repair protein RecN	7	1	0.0496	0.0029	16.81	0.0003
Cag_1766 - peptide ABC transporter, permease protein	11	8	0.0779	0.0236	3.30	0.0008
Cag_1768 - TPR repeat-containing protein	1	14	0.0071	0.0413	0.17	0.0059
Cag_1775 - metal dependent phosphohydrolase	1	21	0.0071	0.0619	0.11	0.0004
Cag_1789 - glyA serine hydroxymethyltransferase	19	27	0.1346	0.0796	1.69	0.0003
Cag_1796 - ankyrin	15	14	0.1063	0.0413	2.57	0.0003
Cag_1798 - XRE family transcriptional regulator	5	1	0.0354	0.0029	12.01	0.0036
Cag_1834 - rpsE 30S ribosomal protein S5	10	4	0.0708	0.0118	6.00	0.0002
Cag_1840 - rplX 50S ribosomal protein L24	3	17	0.0213	0.0501	0.42	0.0089
Cag_1846 - rplV 50S ribosomal protein L22	11	2	0.0779	0.0059	13.21	0.0000
Cag_1847 - rpsS 30S ribosomal protein S19	17	2	0.1204	0.0059	20.41	0.0000
Cag_1848 - rplB 50S ribosomal protein L2	27	20	0.1913	0.059	3.24	0.0000
Cag_1852 - rpsJ 30S ribosomal protein S10	5	50	0.0354	0.1475	0.24	0.0000
Cag_1853 - elongation factor Tu	43	61	0.3046	0.1799	1.69	0.0000
Cag_1854 - elongation factor G	13	2	0.0921	0.0059	15.61	0.0000
Cag_1882 - transcription-repair coupling factor	6	22	0.0425	0.0649	0.65	0.0065
Cag_1883 - oligopeptide/dipeptide ABC transporter, ATP-binding protein-like	19	4	0.1346	0.0118	11.41	0.0000
Cag_1893 - dnaK molecular chaperone DnaK	16	20	0.1133	0.059	1.92	0.0006
Cag_1906 - dihydroxy-acid dehydratase	12	5	0.085	0.0147	5.76	0.0001
Cag_1925 - sulfur oxidation protein SoxA	8	3	0.0567	0.0088	6.40	0.0008
Cag_1927 - twin-arginine translocation pathway signal	5	23	0.0354	0.0678	0.52	0.0041
Cag_1934 - sdhA succinate dehydrogenase flavoprotein subunit	4	19	0.0283	0.056	0.51	0.0078
Cag_1947 - DsrK protein	8	4	0.0567	0.0118	4.80	0.0016
Cag_1973 - ATPase	5	24	0.0354	0.0708	0.50	0.0032
Cag_1974 - 2-dehydro-3-deoxyphosphooctonate aldolase	1	16	0.0071	0.0472	0.15	0.0028
Cag_1978 - rubrerythrin	8	3	0.0567	0.0088	6.40	0.0008
Cag_1983 - alpha amylase domain-containing protein	2	17	0.0142	0.0501	0.28	0.0049
Cag_1997 - TPR repeat-containing protein	7	6	0.0496	0.0177	2.80	0.0084
Cag_2014 - F0F1 ATP synthase subunit beta	9	5	0.0638	0.0147	4.32	0.0011
Cag_2018 - hypothetical protein	2	15	0.0142	0.0442	0.32	0.0094
Cag_2028 - 23S rRNA methyltransferase/RumA	6	2	0.0425	0.0059	7.20	0.0029
Detected in Consortium Only						
Cag_0003 - RecF protein	5	0	0.0354	0	N.A.	0.0009
Cag_0004 - hypothetical protein	5	0	0.0354	0	N.A.	0.0009

Gene						
Cag_0094 - rpsF 30S ribosomal protein S6	20	0	0.1417	0	N.A.	0.0000
Cag_0124 - hypothetical protein	5	0	0.0354	0	N.A.	0.0009
Cag_0137 - 1-acyl-sn-glycerol-3-phosphate acyltransferase	4	0	0.0283	0	N.A.	0.0034
Cag_0140 - F0F1 ATP synthase subunit alpha	4	0	0.0283	0	N.A.	0.0034
Cag_0142 - thrA bifunctional aspartokinase I/homeserine dehydrogenase I	12	0	0.085	0	N.A.	0.0000
Cag_0146 - monofunctional biosynthetic peptidoglycan transglycosylase	4	0	0.0283	0	N.A.	0.0034
Cag_0149 - NusG antitermination factor	5	0	0.0354	0	N.A.	0.0009
Cag_0162 - bifunctional UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase/(3R)-hydroxymyristoyl-[acyl-carrier-protein] dehydratase	6	0	0.0425	0	N.A.	0.0002
Cag_0180 - prephenate dehydratase	4	0	0.0283	0	N.A.	0.0034
Cag_0219 - chlorosome envelope protein A	12	0	0.085	0	N.A.	0.0000
Cag_0253 - hypothetical protein	6	0	0.0425	0	N.A.	0.0002
Cag_0257 - ATPase	4	0	0.0283	0	N.A.	0.0034
Cag_0258 - purH bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	4	0	0.0283	0	N.A.	0.0034
Cag_0318 - gatC aspartyl/glutamyl-tRNA amidotransferase subunit C	6	0	0.0425	0	N.A.	0.0002
Cag_0334 - ATPase	6	0	0.0425	0	N.A.	0.0002
Cag_0335 - FusA/NodT family protein	4	0	0.0283	0	N.A.	0.0034
Cag_0356 - rplL 50S ribosomal protein L7/L12	6	0	0.0425	0	N.A.	0.0002
Cag_0369 - alaS alanyl-tRNA synthetase	6	0	0.0425	0	N.A.	0.0002
Cag_0430 - hypothetical protein	7	0	0.0496	0	N.A.	0.0001
Cag_0434 - hypothetical protein	6	0	0.0425	0	N.A.	0.0002
Cag_0451 - tryptophan synthase subunit beta	6	0	0.0425	0	N.A.	0.0002
Cag_0464 - di-trans-poly-cis-decaprenylcistransferase	4	0	0.0283	0	N.A.	0.0034
Cag_0478 - VpsC protein	5	0	0.0354	0	N.A.	0.0009
Cag_0498 - phosphate ABC transporter permease	5	0	0.0354	0	N.A.	0.0009
Cag_0503 - acetylornithine aminotransferase	4	0	0.0283	0	N.A.	0.0034
Cag_0537 - glutamate synthase	6	0	0.0425	0	N.A.	0.0002
Cag_0553 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_0601 - hypothetical protein	6	0	0.0425	0	N.A.	0.0002
Cag_0602 - fructose-1,6-bisphosphatase	5	0	0.0354	0	N.A.	0.0009
Cag_0604 - hypothetical protein	6	0	0.0425	0	N.A.	0.0002
Cag_0620 - twin-arginine translocation pathway signal	11	0	0.0779	0	N.A.	0.0000
Cag_0622 - deoxycytidylate deaminase, putative	4	0	0.0283	0	N.A.	0.0034
Cag_0644 - proton-translocating NADH-quinone oxidoreductase, chain N	4	0	0.0283	0	N.A.	0.0034
Cag_0647 - glycosyl transferase family protein	4	0	0.0283	0	N.A.	0.0034
Cag_0657 - putative transcriptional regulator	4	0	0.0283	0	N.A.	0.0034

Gene					
Cag_0722 - HDIG	4	0	0.0283	0	N.A.	0.0034
Cag_0726 - ABC transporter, periplasmic substrate-binding protein	7	0	0.0496	0	N.A.	0.0001
Cag_0787 - alkaline phosphatase	4	0	0.0283	0	N.A.	0.0034
Cag_0796 - citrate lyase, subunit 1	6	0	0.0425	0	N.A.	0.0002
Cag_0797 - citrate lyase, subunit 2	9	0	0.0638	0	N.A.	0.0000
Cag_0838 - hypothetical protein	11	0	0.0779	0	N.A.	0.0000
Cag_0849 - ATPase	11	0	0.0779	0	N.A.	0.0000
Cag_0853 - branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein, putative	8	0	0.0567	0	N.A.	0.0000
Cag_0916 - C-type cytochrome, putative	5	0	0.0354	0	N.A.	0.0009
Cag_0932 - aspS aspartyl-tRNA synthetase	5	0	0.0354	0	N.A.	0.0009
Cag_0950 - K+-dependent Na+/Ca+ exchanger related-protein	6	0	0.0425	0	N.A.	0.0002
Cag_0963 - polyprenyl synthetase	4	0	0.0283	0	N.A.	0.0034
Cag_0967 - superoxide dismutase	7	0	0.0496	0	N.A.	0.0001
Cag_0978 - ParaA family ATPase	4	0	0.0283	0	N.A.	0.0034
Cag_0993 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_1034 - hypothetical protein	5	0	0.0354	0	N.A.	0.0009
Cag_1036 - hypothetical protein	5	0	0.0354	0	N.A.	0.0009
Cag_1054 - activation/secretion signal peptide protein	5	0	0.0354	0	N.A.	0.0009
Cag_1079 - ATPase	7	0	0.0496	0	N.A.	0.0001
Cag_1086 - histidinol-phosphate aminotransferase	7	0	0.0496	0	N.A.	0.0001
Cag_1090 - hppA membrane-bound proton-translocating pyrophosphatase	5	0	0.0354	0	N.A.	0.0009
Cag_1094 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_1103 - hypothetical protein	6	0	0.0425	0	N.A.	0.0002
Cag_1111 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_1119 - Fmu, rRNA SAM-dependent methyltransferase	5	0	0.0354	0	N.A.	0.0009
Cag_1137 - HhH-GPD	9	0	0.0638	0	N.A.	0.0000
Cag_1159 - L-aspartate oxidase	4	0	0.0283	0	N.A.	0.0034
Cag_1207 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_1244 - nifH nitrogenase reductase	8	0	0.0567	0	N.A.	0.0000
Cag_1249 - nitrogenase MoFe cofactor biosynthesis protein NifE	13	0	0.0921	0	N.A.	0.0000
Cag_1251 - nitrogenase cofactor biosynthesis protein NifB	5	0	0.0354	0	N.A.	0.0009
Cag_1265 - putative membrane-located cell surface saccharide saccharide acetylase protein	6	0	0.0425	0	N.A.	0.0002
Cag_1268 - thiH thiamine biosynthesis protein ThiH	5	0	0.0354	0	N.A.	0.0009
Cag_1281 - membrane-bound metallopeptidase-like	5	0	0.0354	0	N.A.	0.0009
Cag_1316 - glycosyl transferase	11	0	0.0779	0	N.A.	0.0000
Cag_1322 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034

Gene						
Cag_1339 - putative sugar transport protein	10	0	0.0708	0	N.A.	0.0000
Cag_1419 - CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	5	0	0.0354	0	N.A.	0.0009
Cag_1443 - hypothetical protein	7	0	0.0496	0	N.A.	0.0001
Cag_1466 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_1467 - hisS histidyl-tRNA synthetase	4	0	0.0283	0	N.A.	0.0034
Cag_1470 - phosphatidate cytidylyltransferase	7	0	0.0496	0	N.A.	0.0001
Cag_1483 - acyltransferase	4	0	0.0283	0	N.A.	0.0034
Cag_1495 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_1504 - TPR repeat-containing protein	4	0	0.0283	0	N.A.	0.0034
Cag_1542 - hypothetical protein	5	0	0.0354	0	N.A.	0.0009
Cag_1558 - hypothetical protein	20	0	0.1417	0	N.A.	0.0000
Cag_1561 - zinc protease, putative	8	0	0.0567	0	N.A.	0.0000
Cag_1588 - glutamine synthetase	5	0	0.0354	0	N.A.	0.0009
Cag_1652 - ATP-binding Mrp/Nbp35 family protein	6	0	0.0425	0	N.A.	0.0002
Cag_1654 - ATPase	8	0	0.0567	0	N.A.	0.0000
Cag_1658 - glycine dehydrogenase subunit 2	5	0	0.0354	0	N.A.	0.0009
Cag_1663 - malonyl CoA-acyl carrier protein transacylase	5	0	0.0354	0	N.A.	0.0009
Cag_1742 - bacteriochlorophyll/chlorophyll a synthase	9	0	0.0638	0	N.A.	0.0000
Cag_1746 - MesJ protein	7	0	0.0496	0	N.A.	0.0001
Cag_1818 - cysS cysteinyl-tRNA synthetase	14	0	0.0992	0	N.A.	0.0000
Cag_1824 - DNA-directed RNA polymerase subunit alpha	8	0	0.0567	0	N.A.	0.0000
Cag_1830 - peptidase M24A	8	0	0.0567	0	N.A.	0.0000
Cag_1849 - rplW 50S ribosomal protein L23	5	0	0.0354	0	N.A.	0.0009
Cag_1855 - 30S ribosomal protein S7	8	0	0.0567	0	N.A.	0.0000
Cag_1860 - hypothetical protein	7	0	0.0496	0	N.A.	0.0001
Cag_1867 - glycosyl transferase	5	0	0.0354	0	N.A.	0.0009
Cag_1892 - HSP20 family protein	6	0	0.0425	0	N.A.	0.0002
Cag_1902 - ketol-acid reductoisomerase	6	0	0.0425	0	N.A.	0.0002
Cag_1905 - acetolactate synthase large subunit biosynthetic type	7	0	0.0496	0	N.A.	0.0001
Cag_1930 - ferrous iron transport protein A	5	0	0.0354	0	N.A.	0.0009
Cag_1936 - hypothetical protein	9	0	0.0638	0	N.A.	0.0000
Cag_1938 - Sel1 repeat-containing protein	5	0	0.0354	0	N.A.	0.0009
Cag_1962 - hypothetical protein	4	0	0.0283	0	N.A.	0.0034
Cag_1996 - putative segregation and condensation protein A	7	0	0.0496	0	N.A.	0.0001
Cag_2033 - rpmH 50S ribosomal protein L34	15	0	0.1063	0	N.A.	0.0000

Detected in Free Living Only

110

Gene						
Cag_0053 - UDP-N-acetylmuramoylalanine-D-glutamate ligase	0	21	0	0.0619	N.A.	0.0001
Cag_0054 - cell cycle protein FtsW	0	21	0	0.0619	N.A.	0.0001
Cag_0080 - hypothetical protein	0	17	0	0.0501	N.A.	0.0004
Cag_0093 - heptosyltransferase	0	13	0	0.0383	N.A.	0.0023
Cag_0106 - aminotransferase, class V	0	12	0	0.0354	N.A.	0.0035
Cag_0112 - von Willebrand factor, type A	0	18	0	0.0531	N.A.	0.0003
Cag_0141 - F0F1 ATP synthase subunit gamma	0	11	0	0.0324	N.A.	0.0054
Cag_0158 - CrtK protein	0	14	0	0.0413	N.A.	0.0015
Cag_0178 - membrane proteins related to metalloendopeptidase-like	0	13	0	0.0383	N.A.	0.0023
Cag_0207 - IscU protein	0	15	0	0.0442	N.A.	0.0009
Cag_0251 - ATPase	0	13	0	0.0383	N.A.	0.0023
Cag_0286 - 1-(5-phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR) carboxylase	0	13	0	0.0383	N.A.	0.0023
Cag_0303 - M16 family peptidase	0	10	0	0.0295	N.A.	0.0083
Cag_0321 - xseA exodeoxyribonuclease VII large subunit	0	12	0	0.0354	N.A.	0.0035
Cag_0329 - GTPase EngC	0	12	0	0.0354	N.A.	0.0035
Cag_0381 - XRE family transcriptional regulator	0	11	0	0.0324	N.A.	0.0054
Cag_0393 - hydroxyneurosporene synthase CrtC	0	13	0	0.0383	N.A.	0.0023
Cag_0413 - molybdenum ABC transporter periplasmic-binding protein	0	12	0	0.0354	N.A.	0.0035
Cag_0419 - chromosome segregation protein SMC	0	27	0	0.0796	N.A.	0.0000
Cag_0463 - surface antigen family protein	0	34	0	0.1003	N.A.	0.0000
Cag_0471 - hypothetical protein	0	15	0	0.0442	N.A.	0.0009
Cag_0480 - putative ABC transport system permease protein	0	11	0	0.0324	N.A.	0.0054
Cag_0493 - anthranilate synthase component I	0	15	0	0.0442	N.A.	0.0009
Cag_0494 - peptidase S1C, Do	0	14	0	0.0413	N.A.	0.0015
Cag_0497 - tpiA triosephosphate isomerase	0	20	0	0.059	N.A.	0.0001
Cag_0509 - hypothetical protein	0	12	0	0.0354	N.A.	0.0035
Cag_0511 - metal dependent phosphohydrolase	0	11	0	0.0324	N.A.	0.0054
Cag_0523 - ileS isoleucyl-tRNA synthetase	0	10	0	0.0295	N.A.	0.0083
Cag_0526 - hypothetical protein	0	16	0	0.0472	N.A.	0.0006
Cag_0556 - hydrogenase expression/formation protein HypE	0	16	0	0.0472	N.A.	0.0006
Cag_0559 - hisI phosphoribosyl-AMP cyclohydrolase	0	17	0	0.0501	N.A.	0.0004
Cag_0577 - hypothetical protein	0	15	0	0.0442	N.A.	0.0009
Cag_0684 - hypothetical protein	0	11	0	0.0324	N.A.	0.0054
Cag_0685 - hypothetical protein	0	11	0	0.0324	N.A.	0.0054
Cag_0732 - TPR repeat-containing protein	0	17	0	0.0501	N.A.	0.0004
Cag_0752 - ribonuclease R	0	14	0	0.0413	N.A.	0.0015
Cag_0868 - RND efflux system, outer membrane lipoprotein, NodT	0	15	0	0.0442	N.A.	0.0009

Gene						
Cag_0896 - methyltransferase, putative	0	16	0	0.0472	N.A.	0.0006
Cag_0931 - Elongator protein 3/MiaB/NifB	0	10	0	0.0295	N.A.	0.0083
Cag_0959 - heavy metal translocating P-type ATPase	0	13	0	0.0383	N.A.	0.0023
Cag_1014 - TonB-dependent receptor-related protein	0	14	0	0.0413	N.A.	0.0015
Cag_1056 - hemolysin activation/secretion protein-like	0	25	0	0.0737	N.A.	0.0000
Cag_1064 - molybdopterin binding domain-containing protein	0	11	0	0.0324	N.A.	0.0054
Cag_1110 - rpmG 50S ribosomal protein L33	0	11	0	0.0324	N.A.	0.0054
Cag_1114 - DegT/DnrJ/EryC1/StrS family protein	0	11	0	0.0324	N.A.	0.0054
Cag_1116 - DEAD/DEAH box helicase-like	0	11	0	0.0324	N.A.	0.0054
Cag_1135 - long-chain fatty-acid-CoA ligase	0	10	0	0.0295	N.A.	0.0083
Cag_1145 - hypothetical protein	0	11	0	0.0324	N.A.	0.0054
Cag_1149 - hypothetical protein	0	13	0	0.0383	N.A.	0.0023
Cag_1155 - ribose-phosphate pyrophosphokinase	0	13	0	0.0383	N.A.	0.0023
Cag_1237 - arginine/ornithine transport system ATPase	0	11	0	0.0324	N.A.	0.0054
Cag_1276 - dihydrolipoamide dehydrogenase	0	10	0	0.0295	N.A.	0.0083
Cag_1331 - Fis family transcriptional regulator	0	17	0	0.0501	N.A.	0.0004
Cag_1345 - quinate/shikimate 5-dehydrogenase	0	10	0	0.0295	N.A.	0.0083
Cag_1349 - peptidyl-prolyl cis-trans isomerase, cyclophilin-type	0	19	0	0.056	N.A.	0.0002
Cag_1361 - hypothetical protein	0	10	0	0.0295	N.A.	0.0083
Cag_1368 - hypothetical protein	0	14	0	0.0413	N.A.	0.0015
Cag_1384 - sodium:solute symporter family protein	0	18	0	0.0531	N.A.	0.0003
Cag_1403 - metG methionyl-tRNA synthetase	0	12	0	0.0354	N.A.	0.0035
Cag_1411 - enoyl-(acyl-carrier protein) reductase (NADH)	0	10	0	0.0295	N.A.	0.0083
Cag_1440 - aspartate aminotransferase, putative	0	12	0	0.0354	N.A.	0.0035
Cag_1446 - isoprenyl synthetase	0	10	0	0.0295	N.A.	0.0083
Cag_1454 - glutathione S-transferase, fosfomycin resistance protein, putative	0	12	0	0.0354	N.A.	0.0035
Cag_1526 - restriction endonuclease S subunits-like	0	11	0	0.0324	N.A.	0.0054
Cag_1537 - Elongator protein 3/MiaB/NifB	0	19	0	0.056	N.A.	0.0002
Cag_1545 - NUDIX/MutT family protein	0	11	0	0.0324	N.A.	0.0054
Cag_1546 - glucose-1-phosphate thymidylyltransferase	0	12	0	0.0354	N.A.	0.0035
Cag_1553 - sulphate anion transporter	0	17	0	0.0501	N.A.	0.0004
Cag_1616 - bifunctional ADP-heptose synthase	0	17	0	0.0501	N.A.	0.0004
Cag_1641 - aldolase	0	13	0	0.0383	N.A.	0.0023
Cag_1664 - 3-oxoacyl-(acyl carrier protein) synthase III	0	12	0	0.0354	N.A.	0.0035
Cag_1703 - hypothetical protein	0	15	0	0.0442	N.A.	0.0009
Cag_1710 - NADH-dependent butanol dehydrogenase	0	17	0	0.0501	N.A.	0.0004
Cag_1750 - hypothetical protein	0	12	0	0.0354	N.A.	0.0035

Cag_1764 - HPr kinase/phosphorylase	0	17	0	0.0501	N.A.	0.0004
Cag_1784 - Alpha-glucan phosphorylase	0	12	0	0.0354	N.A.	0.0035
Cag_1785 - hypothetical protein	0	14	0	0.0413	N.A.	0.0015
Cag_1863 - glycosyl transferase, group 1 family protein	0	15	0	0.0442	N.A.	0.0009
Cag_1910 - TolB protein, putative	0	14	0	0.0413	N.A.	0.0015
Cag_1923 - twin-arginine translocation pathway signal	0	11	0	0.0324	N.A.	0.0054
Cag_1937 - hypothetical protein	0	17	0	0.0501	N.A.	0.0004
Cag_1977 - mazG nucleoside triphosphate pyrophosphohydrolase	0	13	0	0.0383	N.A.	0.0023
Cag_2005 - sulfide dehydrogenase, cytochrome subunit	0	17	0	0.0501	N.A.	0.0004
Cag_2024 - phosphoribosylaminoimidazole synthetase	0	14	0	0.0413	N.A.	0.0015
Cag_2026 - gpsA NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	0	14	0	0.0413	N.A.	0.0015

Suppl. Table S3: Gene specific oligonucleotide primers used for PCR/RT-qPCR *

Primer name /target ORF	Oligonucleotide sequence (5´-3´)
rpoDf/RNA polymerase sigma factor rpoD (Cag_0387)	CGACAACCGCTTGCTTGAC
rpoDr	CCAATGCCAAAGTAAGAGCGA
CC1f/Chlorosome protein (Cag_1285)	GCAAGCTGATGAAGCTGAC
CC1r	TCAGCAGCAGGTGCTTC
CM2f/Binding protein of ABC-transporter (Cag_0853)	CTCAAGCGGCTCCCATTA
CM2r	TCGCCACCATCATCTTCA
C2f/Nitrogen regulatory protein P-II (Cag_1245)	TTGTAGGGCGTGGCAAAC
C2r	CCATCAGCACCCGACTTT
C10f/Glutamate synthase (Cag_0537)	TTTGGGGGTGGCAATACT
C10r	ATACGGGCAGGCATTTCA
C25f/Glyceraldehyde-3-phosphate dehydrogenase (Cag_1420)	CTTAACCGTGAGGGCTTCCAA
C25r	CGTACCAGCCAACAACCTTT
C27f/PTS IIA-domain protein (Cag_1468)	CTGCCGTTAGCGACAATG
C27r	CGCCATCAATGGAATCAA
C30f/UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (Cag_1154)	GGCGAAGAGCTTATTCATGC
C30r	CGCTAATGAACGCTCAACAA

* an annealing temperature of 59.5 °C was used for all primers

Chapter 5

Ultrastructure, tactic behaviour and potential for sulfate reduction of a novel multicellular magnetotactic prokaryote from North Sea sediments

Summary

Multicellular magnetotactic prokaryotes (MMPs) represent highly organised, spherical and motile aggregates of 10-40 bacterial cells containing magnetosomes. Although consisting of different cells, each with its own magnetosomes and flagellation, MMPs orient themselves within a magnetic field and exhibit magnetotaxis. So far, MMPs have only been found in several North and South American coastal lagoons and salt marshes. In the present study, a novel type of MMP was discovered in coastal tidal sand flats of the North Sea. High resolution scanning electron microscopy revealed the presence of bullet-shaped magnetosomes which were aligned in several parallel chains. Within each aggregate, the magnetosome chains of individual cells were oriented in the same direction. Energy dispersive X-ray analysis (EDX) showed that the magnetosomes are composed of iron sulfide. This particular morphology and arrangement of magnetosomes has previously not been reported for other MMPs. 16S rRNA gene sequence analysis revealed a single phylotype which represented a novel phylogenetic lineage with $\geq 4\%$ sequence divergence to all previously described MMP sequences and was related to the dissimilatory sulfate reducing *Desulfosarcina variabilis* within the family *Desulfobacteraceae* of the subphylum *Deltaproteobacteria*. Fluorescence *in situ* hybridisation with a specific oligonucleotide probe revealed that all MMPs in the tidal flat sediments studied belonged to the novel phylotype. Within each MMP, all bacterial cells showed a hybridisation signal, indicating that the aggregates are composed of cells of the same phylotype. Genes for dissimilatory sulfite reductase (*dsr*AB) and dissimilatory adenosine-5´-phosphate reductase (*apr*A) could be detected in purified MMP samples, suggesting that MMPs are capable of sulfate reduction. Chemotaxis assays with 41 different test compounds yielded strong responses towards acetate and propionate, whereas other organic acids, alcohols, sugars, sugar alcohols or sulfide did not elicit any response. By means of its coordinated magnetotaxis and chemotaxis, the novel type of MMP is well adapted to the steep chemical gradients which are characteristic for intertidal marine sediments.

Introduction

Mulberry-shaped aggregates of Gram-negative cells which move magnetotactically as an entire unit were first discovered in a coastal brackish lagoon near Rio de Janeiro City (Brasil) (Esquivel *et al.*, 1983; Farina *et al.*, 1983) and have been named multicellular magnetotactic prokaryotes (MMPs) (Rodgers *et al.*, 1990b). So far, the most detailed phenotypic characterisation is available for the MMP '*Candidatus* Magnetoglobus multicellularis' from Araruama lagoon in Brazil (Abreu *et al.*, 2007).

MMPs are spherical with a diameter varying between 3 and 12 μm. They consist of 10-40 individual cells, which are arranged around a central, acellular compartment (Keim *et al.*, 2004a). Magnetosome crystals of MMPs are typically composed of iron sulfides with cubical to rectangular shape and a size varying between 75 - 120 nm. The crystals are predominantly composed of the magnetic iron-sulfur mineral greigite (Fe_3S_4) (Mann *et al.*, 1990a; Pósfai *et al.*, 1998b) and form clusters inside each cell (Farina *et al.*, 1983; Mann *et al.*, 1990a). Within each MMP, the outer membranes of adjacent cells form contact zones (Keim *et al.*, 2004a). The cells carry tufts of flagella which are inserted only at the aggregate surface and thus protrude in the environment (Rodgers *et al.*, 1990a; Silva *et al.*, 2007).

Individual MMPs move in an oriented fashion (Greenberg *et al.*, 2005) which indicates that the flagellar motion is highly coordinated between cells of the same aggregate. The life cycle of MMPs comprises a synchronous division of all cells, which leads to two identical daughter aggregates without a single celled stage (Keim *et al.*, 2004b, 2007). A close cell-cell interaction and interdependence is also suggested by the observations that individual cells of disaggregated MMPs rapidly loose their viability (Abreu *et al.*, 2006) and that free-living single cells were never observed (Rodgers *et al.*, 1990a; Lins and Farina, 1999). Intercellular communication has been suggested to proceed via the internal acellular compartment and the contact sites between the cells (Keim *et al.*, 2004a). MMPs therefore may serve as future model systems for the study of bacterial multicellularity.

Meanwhile, multicellular magnetotactic prokaryotes have been detected in various aquatic environments (Bazylinski *et al.*, 1990; Mann *et al.*, 1990a; Rodgers *et al.*, 1990a; Pósfai *et al.*, 1998a; Lins and Farina, 1999; Simmons *et al.*, 2004; Keim *et al.*, 2007; Simmons and Edwards, 2007). MMPs have been found in the anoxic zone of brackish to hypersaline aquatic environments including stratified sediments of salt marshes (DeLong *et al.*, 1993; Pósfai *et al.*, 1998a; Simmons and Edwards, 2007) and coastal lagoons (Farina *et al.*, 1983, 1990), as well as the water column of meromictic salt water lagoons (Pósfai *et al.*, 1998a; Simmons *et al.*, 2004). So far, MMPs have only been reported from locations in North and South America. Therefore, their biogeographical distribution and the full extent of their diversity is presently unknown.

Based on available 16S rRNA gene sequence data, MMPs are related to dissimilatory sulfate reducers within the *Deltaproteobacteria* (DeLong *et al.*, 1993; Abreu *et al.*, 2007; Simmons and Edwards, 2007). Previous phylogenetic analyses suggest that MMPs constitute a single cluster within the *Desulfobacteraceae* (Abreu *et al.*, 2007; Simmons and Edwards, 2007). However, the phylogenetic breadth and the origin of multicellularity of MMPs cannot be assessed further due to the limited knowledge of their diversity.

In the present communication, we report on the ultrastructure, phylogenetic affiliation, functional genes for sulfate reduction and tactic behaviour of MMPs from North Sea intertidal flat sediments. This represents the first report of the occurrence of MMPs outside of North and South America and the first analysis of the chemotactic behaviour of MMPs.

Results and Discussion

Enrichment and ultrastructure of a novel type of MMP

Out of 20 different sites sampled along a 50 km transect in the German Wadden Sea, MMPs were only detected at the five locations with sandy sediments but were never observed in the dominant silty or clayey areas. Enrichments of MMPs were established in sediment cores recovered from the sandy sites. These enrichments could be maintained without a decrease in MMP numbers for up to six months. Subsequent application of a magnetic field yielded a visible greyish pellet at the inner wall of the Plexiglas corer (Fig. 1A, insert) which typically contained about 600 aggregates. Phase contrast microscopic inspection showed that the pellet consisted mostly of magnetotactic cocci and MMPs. The MMPs exhibited the characteristic mulberry-like, multicellular morphology (Fig. 1B) of previously described MMPs (Farina *et al.*, 1983; Rodgers *et al.*, 1990a). Phase contrast microscopy of the stratified sediment revealed that MMPs reached an abundance of 300 - 400 cm^{-3}. 95% of the entire population of MMPs was present within a 5-10 mm thick layer at the upper boundary of the black-coloured FeS-containing sediment layers. This depth horizon of the Wadden Sea sediment has been shown to remain anoxic throughout the tidal cycle and to contain free sulfide (> 200 μM H$_2$S) (de Beer *et al.*, 2005) as a result of high rates of dissimilatory sulfate reduction. When the hanging drop assay was performed under oxic conditions, MMPs remained viable for 15-30 min.
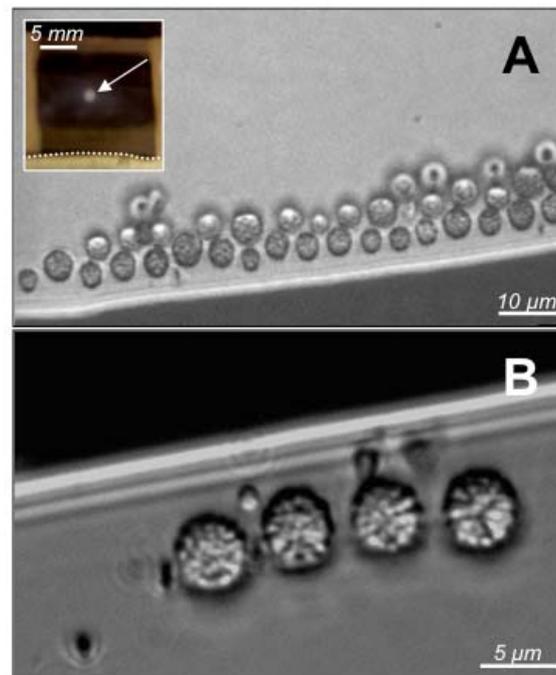
Fig. 1. A. Phase contrast photomicrograph of MMPs from the North Sea accumulating at the edge of a water drop. *Insert:* pellet of magnetotactic bacteria (arrow) formed at the south pole of a ferrite bar magnet at the inner wall of the Plexiglas liner. Dashed line indicates sediment-water interface. **B.** Close up phase contrast photomicrograph showing the mulberry-like, multicellular morphology of the MMPs.

The diameter of the aggregates varied between 4.7 and 6.9 µm with a mean of 5.7 (±1.1) µm. The mean cell number per MMP was 25 (±6; n=102). Based on the *in situ* abundance of the MMP aggregates (see above) and the total bacterial cell numbers (average, $3 \cdot 10^9$ cm^{-3}; Musat *et al.*, 2006; Mussmann *et al.*, 2005), the bacterial cells constituting MMPs account for $3 \cdot 10^{-4}$ % of the total bacterial community in the sediments. The MMPs from the North Sea sediments were highly motile and exhibited clockwise and counter-clockwise rotation as well as a "ping-pong" motion (Rodgers *et al.*, 1990a), also called escape motility (Keim *et al.*, 2007). This latter type of movement is typical for MMPs but unique among all magnetotactic bacteria. It consists of extended phases of swimming along the magnetic field lines interrupted by excursions in the reverse direction during which the aggregates swim at about double the speed as during forward motion (compare the Supplementary Movie).

Ultrastructural analysis of the MMPs by scanning electron microscopy (SEM) revealed that their surface was covered by conspicuous filamentous structures (Fig. 2A) forming a thick capsule around the entire aggregate. So far, only thin filaments, possibly composed of sugars and/or glycoproteins, were observed on the surface of MMPs from Brazilian sites (Farina *et al.*, 1983; Keim *et al.*, 2004a). 3D reconstruction of MMPs and their magnetosome arrangement is difficult with TEM since aggregates would have to be cut in up to 150 thin sections. Applying an

accelerating voltage of 30 kV in combination with the detection of backscattered electrons and a tilting of the specimen allowed 3D visualisation of i) the position of single cells within the aggregates, ii) the intercellular spaces, and iii) the shape and arrangement of the electron-dense magnetosomes crystals (Fig. 2 B-D; SupplFig. 1). All magnetosome crystals were consistently bullet-shaped (Fig. 2C,D). Individual crystals measured 91 (±21) nm in length and 40 (±6) nm in width. The magnetosomes exhibited the same orientation and were arranged in two to three parallel chains that were preferentially located at the periphery of the cells. Most notably, magnetosome chains of different cells within the same multicellular aggregate were aligned with each other such that most chains within one aggregate pointed in the same direction (Fig. 2B,C).
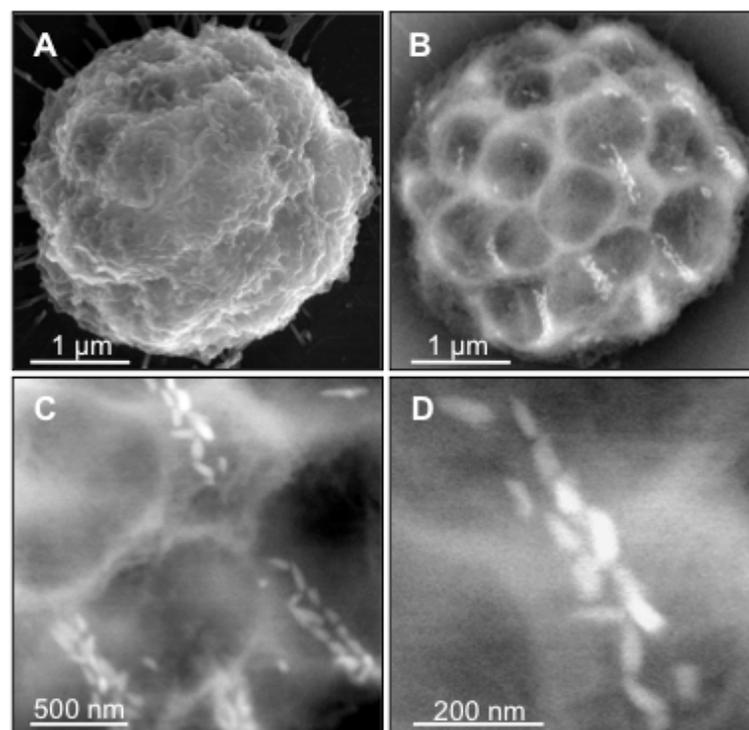

Fig. 2. Ultrastructural analysis of the MMP by SEM at 30 keV **A.** Filamentous surface structures observed in the secondary electron image. **B.** The back-scattered electron image reveals the arrangement of single cells within an individual MMP. The rows of magnetosomes can be distinguished based on their high yield of backscattered electrons which results in a bright signal. **C.** Magnetosomes are typically arranged in 2 or 3 parallel chains. **D.** Close up of two magnetosome chains. All magnetosome crystals are bullet-shaped, 91±21 nm long and 40±6 nm wide, and show a similar orientation.

If the magnetosome chains of the individual cells within one MMP were oriented in a random fashion, the resulting magnetic moment of the entire aggregate would be significantly decreased. However, previous measurements of the magnetisation of intact MMPs revealed a high degree of magnetic optimisation (Winklhofer *et al.*, 2007). The results of our SEM investigations provide the

first direct evidence for a highly ordered arrangement of the magnetosome chains across the entire MMP aggregate and indicate a tight coupling between bacteria of the same aggregate. This coupling may be based on an intercellular signal exchange between the cells.

High resolution energy-dispersive X-ray (EDX) analysis of single magnetosomes revealed a high iron and sulfur contents of the magnetosomes (Fig. 3). Since entire cells were used for these analyses, the electron beam passed through organic cell material, resulting in additional signals of carbon, oxygen, phosphorous and chloride similar to previous studies (Bazylinski *et al.*, 1995; Spring *et al.*, 1998; Keim *et al.*, 2004a, 2005; Lins *et al.*, 2007). Signals of Al and Si are typical for preparations of magnetotactic bacteria and MMPs from sediments (e.g., Spring *et al.*, 1998; Lins *et al.*, 2007) and accordingly were also detected in our samples. In order to confirm that the iron and sulfur signals, but not the oxygen peak originated from magnetosomes, EDX spectra of cell areas adjacent to a magnetosome were recorded as a control. In the vicinity of the magnetosomes, iron and sulfur peaks were very low while an oxygen signal of the same height as in magnetosomes was detected (data not shown). These results imply that the magnetosomes of MMPs from the intertidal sediments of the North Sea consisted of an iron sulfide rather than an iron oxide mineral.
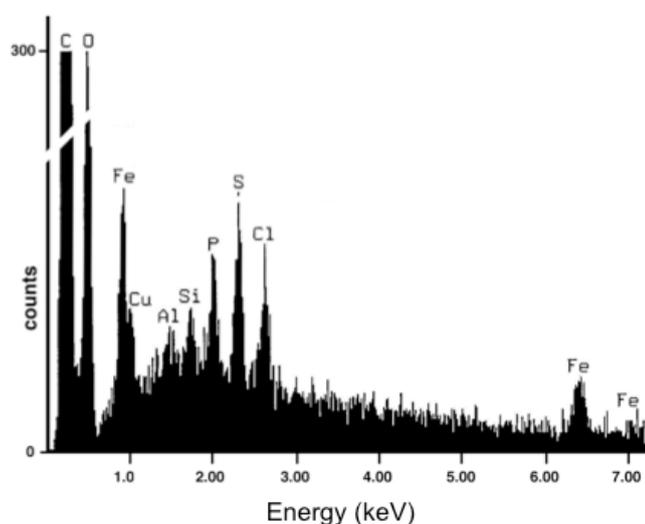
Fig. 3. Energy dispersive X-ray (EDX) analysis of an individual bullet-shaped magnetosome crystal. The prominent iron and sulfur peaks indicate the presence of an iron-sulfur mineral. Carbon, oxygen, aluminium, silicium, phosphor and chloride signals in the spectrum originate from the cellular background and the copper peak from the electron microscopy grid.

With respect to magnetosome morphology and/or composition, the MMPs investigated in the present study differ from all other MMPs characterised to date. Rectangular crystals were observed in various MMPs recovered from Brazilian lagoons and North American sulfidic marine sediments (Farina *et al.*, 1990; Bazylinsky *et al.*, 1993; Keim *et al.*, 2003, 2004b, 2007; Lins *et al.*, 2007;

Martins *et al.*, 2007). The EDX spectra of these rectangular magnetosomes were almost identical to those obtained from the bullet-shaped crystals analysed in the present study and hence also consisted of an iron-sulfide mineral. Based on selected-area electron diffraction (SAED) patterns of transmission electron microscopic preparations this magnetic iron sulfide mineral was identified as greigite (Mann *et al.*, 1990a; Bazylinzki *et al.*, 1993; Pósfai *et al.*, 1998a,b). In addition to the predominantly rectangular crystals, a few arrowhead-shaped greigite crystals have occasionally been observed in some MMPs (Pósfai *et al.*, 1998b). So far, bullet-shaped crystals have only been detected in MMPs from the Itaipu Lagoon in Brazil, but these magnetosomes consisted of an iron oxide (Keim *et al.*, 2003, 2007; Lins *et al.,* 2007). In conclusion, the bullet-shaped iron-sulfide magnetosomes of the MMPs from North Sea intertidal sediments are unique with respect to their specific combination of morphology and chemical composition. Therefore, the phylogenetic position of the North Sea MMPs was compared to the previously described lineages.

MMPs from the North Sea represent a novel phylogenetic lineage

Analyses of almost full-length 16S rRNA genes from magnetotactically enriched MMPs of different sampling dates revealed the presence of the same, single and novel sequence type in samples obtained in the years 2005, 2006 and 2007. The novel phylotype falls within the family *Desulfobacteraceae* of the bacterial subphylum *Deltaproteobacteria*, and represents a novel phylogenetic lineage with $\geq 4\%$ sequence divergence to all previously described MMP sequences (Fig. 4). Thus, the novel phylotype can be considered as a new species according to the commonly accepted criterion of 1.0-1.3% sequence divergence of 16S rRNA gene sequences between different bacterial species (Stackebrandt and Ebers, 2006). Based on the large phylogenetic distance to the recently described '*Candidatus* Magnetoglobus multicellularis' (Abreu *et al.*, 2007) and described genera of the *Desulfobacteraceae*, we propose the name '*Candidatus* Magnetomorum litorale' for the novel type of MMP detected in the North Sea sediments.
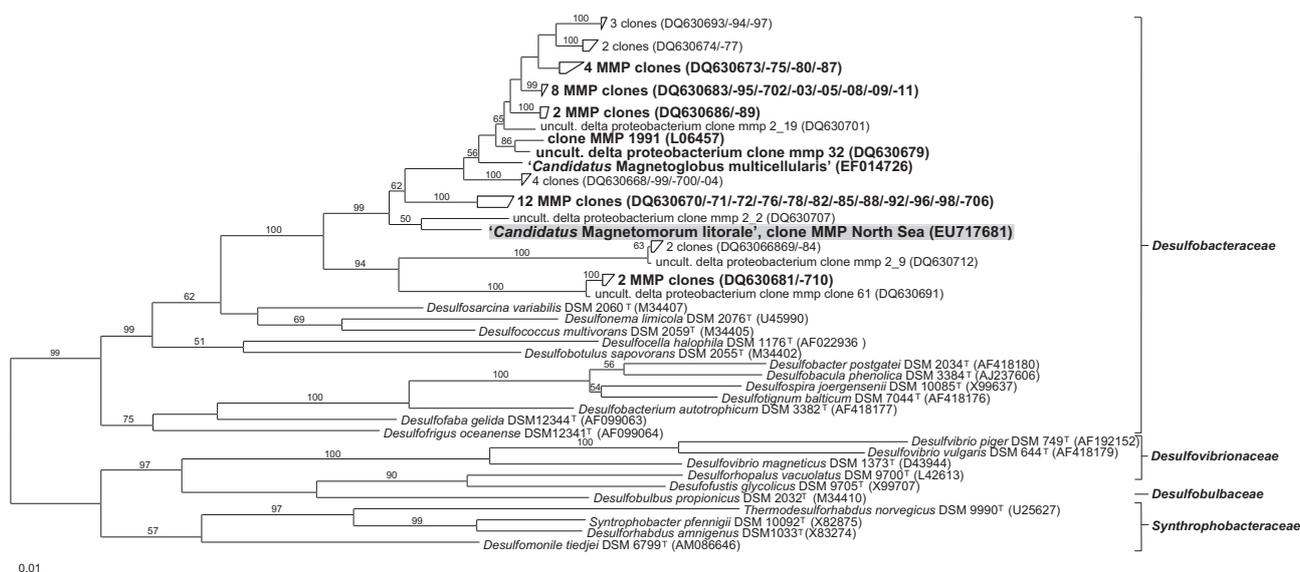
Fig. 4. Phylogenetic tree of 16S rRNA gene sequences showing the relationship of the MMP obtained from the North Sea sediment within its closest relatives. The tree was constructed using the FastdnaML maximum likelihood algorithm as implemented in the ARB program package. Bootstrap values ≥ 50% are indicated at nodes and represent percentages of 100 replicates. Sequences which have already been confirmed by FISH analysis to originate from MMPs are highlighted in bold. Gen Bank accession numbers are given in parentheses. Type species of *Desulfovibrionacae*, *Desulfobulbaceae* and *Synthrophobacteracae* were used as an out-group. Scale bar indicates 1% fixed point mutations per nucleotide base.

Fluorescence *in situ* hybridisation (FISH) was used to assign the 16S rRNA gene obtained from the magnetotactic enrichment to the novel MMP morphotype discovered in the North Sea (Fig. 5A,B). For this purpose the Cy3-labeled oligonucleotide probe mmp-189 (5´-GGCCACCTTTGATAAATTAGC-3´) was established. This probe fell into fluorescence class II (Fuchs *et al.*, 1998) and yielded a strong signal when hybridised in the presence of 30 % formamide (Fig.5 B). Even when allowing for two mismatches, no hybridisation to any other sequence of the databases was predicted using the probe match tool of the RDP-II database. Also, no hybridization to cells of *Desulfobacter hydrogenophilus* DSMZ 3380[T] was observed throughout all our FISH experiments. Counterstaining with DAPI confirmed that probe mmp-189 hybridized exclusively to MMPs in the natural samples (compare Figs. 5A, B). All DAPI-stained MMPs showed a hybridisation signal which indicates that MMPs present in the sediment samples constituted a monospecific population. Similar to other MMPs (Simmons and Edwards, 2007), all cells within a single MMP hybridised with probe mmp-189 (Fig. 5B), suggesting that the aggregate consists of only one phylotype. In the sediment samples, MMPs as well as 70% of the accompanying bacteria yielded hybridisation signals with the eubacterial probes EUBI-III (not shown).
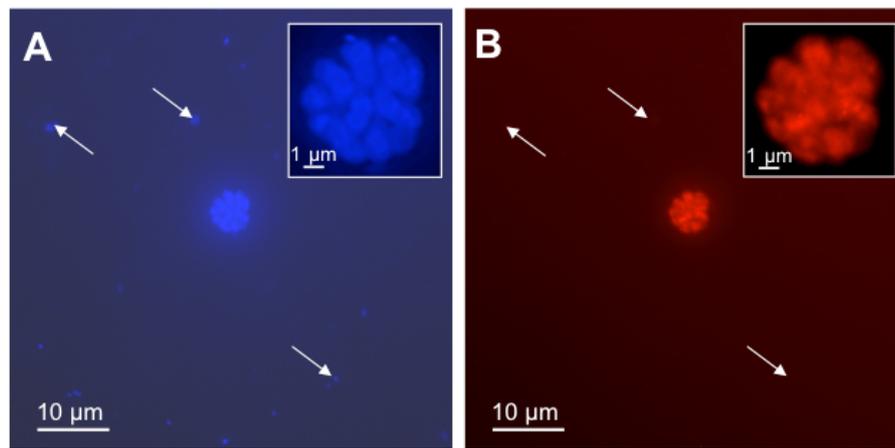
Fig. 5. FISH of MMPs with the specific oligonucleotide probe mmp-189. Cells in magnetotactic enrichments were double stained with DAPI and the carboxymethylrhodamine-labeled probe. **A.** Fluorescence of DAPI; all cells are stained. **B.** Epifluorescence image of Cy3-labeled oligonucleotide probe. All cells of the MMP hybridised with the specific probe mmp-189, whereas the accompanying cells were not stained. Inserts show magnified images of the MMP.

The novel phylotype from the North Sea groups with the 16S rRNA gene sequence of clone mmp 2_2 of an uncultivated *Deltaproteobacterium* (Fig. 4) obtained from a Massachusetts salt marsh across the Atlantic (Simmons and Edwards, 2007). It is unclear, however, whether this phylotype originated from a MMP, since FISH studies with a specific probe were not performed for clone mmp2_2. All MMP sequence types which so far have been confirmed by FISH (depicted in bold face in Fig. 4) were more distantly related to the North Sea MMP phylotype. Interestingly, the closest cultivated relative of the MMPs is the dissimilatory sulfate reducing *Desulfosarcina variabilis* which typically also forms dense aggregates consisting of a variable number of cells (Widdel, 1980).

Similar to our results, natural populations of MMPs from various sites in New England and a Brazilian lagoon were reported to consist of only a single phylotype (DeLong *et al.*, 1993; Abreu *et al.*, 2007). In contrast, a thorough study of MMPs in sediments of a shallow saltwater pool in Little Sippewissett salt marsh (MA, USA) revealed the presence of 9 different sequence types in this latter environment which comprises a variety of different microbial habitats like extensive microbial mats, roots of marsh grass, sulfur filaments of sulfide-oxidizing bacteria and macroscopic aggregates of purple sulfur bacteria. In addition, MMPs may occupy sediment layers with different redox conditions (Simmons and Edwards, 2007). On the 5% sequence divergence level, the sequences present in the salt marsh were considered to be fully sampled and fell into five lineages which diverged sufficiently to be considered different species of MMPs (Simmons and Edwards, 2007). Our first analysis of MMPs from a habitat outside of the American continents yielded a different, novel and well-separated phylogenetic lineage, indicating an even larger phylogenetic

breadth of MMPs. The low diversity of MMPs in the intertidal sediments of the North Sea maybe related to the rather uniform structure of this habitat as compared to Little Sippewissett salt marsh environment.

The presence of dsrAB and aprA genes indicates a potential for sulfate reduction

Since the 16S rRNA gene sequence of MMPs from North Sea sediments was most closely related to sulfate-reducing *Deltaproteobacteria*, we investigated whether our enrichments also harbored the corresponding functional genes, dissimilatory sulfite reductase (*dsr*AB) and adenosine-5´-phosphate reductase (*apr*A). *Desulfosarcina* sp. and *Desulfobulbaceae* represent numerically dominant groups of sulfate-reducing bacteria in the sandy sediments of the Wadden Sea (Mussmann *et al.*, 2005). Therefore, we initially tested whether 16S rRNA gene sequences of other *Deltaproteobacteria* were present in our purified MMP enrichments. However, out of a total of 96 different 16S rRNA gene clones from the enrichments obtained in the years 2005, 2006 and 2007, none was affiliated with sulfate-reducing *Deltaproteobacterium* except the MMP sequence type.

Fragments of the *dsr*AB and *apr*A genes could be amplified employing specific primer pairs. Cloning, sequencing and subsequent phylogenetic analysis of the 2005 and 2006 enrichments revealed that 9 of 10 *dsr*AB-clones as well as 8 of 10 *apr*A-clones belonged to the same *dsr*AB or *dsr*AB sequence type, respectively. The two remaining *apr*A-clones were related to a sequence of the *Gammaproteobacterium Thiothrix* sp. DSM 12730 which is consistent with the observation that 16S rRNA gene sequences affiliated with *Thiothrix* were also detected in the enrichments. The single different sequence type present in among the *dsr*AB clones was distantly affiliated with *Desulfotalea psychrophila* LSv54 of the *Desulfobulbaceae*. However, the prevalent *dsr*AB and *apr*A sequences were found to group with those of the family *Desulfobacteracae* (Fig. 6A, B). The prevalent *apr*A sequence type showed highest sequence identity to the *apr*A gene of *Desulfofrigus* sp. strain HRS La3X (Fig. 6A). The closest relative of the dominant *dsr*AB sequence was clone PIMO 1D01 of an uncultured sulfate reducing bacterium of the Plum island estuary salt marsh (MA, USA) (Fig. 6 B), where MMPs have been observed previously (Bazylinski *et al.*, 1993). The closest cultured sulfate reducer was *Desulfosarcina variabilis* DSM 2060^T (Fig. 6 B).
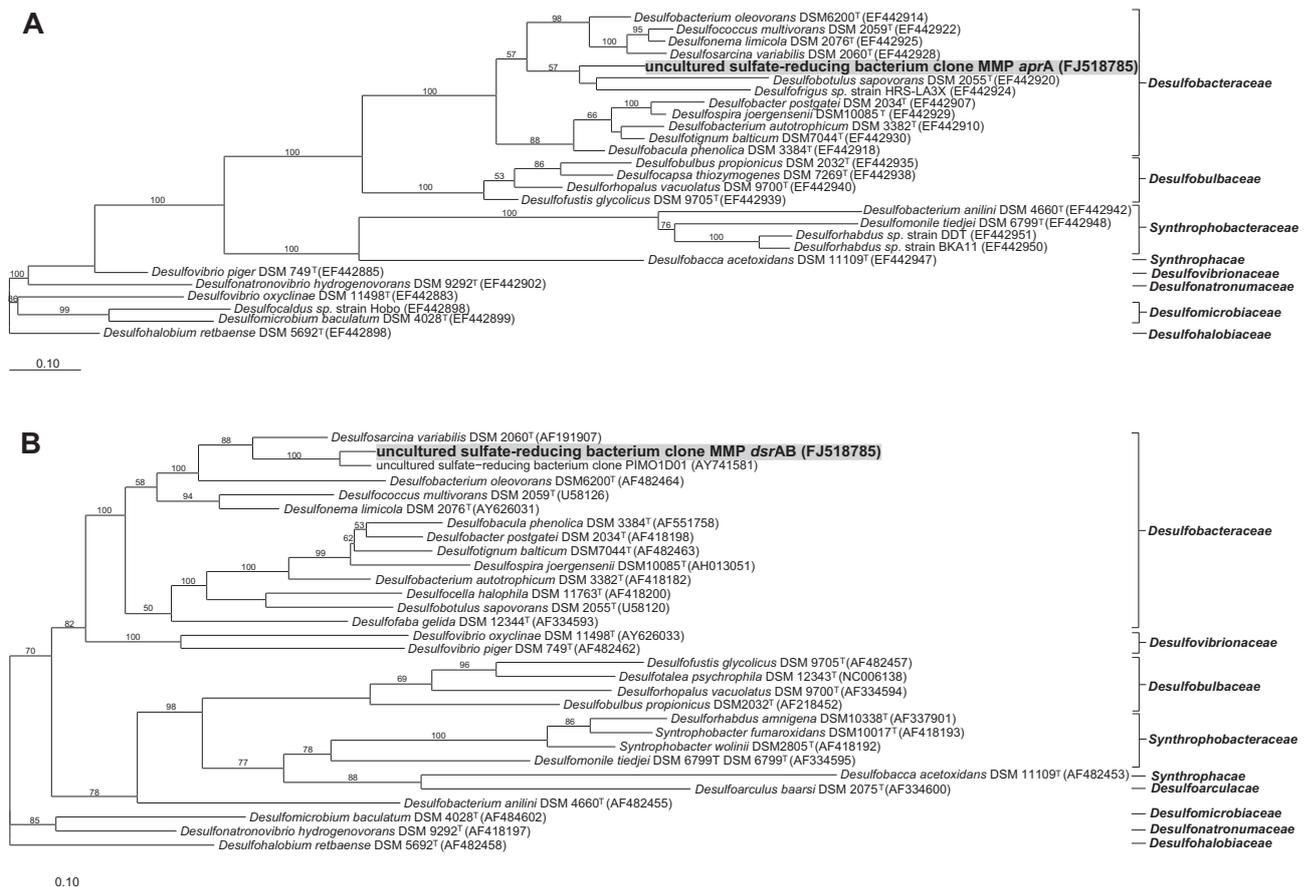
Fig. 6. Phylogenetic affiliation of gene products deduced from genes of dissimilatory sulfate reduction obtained from magnetotactically purified MMP samples. **A.** Phylogenetic tree of AprA. **B.** phylogenetic tree of DsrAB. Trees were constructed using the ProteinML maximum likelihood algorithm as implemented in the ARB program package. Bootstrap values ≥ 50% are indicated at nodes and represent percentages of 100 replicates. Gen Bank accession numbers for the nucleotide sequences are given in parentheses. The DsrAB and AprA sequences of *Desulfohalobium retbaense* DSM 5692[T] (*Desulfohalobiacae*) were used as outgroups. Scale bar indicates 10% fixed point mutations per amino acid position.

Similar to the 16S rRNA gene sequence of the MMP, the prevalent *dsr*AB and *apr*A sequences obtained from the enrichment clustered with the genera *Desulfococcus*, *Desulfonema* and *Desulfosarcina* within the *Desulfobacteracae*. For the family *Desulfobacteracae*, the phylogeny of DsrAB and AprA has been shown to be congruent with the 16S rRNA phylogeny and no evidence for lateral gene transfer has been found for *apr*A or (with the exception of *Desulfobaccula toluolica*) for *dsr*AB (Klein *et al.*, 2001; Mayer and Kuever, 2007). Taken together, our results suggest that MMPs from the North Sea retained the potential for dissimilatory sulfate reduction.

Chemotactic response of MMPs

Without exception, MMPs so far have escaped all laboratory cultivation attempts. In addition, previous studies of natural populations have focused on the ultrastructure, magnetotactic behaviour and phylogenetic diversity, whereas no information on the ecophysiology of MMPs is available. It has been speculated that MMPs occurring in highly productive marine sediments utilise organic carbon substrates for growth (Simmons and Edwards, 2007). The chemotaxis towards different inorganic or organic substrates increases substrate availability (Kiørboe and Jackson, 2001) thereby providing a selective advantage particularly in stratified environments. In addition, analyses of the chemotactic behaviour of motile bacteria yield information on potential growth substrates (Fröstl and Overmann, 1998; Overmann, 2005).

In order to assess the chemotactic behaviour of MMPs from the North Sea sediments, we tested 41 different substrates in parallel capillary assays, following the movement of intact MMPs into the capillaries (Fröstl and Overmann, 1998). The sediment samples for the assays were directly obtained from a single sediment core without magnetic preenrichment to assure identical conditions in all parallels. Since the number of MMPs which could be harvested from the sediments were limited, we initially tested mixtures of substrates. Subsequently, compounds of those mixtures which elicited a positive response were tested individually.

Compared to the control without substrate, no accumulation of MMPs could be observed in capillaries containing lactate, pyruvate, benzoate as well as in those filled with mixtures of sugars, tricarboxylic acid cycle intermediates, proteinogenic aminoacids or alcohols. Similarly, no response was observed towards sulfide (Fig. 7). In contrast, a mixture of formate, acetate, propionate, and butyrate resulted in a significant chemotactic accumulation of MMPs. Tests with the individual constituents of the fatty acid mixture revealed that only acetate and propionate acted as chemoattractants for MMPs and resulted in an up to thousandfold increase of motile aggregates in the respective capillaries (Fig. 7). Similar results were obtained with samples from the years 2005 and 2007. Notably, *Desulfosarcina variabilis*, the closest cultured relative of all known MMPs, is capable of utilising and completely oxidising formate, acetate, propionate and higher fatty acids (Widdel, 1980).
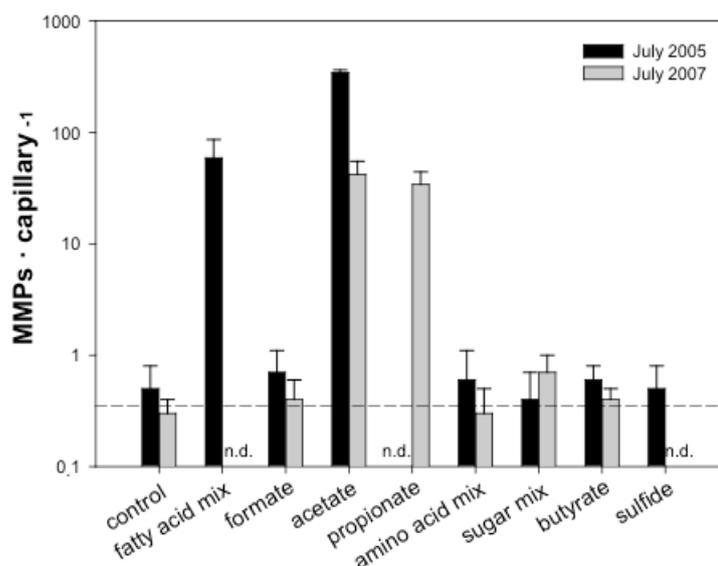
Fig. 7. Chemotactic response of MMPs towards different organic carbon compounds and sulfide. The bars represent mean numbers of MMPs counted per capillaries; error bars depict one standard deviation. Formate, butyrate, the amino acid mix (20 proteinogenic amino acids), the sugar mix (consisting of fructose, glucose, lactose, mannit), and sulfide did not elicit any chemotactic response. Similarly, no accumulation of MMPs was observed in capillaries containing lactate, pyruvate, benzoate, protocatechuate, or TCA intermediates (citrate, fumarate, malate, succinate, 2-oxoglutarate) and alcohols (2,3-butanediol, butanol, ethanol) (data not shown). n.d., not determined. For comparison, dashed line indicates mean value of controls.

Chemotaxis of magnetotactic bacteria is thought to be the reason for their accumulation at particular depths of stratified sediments (Blakemore, 1982; Mann *et al.*, 1990b). Acetate and propionate represent typical, ubiquitous substrates in the pore water of intertidal estuarine sediments where acetate concentrations range between 20 µM (Wellsbury and Parkes, 1995; Albert and Martens, 1997) and 2.5 mM (Albert and Martens, 1997; Hoehler *et al.*, 1999). Typically, carboxylic acid concentrations reach a pronounced peak just below the oxic/anoxic transition zone. Furthermore, concentrations of acetate and propionate are very dynamic and can change rapidly in marine sediments by following the seasonal vertical excursions of the oxic/anoxic transition zone (Albert and Martens, 1997; Hoehler *et al.*, 1999) and the variations in the input of easily degradable organic matter (Wu *et al.*, 1997). In the North Sea sediments, MMPs were consistently observed just below the oxic/anoxic transition zone which could be localized based on the typical change from a brownish to greyish color of the sediment and confirmed by microelectrode measurements (de Beer *et al.*, 2005). The motile MMPs may therefore employ chemotaxis towards acetate and propionate in combination with magnetotaxis to move along chemical gradients in order to accumulate at suitable concentrations of potential substrates. In the tidal flat sediments of the North Sea, sulfate-reducers of the genera *Desulfosarcina* and *Desulfococcus* are one of the dominant

groups of the microbial community of Wadden Sea sediments (Musat *et al.*, 2006) and represent the numerically most abundant sulfate-reducing bacteria (Mussmann *et al.*, 2005). *Desulfosarcina* and *Desulfococcus* typically utilise short chain fatty acids as electron donor for growth (Kuever *et al.*, 2005). A combined chemotactic-magnetotactic behaviour thus would result in a selective advantage of MMPs over sulfate reducers when competing for volatile fatty acids as growth substrates.

Description of 'Candidatus *Magnetomorum litorale*'

Mag.ne'to L. adj. *magnetis*, magnetic; mo.rum L.N.n. *morum*, mulberry; li.to.ra'le L. adj. *litorale*, of the sea shore. The genus name relates to the magnetotactic behaviour and the conspicuous mulberry-like multicellular organization of the magnetotactic prokaryote. The species epithet refers to its specific habitat.

   The diameter of the aggregates ranges between 4.7 and 6.9 µm. Each aggregate consists of 25 ±6 cells. Individual cells form bullet-shaped, magnetosome crystals composed of an iron sulfide mineral. Within the same multicellular aggregate, crystals are aligned in parallel chains and thus are all oriented in the same direction. The aggregates are covered by filamentous surface structures which form a thick capsule. Aggregates of '*Candidatus* Magnetomorum litorale' exhibit chemotaxis toward acetate and propionate and harbour *aprA* and *dsrAB* genes. The habitats of '*Candidatus* Magnetomorum litorale' are anoxic sediment layers of costal intertidal sand flats of the German Wadden Sea.

   All cells of '*Candidatus* Magnetomorum litorale' contain the same 16S rRNA gene sequence and represent a novel phylogenetic lineage which is related to the genus *Desulfosarcina* of the family *Desulfobacteracae* within the order *Desulfobacterales* (subphylum *Deltaproteobacteria*). The associated 16S rRNA gene sequence of the provisional taxon '*Candidatus* Magnetomorum litorale' has been deposited in the GenBank database under the Accession No. EU717681.

Conclusions

Our results indicate that multicellular magnetotactic prokaryotes occur over a broader geographical range than previously assumed. Furthermore, the open intertidal sand flats investigated in the present study represent a novel type of habitat of these multicellular bacteria which so far have only been found in sheltered coastal lagoons or salt marsh sediments. Whereas the MMP from North Sea intertidal sediments resembled previously described cell aggregates with regard to its overall morphology, ultrastructural analysis revealed that the MMP harboured bullet-shaped magnetosome crystals composed of an iron sulfide mineral. Besides this distinct cytological feature, the MMP from the North Sea represents a novel phylotype and, based on its phylogenetic distance to known

sequence types, a new genus and species, for which the name 'Candidatus Magnetomorum litorale' is proposed.

Although previous magnetisation measurements indicated a nonrandom orientation of magnetic dipoles in MMPs, the arrangement of magnetosomes within the aggregates has so far not been visualised. Based on our results, magnetosomes in different bacterial cells of one aggregate form a highly ordered array of aligned chains. This particular ultrastructure of the aggregate provides an explanation for the observed magnetic optimisation. Based on the detection of *dsrAB* and *aprA* genes, the MMP may be capable of sulfate reduction. The North Sea MMPs exhibited a pronounced chemotactic response towards acetate and propionate. The spectrum of organic carbon compounds is very narrow since only 2 out of the 40 organic compounds tested were perceived. Future investigations will reveal under which conditions this combination of tactic behaviour results in a selective advantage of MMPs over purely chemotactic bacterial competitors in vertically stratified marine sediments.

Experimental Procedures

Sampling procedure and magnetotactic enrichment

Sediment cores were collected in July 2005, 2006 and 2007 at low tide in the Wadden Sea near Cuxhaven, northern Germany (53°53.555′ N, 8°40.565′ E). Plexiglas tubes with a length of 50 cm and a diameter of 9.5 cm were employed for sampling. The sampling location is characterised by extended intertidal flats which consist of stratified, sandy sediments. After sampling, the cores were immediately overlaid with sea water, transported to the laboratory and subsequently stored in the dark at 15 °C, which corresponds to the *in situ* temperature in the sediments at the time of sampling. The sediment color was brownish in the upper 4 mm of oxic sediment, then greyish down to 3-5 cm depth and below changed to black colour due to a high FeS content.

For magnetic enrichment of MMPs, a rectangular ferrite magnet (IBS Magnet GmbH, Berlin, Germany) was positioned 3 mm above the water-sediment interface at the outside of the Plexiglas tube of each core with the south pole of the magnet (corresponding to geomagnetic North) facing the core. In this experimental setup, north-seeking magnetotactic bacteria left the sediment and accumulated in front of the magnet such that bacterial cells could be separated from the sediment particles (Fig. 1A, insert). After 1-2 h of collection, the accumulated magnetotactic bacteria were removed with an automatic pipette and resuspended in 200 µl of anoxic seawater.

Phase contrast and scanning electron microscopy

The presence and quantity of MMPs in the magnetotactic enrichments was checked on the basis of their unique morphology using a phase contrast microscope (Axiostar Plus, Carl Zeiss AG, Oberkochen, Germany) at a magnification of 400x and employing the hanging drop method. In this method, 30 µl of pre-enriched sample are placed on a coverslip, the latter is inverted and positioned onto an O-ring on top of a microscope slide such that the drop is enclosed in a chamber which prevents evaporation. In order to increase the sensitivity of the microscopic detection, the magnetotactic bacteria were further concentrated by placing the south pole of a bar magnet at one side of the drop. To analyse the vertical distribution of MMPs in their natural habitat, fresh sediment cores were subsampled in vertical intervals of 5 mm down to a depth of 10 cm employing 20 ml syringes which had their ends cut off. Prior to analysis, the sediment samples were diluted with filter-sterilized habitat water (1:2). For enumeration, 15 µl of the resulting sediment slurry were analysed with the hanging drop method as described above.

The ultrastructure of the MMP from the North Sea was analysed by scanning electron microscopy. Magnetically enriched MMPs were fixed at a dilution of 1:50 in in 140 mM cacodylate buffer (pH 7) containing 2.5% glutardialdhyde and 2.5% NaCl. Subsequently, cells were concentrated by centrifugation at 8.500 x g for 5 min (Centrifuge 5417R, Eppendorf AG, Hamburg, Germany). Two different methods were used for further processing. For scanning electron microscopy of whole cells, drops of the cell suspensions were placed onto polylysine-coated glass slides (Menzel-Glaeser GmbH, Braunschweig, Germany), covered with a coverslip and rapidly frozen with liquid nitrogen. The coverslip was removed with a razor blade and the slide was postfixed with glutardialdehyde in fixative buffer, dehydrated in a graded series of acetone solutions, critical-point dried with liquid CO_2, mounted on stubs and carbon coated (approx. 5 nm) by evaporation. For energy-dispersive X-ray (EDX) analysis, samples were dropped onto formvar-coated electron microscopy grids and air dried. Specimens were examined with a field emission scanning electron microscope (S-4100, Hitachi, Tokyo, Japan) operated at 30 keV, equipped with a back scattered electron detector of the YAG type (Autrata) and a Noran Vantage system for EDX analysis (Thermo Fisher Scientific, Waltham, USA).

DNA extraction and sequence analysis of the 16S rRNA, aprA and dsrAB genes

Cells from magnetically enriched samples were lysed by three consecutive rounds of freezing in liquid nitrogen and thawing at 100°C. Subsequently, DNA was extracted with the DNeasy Tissue Kit (Qiagen, Hilden, Germany) according to the instructions of the manufacturer.

Polymerase chain reactions were performed in the Verti^{TM} Fast Thermal Cycler (Applied Biosystems, Foster City, USA). For amplification of the almost complete 16S rRNA gene the

universal primers 8f and 1492r (Lane, 1991) were used. Fragments of the dissimilatory adenosine-5′-phosphate reductase encoding gene *apr*A (1.4 kb) and dissimilatory sulfite reductase encoding genes *dsr*AB (1.9 kb) were amplified with the specific primer pairs AprA-1-FW/AprA-10-RV (Mayer and Kuever, 2007) and DSR1F/DSR4R (Wagner *et al.*, 1998), respectively. Each reaction contained 1 µM of the primers, 1x buffer, 1x Q-solution (Qiagen), 200µM of each dNTP, 1.75 µM MgCl$_2$, 1.25 U Taq DNA Polymerase (Qiagen), 20 µg BSA and nuclease-free water added to a final volume of 50 µl. Cycling conditions of the step-down PCR with primers 8f and 1492r included an initial denaturation step at 94°C for 3 min, 10 cycles of denaturation for 30 s, primer annealing at 59 °C for 45 s, and elongation at 72°C for 1 min, followed by 20 cycles with the annealing temperature changed to 54°C, and a final extension lasting for 10 min at 72°C. The touch-down PCR protocol with primer pair AprA-1-FW/AprA-10-RV included an initial denaturation step at 94°C for 3 min, 20 cycles of denaturation for 30 s, primer annealing for 60 s which started at 62 °C and decreased by 0.5 °C per cycle, and elongation at 72°C for 90 s, followed by 15 cycles with a constant annealing temperature of 52°C, and a final extension lasting for 10 min at 72°C. PCR with primer pair DSR1F/DSR4R was performed with an initial denaturation step at 94°C for 3 min, 35 cycles of denaturation for 30 s, primer annealing at 54 °C for 30 s, elongation at 72°C for 90s and a final extension lasting for 10 min at 72°C.

The PCR products were cloned into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, USA) and transformed into competent *Escherichia coli* TOP 10 cells (Invitrogen). Plasmids were extracted with the QIAprep Spin Miniprep-Kit (Qiagen), and the presence of inserts was verified by digestion with EcoRI (Fermentas, St. Leon-Rot, Germany). Sequencing reactions were run using the Big Dye V3.1 Terminator Cycle Sequencing Kit (Applied Biosystems) on an ABI Prism 3730 Genetic Analyzer (Applied Biosystems). Related 16S rRNA, *apr*A and *dsr*AB gene sequences were retrieved from the GenBank database using BLAST version 2.0.4 (Altschul *et al.*, 1997). Clones of interest were double pass sequenced with the primers M13f and M13r (Invitrogen). Universal primers 926f (Lane, 1991) and 1055r (Amann *et al.*, 1995) as well as the *dsr*AB clone-specific nested primer D52N (5′-GTGCTTGTCGGCAATCTC-3`) were used for complete coverage of the 16S rRNA and *dsr*AB gene sequence, respectively. The Vector NTI computer package (Invitrogen) was used for sequence assembly and editing. The 16S rRNA, *apr*A and *dsrAB* gene sequences obtained in the present study were checked for chimeras using the CHIMERA-CHECK online analysis program of the RDP-II database (Cole *et al.*, 2005).

Currently, a total of 47 different 16S rRNA gene sequences of a length >1200 bp are assigned to MMPs. These sequences, together with a representative selection of sequences from *Desulfobacteraceae*, *Desulfovibrionaceae*, *Desulfobulbaceae* and *Synthrophobacteracae* type species were recovered from GenBank. After importing them in the 16S rRNA database of the ARB

package (Ludwig *et al.*, 2004), all sequences were aligned with the newly obtained sequence from the North Sea MMP using the ARB FastAligner tool. Manual refinement of the alignment was carried out taking into account the secondary structure information of the rRNA. A phylogenetic tree was inferred using the FastdnaML maximum likelihood algorithm of the ARB program and a termini filter. Bootstrap values were calculated with 100 replicates using the PhyML program. Deduced aminoacid sequences of the *apr*A and *dsr*AB gene fragments obtained in this study (GenBank Accession No. FJ518784 and FJ518785), together with sequences recovered from the GeneBank database, were imported in ARB and aligned with the CLUSTALW Protein alignment tool. Maximum-likelihood trees were calculated based on 536 (AprA) and 773 (DsrAB) amino acid positions using the ProteinML program implemented in ARB and applying the Jones, Taylor and Thornton (JTT) amino acid substitution model matrix. The robustness of trees was inferred by bootstrap analysis after 100 resamplings using PhyML.

Fluorescence in situ *hybridisation*

The oligonucleotide probe mmp-189 targeting the novel MMP phylotype from the North Sea sediment was designed using the probe design tool of the ARB program package. The probe was labeled with carboxymethylrhodamine as the fluorescent dye. Probe specificity was initially assessed using the Probe Match tool of the RDP-II database (Cole *et al.*, 2005). Accessibility of the target site and the fluorescence class of the probe was checked based on data available for *Escherichia coli* (Fuchs *et al.*, 1998; Behrens *et al.*, 2003). Throughout all fluorescence *in situ* hybridisation (FISH) experiments, *Desulfobacter hydrogenophilus* DSMZ 3380[T] was used as a negative control. Furthermore, probe mmp-189 hybridized exclusively to the few MMP morphotypes but never to any other of the ~30,000 bacterial cells inspected during the FISH experiments, indicating that the novel oligonucleotide probe is specific for the MMP phylotype from the North Sea.

Cells were magnetotactically collected as described above, fixed with 2% paraformaldehyde for 30 min at room temperature and filtered onto white polycarbonate filters of 0.2 μm pore size (Millipore, Billerica, USA). Filters were cut into pieces and immersed in 300 μl hybridisation buffer containing 900 mM NaCl, 20 mM Tris-HCl (pH 7.5), 0% to 50 % formamide, 10x Denhardts reagent, 0.01% SDS and 167 nM of the Cy3-labelled probe. Hybridisation lasted for 2h at 46°C and was followed by a washing step in hybridisation buffer for 15 min at 48 °C. Filter sections were rinsed with sterile filtered dd H_2O, then dipped in 80% EtOH for 1 min to reduce background staining and air dried on a microscope slide. After counterstaining with 4´,6-diamidino-2-phenylindole (DAPI), hybridisations were analysed by epifluorescence microscopy with a Zeiss AxioStar Plus microscope equipped with a 100x oil immersion objective (Plan Apochromat

100x/1.4, Zeiss) and a high resolution digital camera (Visitron RT, Visitron Systems GmbH, Puchheim, Germany). Image analysis was performed with Adobe Photoshop (Adobe, San Jose, CA, USA). Diameters of the aggregates were measured using Image J (http://rsb.info.nih.gov/ij/). Hybridisation stringency was tested and optimised by varying the formamide concentrations between 0% and 50%. The highest stringency producing sufficient fluorescence was used. Cy3-labeled eubacterial probes EUB338 I (Amann *et al.*, 1990), II and III (Daims *et al.*, 1999) were used in equimolar mixture for positive hybridisation controls and probe NONEUB (Wallner *et al.*, 1993) for negative controls at the same formamide concentration as mmp-189.

Chemotaxis assays

Chemotactic responses of the MMP were analysed by capillary assays (Fröstl and Overmann, 1998; Overmann, 2005). Ten holes of 3-mm diameter were drilled into the side-wall of 100 ml Meplats glass bottles. The bottles were filled with slurry from a sediment core by taking subcores with a 20 ml syringe which had its end cut off. The sediment was diluted two times with anoxic seawater, then dispersed equally along the bottom of the bottle and allowed to settle for about 1h.

Sterile 100 mM stock solutions of sulfide and various carbon compounds were diluted in filter sterilised (0.2 μm pore size) sample water to final concentrations between 0.5 to 5 mM. Sulfide was tested at 2 mM, and glucose, fructose, mannite and lactose at concentrations of 5 mM. The tricarboxylic acid cycle intermediates 2-oxoglutarate, fumarate, malate, citrate and succinate as well as the 20 proteinogenic L-amino acids were employed at concentrations of 2 mM. The fatty acids formate and acetate were used at 2 mM, butyrate and propionate at 5 mM. Benzoate, lactate and pyruvate were tested at 2 mM, protocatechuate at 0.5 mM and the alcohols ethanol, butanol, and 2,3-butanediol at 5 mM.

Flat rectangular glass capillaries with a length of 50 mm, inside dimensions of 0.1 x 1.0 mm and a capacity of 5 μl (VitroCom, Mountain Lakes, USA) were filled by capillary action with the different substrate solutions and then sealed at one end with plasticine. The open end of each capillary was inserted into a hole of the Meplats bottle such that it extended into the surface water but did not touch the sediment. The capillary was then fixed with plasticine in the appropriate position. Two parallel capillaries were used per chemoattractant. Control capillaries were filled with sterile sample water only. Two separate rounds of experiments were performed in the years 2005 und 2007. The chemotaxis chambers were incubated at room temperature for 12 h in the dark. In order to prevent an interference with the magnetotactic behavior of the bacteria, chemotaxis chambers were incubated inside a box made of permalloy (Mumetal, Vakuumschmelze GmbH, Hanau, Germany) to shield the earth magnetic field. This material represents a magnetic alloy made of about 75% nickel, 15% iron plus copper and molybdenum. Mumetal has a very high magnetic

permeability μ, causing a concentration of the magnetic flow of static or low-frequency magnetic fields in the material which results in a high shielding effectiveness.

After incubation, the capillaries were removed and their open end was sealed immediately with plasticine. Subsequently, the numbers of MMPs per capillary were determined by dark-field microscopy directly inside the capillaries. MMPs that accumulated in the capillaries were mostly non motile at the end of the experiment which facilitated counting. Because of their large size and conspicuous morphology MMPs were easily distinguishable from single bacterial cells or particles.

Acknowledgements

References

Abreu, F., Silva, K.T., Martins, J.L. and Lins, U. (2006) Cell viability in multicellular magnetotactic prokaryotes. *Int Microbiol* 9: 267-271

Abreu, F., Martins, J.L., Silveira, T.S., Keim, C.N., Lins de Barros, H.G.P., Filho, F.J.G. and Lins, U. (2007) '*Candidatus* Magnetoglobus multicellularis', a multicellular, magnetotactic prokaryote from a hypersaline environment. *Int J Syst Evol Microbiol* 57: 1318-1322

Albert, D.B. and Martens, C.S. (1997) Determination of low-molecular-weight organic acid concentrations in seawater and pore-water samples via HPLC. *Marine Chemistry* 56: 27-37

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Stahl, D.A. (1997) Gappen BLAST and PSI BLAST: a new generation of protein databases search programs. *Nucleic Acids Res* 25: 3389-3402

Amann R., Binder B. J., Olson R. J., Chisholm S. W., Devereux R. and Stahl D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56: 1919-1925.

Amann, R., Ludwig, W. and Schleifer, K-H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169

Bazylinski, D.A., Frankel, R.B., Garrat-Reed, A.J. and Mann, S. (1990) Biomineralisation of iron sulfides in magnetotactic bacteria from sulfidic environments. In *Iron Biominerals*. Frankel, R.B., Blakemore, R.P. (eds.). New York, NY, USA: Plenum Press, pp. 239-255

Bazylinski, D.A., Garrat-Reed, A.J., Abedi, A. and Frankel, R.B. (1993) Copper association with iron sulfide magnetosomes in a magnetotactic bacterium. *Arch Microbiol* 160: 35-42

Bazylinski, D.A., Frankel, R.B., Heywood, B.R., Mann, S., King, J.W., Donaghay, P.L. and Hanson, A.K. (1995) Controlled biomineralisation of magnetite (Fe_3O_4) and greigite (Fe_3S_4) in a magnetotactic bacterium. *Appl Environ Microbiol* 61: 3232-3239

Behrens, S., Fuchs, B.M., Mueller, F. and Amann, R. (2003) Is the in situ accessibility of the 16S rRNA of Escherichia coli for Cy3-labeled oligonucleotide probes predicted by a three-dimensional structure model of the 30S ribosomal subunit? *Appl Environ Microbiol* 69: 2899-2905

Blakemore, R.P. (1982) Magnetotactic bacteria. *Ann Rev Microbiol* 36: 217-238

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., Garrity, G.M. and Tiedje, J.M. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* 33: 294-296

Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. and Wagner, M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: Development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* 22: 434-444.

de Beer, D., Wenzhöfer, T., Ferdelman, T., Boehme, S.E., Huettel, M., van Beusekom, J.E.E., Böttcher, M.E., Musat, N. and Dubilier, N. (2005) Transport and mineralization rates in North Sea sandy intertidal sediments, Sylt-Rømø Basin, Wadden Sea. *Limnol Oceanogr* 50: 113-127

DeLong, E.F., Frankel, R.B. and Bazylinski, D.A. (1993) Multiple evolutionary origins of magnetotaxis in bacteria. *Science* 259: 803-806

Esquivel, D.M.S., Lins de Barros, H.G.P., Farina, M., Aragaõ, P.H.A. and Danon, J. (1983) Microorganismes magnétotactiques de la region de Rio de Janeiro. *Biol. Cell.* 47: 227-234

Farina, M., Lins de Barros, H., Motta de Esquivel, D. and Danon, J. (1983) Ultrastructure of a magnetotactic microorganism. *Biol. Cell.* 48: 85-88

Farina, M., Motta S. Esquivel, D. and Lins de Barros, H.G.P. (1990) Magnetic iron-sulphur crystals from a magnetotactic microorganism. *Nature* 343: 256-258

Fröstl, J.M. and Overmann, J. (1998) Physiology and tactic response of the phototrophic consortium "Chlorochromatium aggregatum" Arch Microbiol 169: 129-135

Fuchs, B.M., Wallner, G., Beisker, W., Schwippl, I., Ludwig, W. and Amann, R. (1998) Flow cytometric analysis of the in situ accessibility of *Escherichia coli* 16S rRNA for fluorescently labeled oligonucleotide probes. *Appl Environ Microbiol* 64: 4973-4982

Greenberg, M., Canter, K., Mahler, I. and Tornheim, A. (2005) Observation of magnetoreceptive behavior in a multicellular magnetotactic prokaryote in higher than geomagnetic fields. *Biophys J* 88: 1496-1499

Hoeler, T.M., Albert, D.B., Alperin, M.J. and Martens, C.S. (1999) Acetogenesis from CO_2 in an anoxic marine sediment. *Limnol Oceanogr* 44: 662-667

Keim, C.N., Lins, U. and Farina, M. (2003) Iron oxide and iron sulfide crystals in magnetotactic multicellular aggregates. *Acta Microscopica* 12: 3-4

Keim, C.N., Abreu, F., Lins, U., Lins de Barros, H. and Farina, M. (2004a) Cell organisation and ultrastructure of a magnetotactic multicellular organism. *J Struct Biol* 145: 254-262

Keim, C.N., Martins, J.L., Abreu, F., Rosado, A.S., Lins de Barros, H., Borojevic, R., Lins, U. and Farina, M. (2004b) Multicellular life cycle of magnetotactic multicellular prokaryotes. *FEMS Microbiol Lett* 240: 203-208

Keim, C.N., Solórzano, G., Farina, M. and Lins, U. (2005) Intracellular inclusions of uncultured magnetotactic bacteria. *Int Micrbiol* 8: 111-117

Keim, C.N., Martins, J.L., Lins de Barros, H., Lins, U. and Farina, M. (2007) Structure, behaviour, ecology and diversity of multicellular magnetotactic prokaryotes. In *Magnetoreception and Magnetosomes in Bacteria*. Schüler, D. (ed.). Berlin, Heidelberg, Germany: Springer, pp. 103-132

Kiørboe, T. and Jackson, G.A. (2001) Marine snow, organic solute plumes, and optimal chemosensory behavior of bacteria. *Limnol Oceanogr* 46: 1309-1318

Klein, M., Friedrich, M., Roger A.J., Hugenholtz, P., Fishbain, S., Abicht, H., Blackall,L.L., Stahl, D.A. and Wagner, M. (2001) Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J Bacteriol* 183: 6028-6035

Kuever, J., Rainey, F.A. and Widdel, F. (2005) Family I. *Desulfobacteriaceae* fam. nov. In *Bergey´s Manual of Systematic Bacteriology*. Brenner D.J., Krieg, N.R., Staley, J.T. (eds.). 2nd ed. Vol. 2, Part C. p. 959-988

Lane, D.J. (1991) 16S/23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*. Stackebrandt, E., Goodfellow, M (eds.). Wiley, Chichster. p. 115-175

Lins, U. and Farina, M. (1999) Organisation of cells in magnetotactic multicellular aggregates. *Microbiol Res* 154: 9-13

Lins, U., Keim, C.N., Evans, F.F., Buseck, P.R. and Farina, M. (2007) Magnetite (Fe$_3$O$_4$) and greigite (Fe$_3$S$_4$) crystals in multicellular magnetotactic prokaryotes. *Geomicrobiol J* 24: 43-50

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, H., Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüßmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Struckmann, N., Viklbig, A., Lenke, M., Ludwig, T., Bode, A. and Schleifer, K-H. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32: 1363-1371

Mann, S., Sparks, N.H.C., Frankel, R.B., Bazylinski, D.A. and Jannsch, H.W. (1990a) Biomineralisation of ferrimagnetic greigite (Fe$_3$S$_4$) and iron pyrite (FeS$_2$) in a magnetotactic bacterium. *Nature* 343: 258-261

Mann, S., Sparks, N.H.C. and Board, R.G. (1990b) Magnetotactic bacteria: microbiology, biomineralisation, paleomagnetism and biotechnology. *Adv Microbiol Physiol* 31: 125-181

Martins, J.L., Silveira, T.S., Abreu, F., Silva, K.T., da Silva-Neto, I.D. and Lins, U. (2007) Grazing protozoa and magnetosome dissolution in magnetotactic bacteria. *Environ Microbiol* 9: 2775-2781

Mayer, B. and Kuever, J. (2007) Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5′-phosphate (APS) reductase from sulfate-reducing prokaryotes-origin and evolution of the dissimilatory sulfate reduction pathway. *Microbiology* 153: 2026-2044

Musat, N., Werner, U., Knittel, K., Kolb, S., Dodenhof, T., van Beusekom, J.E.E., de Beer, D., Dubilier, N. and Amann, R. (2006) Microbial community structure of sandy intertidal sediments in the North Sea, Sylt-Rømø Basin, Wadden Sea. *Syst Appl Microbiol* 29: 333-348

Mussmann, M., Ishii, K., Rabus, R. and Amann, R. (2005) Diversity and vertical distribution of cultured and uncultured *Deltaproteobacteria* in an intertidal mud flat of the Wadden Sea. *Environ Microbiol* 7: 405-418

Overmann, J. (2005) Chemotaxis and behavioural physiology of not-yet-cultivated microbes. In *Methods in Enzymology* 397. Elsevier, San Diego. p. 133-146.

Pósfai, M., Buseck, P.R., Bazylinski, D.A. and Frankel, R.B. (1998a) Reaction sequence of iron sulfide minerals in bacteria and their use as biomarkers. *Science* 280: 880-883

Pósfai, M., Buseck, P.R., Bazylinski, D.A. and Frankel, R.B. (1998b) Iron sulfides from magnetotactic bacteria: structure, composition and phase transitions. *Am Mineral* 83: 1469-1481

Rodgers, F.G., Blakemore, R.P., Blakemore, N.A., Frankel, R.B., Bazylinski, D.A., Maratea, D. and Rodgers, C. (1990a) Intercellular structure in a many-celled magnetotactic prokaryote. *Arch Microbiol* 145: 18-22

Rodgers, F.G., Blakemore, R.P., Blakemore, N.A., Frankel, R.B., Bazylinski, D.A., Maratea, D. and Rodgers, C. (1990b) Intercellular junctions, motility and magnetosome structure in a multicellular magnetotactic prokaryote. In *Iron Biominerals*. Frankel, R.B., Blakemore, R.P. (eds.). Plenum Press, New York. p. 231-237

Silva, K.T., Abreu, F., Almeida, F.P., Keim, C.N., Farina, M. and Lins, U. (2007) Flagellar apparatus of south seeking many celled magnetotactic prokaryotes. *Microsc Res Tech* 70: 10-17.

Simmons, S.L., Sievert, S.M., Frankel, R.B., Bazylinski, D.A. and Edwards, K.J. (2004) Spatiotemporal distribution of marine magnetotactic bacteria in a seasonally stratified costal salt pond. *Appl Environ Microbiol* 70: 6230-6239

Simmons, S.L. and Edwards, K.J. (2007) Unexpected diversity in populations of the many-celled magnetotactic prokaryote. *Environ Microbiol* 9: 206-215

Spring, S., Lins, U., Amann, R., Schleifer, K-H., Ferrera, L.C.S., Esquivel, D.M.S and Farina, M. (1998) Phylogenetic affiliation and ultrastructure of uncultured magnetic bacteria with unusually large magnetosomes. *Arch Microbiol* 169: 136-147

Stackebrandt, E., and J. Ebers (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today* 33: 152-155

Thamdrup, B. (2000) Bacterial manganese and iron reduction in aquatic systems. *Adv Microb Ecol* 16: 41-84

Wagner, M., Roger, A.J., Flax, J.L., Brusseau, G.A. and Stahl, D.A. (1998) Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J Bacteriol* 180: 2975-2982

Wallner, G., Amann, R. and Beisker, W. (1993) Optimizing fluorescent in situ hybridisation with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* 14: 136-143

Wellsbury, P. and Parkes, R.J. (1995) Acetate bioavailability and turnover in an estuarine sediment. *FEMS Microbiol Ecol* 17:85-94

Winkelhofer, M., Abraçado, L.G., Davila, A.F., Keim, C.N. and Lins de Barros, H.G.P. (2007) Magnetic optimization in a mulicellular magnetotactic organism. *Biophys J* 92: 661-670

Widdel, F. (1980) Anaerober Abbau von Fettsäuren und Benzoesäure durch neu isolierte Arten Sulfat-reduzierender Bakterien. Dissertation. Georg August-Universität zu Göttingen. Lindhorst/Schaumburg-Lippe, Göttingen, Germany, 443 p.

Wu, H., Green, M. and Scranton, M.I. (1997) Acetate cycling in the water column and surface sediment of Long Island Sound following a bloom. *Limnol Oceanogr* 42: 705-713

Supplementary Material
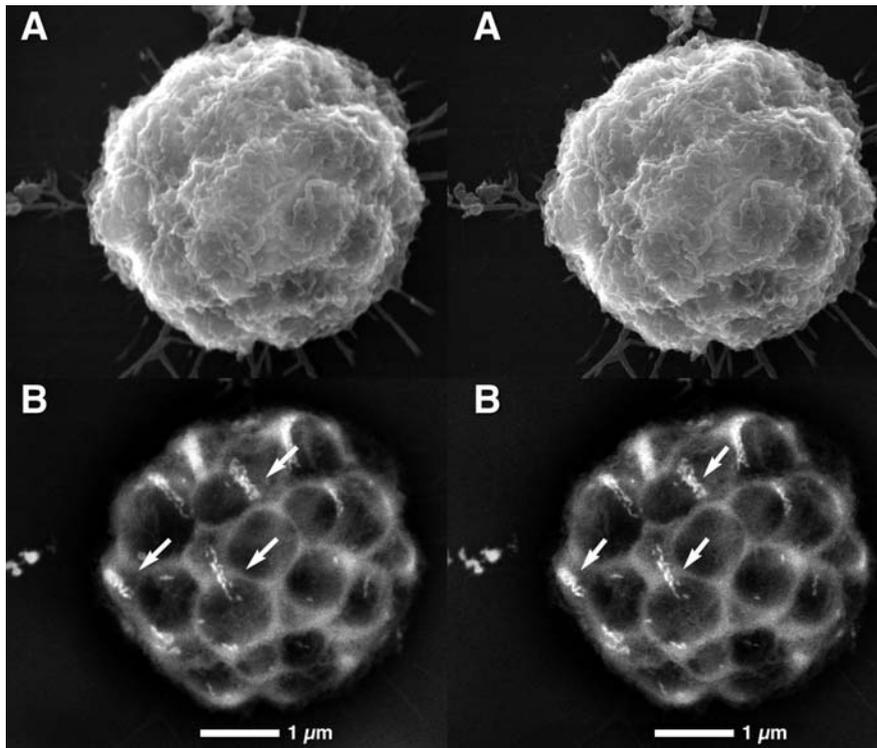
Suppl. Fig. 1. Stereo pairs of scanning electron micrographs of a MMP operated at 30 keV **A.** Filamentous surface structures observed in the secondary electron image. **B.** Back scattered electron image reveals the arrangement of single cells within an individual MMP. The rows of magnetosomes can be distinguished based on their high yield of backscattered electrons resulting in a bright signal.
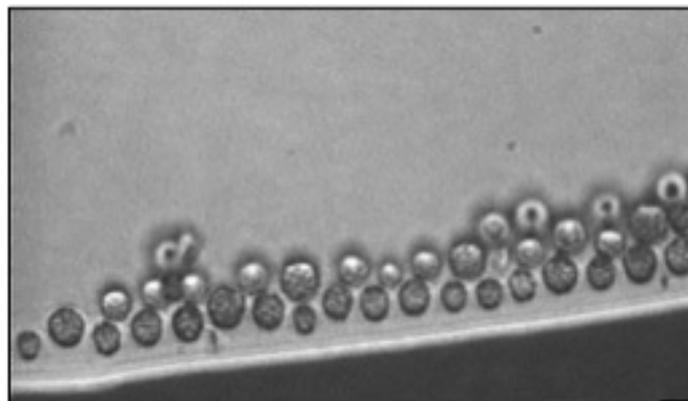
Suppl. Movie. Magnetotaxis and typical "ping-pong" motion of MMPs freshly isolated from North Sea intertidal sandy sediments. This type of motion consists of extended phases of swimming along the magnetic field lines which are interrupted by excursions in the reverse direction

Chapter 6

Discussion

Epibiont symbiosis genes related to bacterial virulence factors and exopolysaccaride synthesis have functional implications for prokaryotic multicellularity

Based on the analysis of all available green sulfur bacterial genomes, the genome of the "*C. aggregatum*" epibiont contains 186 unique open reading frames (ORFs) which are missing in the free-living green sulfur bacteria and hence are of potential relevance for prokaryotic multicellularity. While the majority (i.e., 99) of these epibiont-specific ORFs code for hypothetical proteins with unknown functions, unique epibiont genes that match known bacterial virulence factors provide promising targets for future functional investigations of the interactions between the partner cells in phototrophic consortia. Moreover, this finding indicates that genetic modules exhibiting similarity to virulence factors of typical proteobacterial pathogens have been transferred laterally to nonrelated bacteria and are employed in multicellular interactions between prokaryotes.

The unique genes suggested to encode haemagglutinins and hemolysin-type Ca^{2+}-binding proteins are of particular interest since they are associated with the cell membrane and mediate the attachment of pathogenic bacteria to their host cells (Relman *et al.*, 1989; Kajava *et al.*, 2001; El-Azami-El-Idrissi *et al.*, 2003) or the adherence of non-pathogenic bacteria to surfaces and other microorganisms (Dalisay *et al.*, 2006). Based on their frequent involvement in host-pathogen-interactions and the results of the consortia disaggregation studies using the Ca^{2+} chelating agent EGTA, the haemagglutinin-like as well as the hemolysin-type Ca^{2+}-binding proteins of the epibiont most likely participate in the cell-cell-adhesion between the partner bacteria of phototrophic consortia. Interestingly, Ca^{2+}-ions are also likely to be involved in maintaining the multicellular structure in magnetotactic multicellular prokaryotes (MMPs) as also revealed by their disaggregation upon EGTA treatment, which indicates that the intercellular junctions observed in MMPs (Keim *et al.*, 2004) could be stabilised by calcium binding proteins (Abreu *et al.*, 2006). Furthermore, a unique VCBS-domain (Yousef and Espinosa-Urgel, 2007) containing ORF is preferentially transcribed in the symbiotic state of the epibiont. It resembles its known counterparts in that it contains cadherin-like domains which are known to mediate in Ca^{2+}-dependent cell-cell adhesion in eukaryotes (Takeichi, 1990).

Furthermore, the epibiont genome analysis revealed a conspicuous group of unique genes including seven ORFs closely located to each other which enable the formation of extracellular polysaccharides. Their gene products are likely to participate in capsular exopolysaccharide biosynthesis or transmembrane polysaccharide export and seem to be involved in the production of extracellular material contributing to the formation of cell aggregates. In fact, electron microscopic studies showed that cell attachment in phototrophic consortia is mediated by a dense interconnecting network of up to 150 nm long and 3 nm wide hair-like filaments which cover the epibiont cells and in turn form an elastic capsule enclosing the central bacterium (Fig. 1A.; Wanner *et al.*, 2008). Furthermore, thick capsules have been observed in phototrophic consortia from natural populations (Overmann *et al.*, 1998). Similarily how in phototrophic consortia, ultrastructural analysis of the MMP from the North Sea revealed that its surface is also covered by conspicuous filamentous structures forming a thick capsule around the entire aggregate (Fig. 1B) most likely consisting of exopolysaccarides as its main component (Keim *et al.*, 2004a). The synthesis of extracellular capsular material by the consortia epibionts and the MMP cells therefore may not only contribute to cell-cell adhesion but could also confer an additional protective mechanism like e.g. an osmotic barrier to the bacterial cells representing a major advantage of living in a multicellular state under the continuously changing environmental conditions. Therefore capsule formation could have been an important function for the development of prokaryotic multicellularity in general.
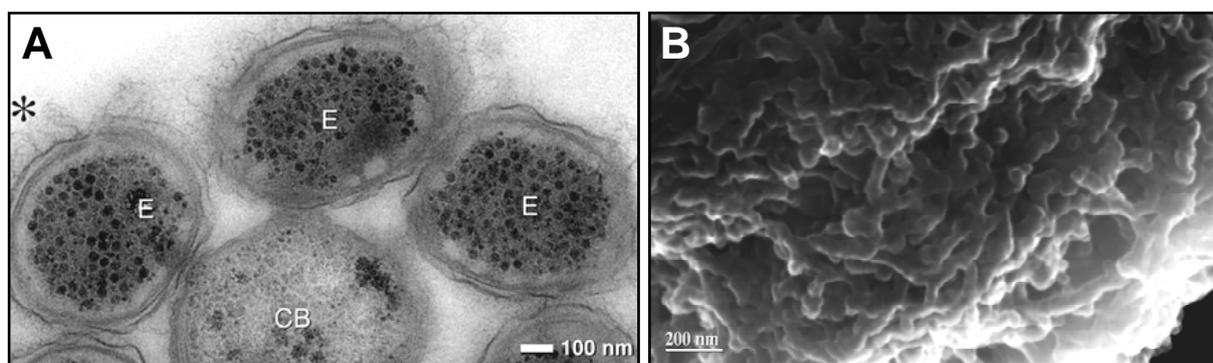
Fig. 1. A. Hair-like filaments (asterisk) forming a capsule interconnecting the epibiont cells in the phototrophic consortium "*C. aggregatum*" (modified after Wanner *et al.*, 2008). **B.** Detail of the filamentous capsule covering the MMP from the North Sea (this study).

A filamentous capsule is also likely to participate in the formation of the dense mucous biofilm of phototrophic consortia invariably observed on the walls of the culture flasks (Pfannes *et al.*, 2007). The extracellular matrix of similar biofilms was found to be composed mainly of exopolysaccarides and proteins (Gottig *et al.*, 2009). The major proteins identified in this matrix were haemagglutinin-like proteins, providing the first indication that multicellular adhesion in highly structured bacterial

aggregates indeed could be mediated by a coordinated action of cell surface haemagglutinins and exopolysaccharides. To reveal whether the presence of genes encoding bacterial virulence factors and exopolysaccharide synthesis constitute a prerequisite for the formation of phototrophic consortia and thus also occur in distantly related types of consortia epibionts, a future approach including the separation of single phototrophic consortia from the environment using micromanipulation followed by multiple displacement amplification of the genomic DNA and subsequent cloning and Southern blot screening will be employed.

The fraction of unique epibiont genes is small due to the preadaptation of free-living relatives to interspecific interaction

Comparisons of the epibiont genome with all available genomes of free-living green sulfur bacteria identified a comparatively low number of unique genes, indicating that the fraction of niche-specific genes in the epibiont is small. Therefore it is likely that the bacterial interaction in phototrophic consortia involves novel, yet unidentified, types of proteins which await future discovery. In order to understand the transition between free-living and symbiotic state, preadaptation of the ancestor of the "*C. aggregatum*" epibiont and access to the existing gene pool of free-living green sulfur bacteria by lateral gene transfer hypothetically represent two key features of the evolution of multicellular bacterial interactions in phototrophic consortia. The possible preadaptation could have caused the limited number of novel symbiosis genes detected in the epibiont genome and most likely predisposed its ancestor to a symbiotic lifestyle. In line with these results, the most promising candidates for genes invoved in the symbiosis had homologues in free-living relatives of the epibiont and represent genes that matched known bacterial virulence factors.

The haemagglutinin-like gene product of the non-unique ORF Cag_1053 cross-linked with a filamentous haemagglutinin/Type IV pilus protein of the central bacterium indicating a direct role in the specific cell-cell-binding to the betaproteobacterial partner. This result provides the first experimental evidence that virulence factor-like proteins likely to be associated with the outer membrane mediate multicellular adhesion in non-pathogenic bacteria. Furthermore, numerous periplasmatic tubules of the central bacterium structurally similar to mannose sensitive hemagglutinin (MSHA) type IV pili mediating cell-cell adhesion (Dalisay *et al.*, 2006) were observed to extend toward the epibiont surface interconnecting both bacterial partners (Fig. 2; Wanner *et al.*, 2008). Based on the these results, the presence of genes encoding proteins resembling virulence factors in the central bacterium draft genome (unpublished data) will be analysed by a similar screening approach conducted in the present study to provide novel insights into evolution, distribution and involvement of virulence factor-like genes and proteins in symbiotic interactions involving betaproteobacteria.
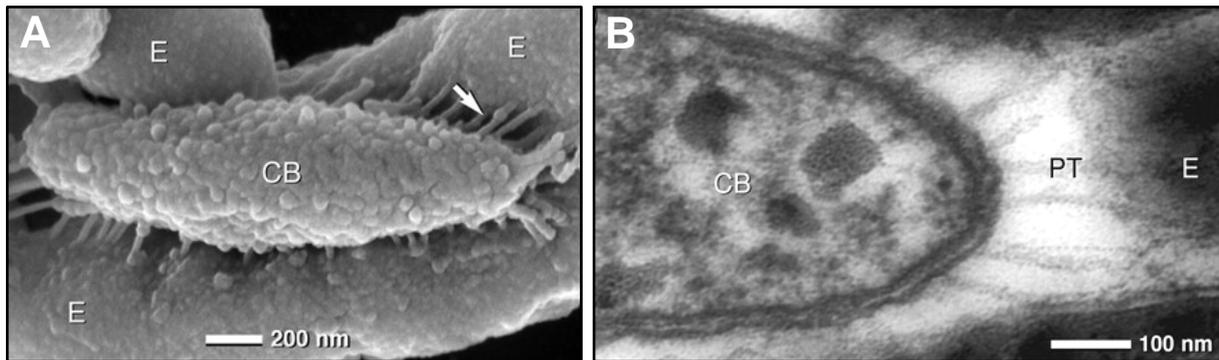
Fig. 2. A. Scanning and **B.** transmission electron microscopy of periplasmatic tubules (PT) formed by the outer membrane of the central bacterium in contact with the epibiont cell surface (modified after Wanner *et al.*, 2008).

Moreover, the non-unique epibiont protein encoded by ORF Cag_1285 was found to be localised in the cytoplasmic membrane or periplasmic space. It is only distantly related to its homologues in green sulfur bacterial relatives of the epibiont. The small protein was detected only in the symbiotic state in which the chlorosomes are absent from the inner face of the epibiont cytoplasmic membrane at the site of attachment to the central bacterium (see Chapter 1/Fig. 3A), whereas chlorosomes are evenly distributed along the cytoplasmic membrane in epibiont cells from pure cultures (Wanner *et al.*, 2008) like in all other green sulfur bacteria. The gene product of Cag_1285 could therefore be involved in the intercellular sorting of chlorosomes in the epibiont by modifying its cytoplasmic membrane in a specific way to exclude chlorosomes from the cell-cell adhesion site. This protein thus represents the first target for the elucidation of this cell biological process in the epibiont. Interestingly, a layered structure composed of regularly arranged elements is attached to the inner side of the cytoplasmatic membrane at all cell-cell contact points of the epibiont and the central bacterium (see Chapter 1/Fig. 3B; Wanner *et al.*, 2008). Since chlorosomes are absent precisely at the positions of this so-called epibiont contact layer, it is likely to be responsible for the chlorosome sorting (Vogl *et al.*, 2006; Wanner *et al.*, 2008). A future subcellular localisation analysis of the Cag_1285 gene product using immunogold staining will reveal whether it is involved in the formation of this contact layer.

The central bacterium contains a similar, exceptionally large layered structure of yet unknown function called central bacterium crystal (see Chapter 1/Fig. 4A; Wanner *et al.*, 2008). Since this paracrystalline structure is located in the cytoplasm, a direct involvement in adhesion to the epibionts seems to be unlikely. Interestingly, the central bacterium crystal exhibits a substructure resembling the overproduced chemotaxis receptor Tsr in complex with histidine autokinase CheA and adaptor protein CheW of *Escherichia coli* (Zhang *et al.*, 2007; Wanner *et al.*, 2008). Almost identical striated structures were also detected in a multicellular magnetotactic prokaryote (MMP)

(see Chapter 1/Fig. 4B; Silva *et al.*, 2007). Since subcellular structures very similar to those found in "*C. aggregatum*" also occur in the nonrelated MMPs, there is a first indication that they could be key features relevant for the evolution of prokaryotic multicellularity which are generally employed in inter- and intraspecific bacterial interactions. Electron microscopic analyses of different environmental populations will reveal if the particular ultrastructures found in "*C. aggregatum*" and the MMP from Brazil (Silva *et al.*, 2007) generally represent characteristic features for multicellular bacteria. Additionally, a future genome comparison of the central bacterium and the MMP with their closest cultivated relatives *Rhodoferax ferrireducens* T118 (Risso *et al.*, 2009) and *Desulfosarcina variabilis* DSM 2060^T (not sequenced), respectively, will reveal if similar genetic modules possibly involved in multicellular interaction have been transferred laterally from nonrelated bacteria or if a preadaptation of the common ancestor exists, which hypothetically could be the case at least with the cell aggregate forming *Desulfosarcina variabilis* (Widdel, 1980).

Future studies of interesting non-unique epibiont genes and their products should also include comparative analyses of their homologues in non-symbiotic green sulfur bacteria in order to elucidate the evolutionary changes which lead to the emergence of the symbiosis and heterologous multicellularity in phototrophic consortia. Such analyses will also have important implications for understanding other bacterial symbioses, can provide the basis for the study of other examples of prokaryotic multicelluar associations like the MMP and will help to elucidate how an existing gene inventory of free-living bacteria was modified to establish mutual interactions with other prokaryotes.

Symbiosis-depended regulation of epibiont gene expression predominantly involves the central metabolism and housekeeping functions

The pronounced changes in gene expression of symbiotic compared to the free-living "*C. aggregatum*" epibionts predominantly involve non-unique genes encoding components of central metabolic pathways and housekeeping genes rather than the unique symbiosis genes. Such an altered expression of basic cellular functions was also observed in the virulence of human pathogens (Bowe *et al.*, 1998) and thus likewise may be important for the formation of bacteria-bacteria symbiosis in phototrophic consortia with regard to a potential physiological coupling between the epibiont and the central bacterium. On the other hand, the regulation of potential symbiosis genes actually may be dispensable since epibionts seem to be specifically adapted to the life in association with the central bacterium and so far have not been detected in a free-living state in nature (Glaeser and Overmann, 2004). Alternatively, the symbiotic association could have been stable in nature for a sufficiently long time and thus a regulation of the symbiosis genes is no longer of selective advantage.

Regardless of the comparatively large fraction of genes differentially expressed in symbiotic epibiont, its genome encodes only a few proteins for environmental sensing and regulatory responses similar to other green sulfur bacteria (Eisen *et al.*, 2002). However, the differences observed between symbiotic and free-living epibionts must involve so far unknown distinct regulatory processes causing cell-cell recognition, adhesion and intracellular sorting as well as the symbiosis-dependent regulation of the cellular metabolism and housekeeping functions. It therefore has to be investigated whether the pronounced changes in gene expression between the symbiotic and free-living state of the epibiont are controlled by the few regulatory proteins identified to date or if a substantial fraction of regulatory proteins so far remained elusive.

Metabolic coupling between the bacterial partners in phototrophic consortia and multicellular magnetotactic prokaryotes

As basis for the symbiosis in phototrophic consortia, an internal synthrophic sulfur cycle was proposed similar to the syntrophic associations of sulfide-oxidising green sulfur bacteria and sulfate-reducing bacteria (Pfennig, 1980). Hypothetically sulfide from sulfate reduction of the central rod thus could be available for the epibiont, which in turn would provide sulfate from the photooxidation of sulfide to the central bacterium (Pfennig, 1980). Since the phylogenetic affiliation of the central bacterium revealed that it belongs to the betaproteobacteria (Fröstl and Overmann 2000; Kanzler *et al.* 2005), a subphylum in which sulfate-reduction was not detected so far, a closed sulfur cycle in phototrophic consortia seems to be unlikely. Moreover, in a preliminary analysis of the central bacterium draft genome no genes encoding key enzymes for sulfate-reduction could be detected (unpublished data). Theoretically interspecies hydrogen transfer may occur between the central bacterium and the epibionts like in aggregates of fermentative bacteria with methonogenic archaea or sulfate reducing bacteria, but there is also no evidence for this hypothesis so far.

The experimental data from previous studies suggested a role of 2-oxoglutarate for the physiological coupling of the partner bacteria in phototrophic consortia (Glaeser and Overmann, 2003). Free-living green sulfur bacteria are known to excrete significant amounts of 2-oxoglutarate (Sirevag and Ormerod, 1970; Czeczuga and Gradski, 1972) which represents a potential carbon substrate of the central bacterium. Whereas intact consortia have been shown to incorporate 2-oxoglutarate (Glaeser and Overmann, 2003), *Chl. chlorochromatii* does not use this compound (Vogl *et al.*, 2006) suggesting that 2-oxoglutarate is indeed utilised by the central bacterium. In the present study, several of the differentially regulated epibiont genes were involved in nitrogen and amino acid metabolism. Of particular interest for a potential metabolic coupling of the partner bacteria in "*C. aggregatum*" involving amino acids is the amino acid binding protein of a branched-chain amino acid ABC transporter which is only expressed in symbiotic state. Interestingly, the

expression of the binding protein in pure epibiont cultures could also be provoked by the addition of consortia culture supernatant indicating that its expression is controlled by some sort of signal exchange with the central bacterium. Analyses of membrane proteins from the close archaeal associations of *Ignicoccus hospitalis* and *Nanoarchaeum equitans* suggested that the binding protein of an ABC transporter of *I. hospitalis* is involved in metabolite exchange between both partners (Burghardt *et al.*, 2009). In other bacteria, a similar ABC-transporter is typically upregulated under nitrogen limiting conditions (Nikodinovic-Runic *et al.*, 2009). Together with the significantly increased expression of the nitrogenase and nitrogen regulatory genes in the symbiotic epibiont cells, the expression pattern of the amino acid binding protein hence suggests that epibionts experience nitrogen limitation in the symbiotic state which maybe caused by an increased synthesis and transfer of branched chain amino acids. A transfer of amino acids has recently been suggested for other interspecific bacterial interactions like the associations of *Rhizobium* sp. with the nitrogen fixing filamentous cyanobacterium *Anabaena* sp. (Behrens *et al.*, 2008) and for the archaeal consortia of *I. hospitalis* and *N. equitans* (Jahn *et al.*, 2008). Interestingly, branched chain amino acids also play an important role in root nodule symbioses (Prell *et al.*, 2009) in which the plant host controls the development and persistence of the *Rhizobium* bacteroids by the transfer of branched chain amino acid via ABC transporters. Recent HPLC analyses revealed that the branched-chain amino acids dominated significantly in the supernatants of exponential and stationary epibiont pure cultures (Wenter, R., Grossart, H.P., Overmann, J., unpublished data). These amino acids are non-polar and thus diffuse through the membranes of Gram-negative bacteria at significant rates (Krämer, 1994). Recent $H^{14}CO_3^-$-radiolabeling studies indicated a rapid transfer of low molecular photosynthesis products from the epibionts to the central bacterium as well as that a specific exchange of branched-chain amino acids and 2-oxoglutarate is indeed likely to occur in phototrophic consortia (Müller, J., Eisenreich, W., Overmann, J., unpublished data). Therefore the precise role of carbon compounds and amino acids in the interaction of the partner bacteria will be elucidated in further detail by an isotopolomics approach aiming to identify the involved products and to determine the kinetics of their exchange.

Since the amino acid binding protein of a branched-chain amino acid ABC transporter could be cross-linked, it may actually be located at the contact sites of the epibionts to the central bacterium and hence be directly involved in a physiological interaction between the bacterial partners in "*C. aggregatum*". As jet, the only protein involved in interspecific interaction in bacterial consortia was detected from the archaeal association constituting of *I. hospitalis* and *N. equitans*, in which the dominating outer membrane protein of *I. hospitalis* forms transmembrane pores likely to be involved in metabolite and signal exchange (Burghardt *et al.*, 2007). Based on these initial results, additional proteins are likely to be involved in cell-cell interaction within phototrophic consortia.

Therefore a future systematic analysis of membrane proteins will we performed using gel-free LC-MS/MS. Additionally, the identification of membrane protein complexes responsible for interspecific interaction could be achieved by employing a size-exclusion chromatography approach according to Burghardt *et al.*, 2007 or blue-native gel electrophoresis (Reisinger and Eichacker, 2008).

In contrast to the results of the present study of "*C. aggregatum*", nothing is known about the possible physiological interactions of the potentially sulfate-reducing single cells of the MMPs, because they could not yet be cultivated and no genome sequence is available so far. Together with the myxobacteria, the only intraspecific multicellular interaction among the deltaproteobacteria, which in contrast to the MMP becomes multicellular by aggregation rather than by growth and cell division (Kaiser, 2001), is represented by the sulfate reducer *Desulfosarcina variabilis* known to form dense cell aggregates (Widdel, 1980). Interestingly, *Desulfosarcina* species represent the closest cultivated relatives of the MMP, but likewise almost nothing is known about their cell-cell interactions. Based on the chemotaxis assays performed in this work, MMPs possibly utilise acetate and propionate for their growth. Together with other substrates these fatty acids will be used in future cultivation approaches since a stable laboratory culture is a prerequisite to learn more about the potential physiological coupling between the single MMP cells. In parallel, separation of single MMPs by micromanipulation followed by multiple displacement amplification of its genomic DNA or alternatively flow cytometric sorting of dense environmental populations and subsequent large insert library construction or direct genome sequencing will be employed in a future approach to study the molecular basis and development of prokaryotic multicellularity in this fascinating microorganism.

Conclusions

Taken together, substantial progress has been made during the course of this study with regard to the understanding of functional implications important for the development of bacterial multicellularity based on the characterisation of molecular determinants underlying the interspecific symbiosis in the phototrophic consortium "*C. aggregatum*". The results of the combined comparative genomic, transcriptomic and proteomic approach yielded a considerable inventory of genes and gene products with significant relevance for the mutualistic interactions in phototrophic consortia and other bacterial symbioses, identified promising targets for future functional studies and generated novel hypotheses with respect to the molecular mechanisms that form the basis of morphologically and physiologically united higher entities of different bacterial species. In particular, unique epibiont genes exhibiting striking similarities to virulence factors of typical proteobacterial pathogens were found to have possible implications for cell-cell adhesion between

bacteria. This finding provides a good indication that genetic modules have been transferred laterally to nonrelated bacteria and are employed in multicellular interactions between prokaryotes. The small fraction of niche-specific genes in the epibiont genome indicates a preadaptation of their free-living relatives to interspecific interactions. Lateral gene transfer as well as the preadaptation of the ancestor thus hypothetically represent two key features in the evolution of bacterial multicellularity in phototrophic consortia and presumably also in MMPs. On the other hand, altered gene expression was detected mainly with regard to basic cellular functions involving central metabolic pathways and housekeeping functions rather than symbiosis genes which may represent one decisive step towards the formation of bacteria-bacteria symbioses as well as host-pathogen interactions. The exchange of branched chain amino acids was found to have a possible regulatory function in the metabolic coupling of bacterial partners in "*C. aggregatum*" and most likely plays an important physiological role in other bacterial symbioses. A protein possibly involved in the intercellular sorting of chlorosomes represents a suitable target for the elucidation of this cell biological process which is uncommon in prokaryotes. Therefore specific symbiosis dependent regulation and signal exchange between the partner bacteria in phototrophic consortia must occur in the coupling of metabolic processes as well as in the intracellular sorting of their photosynthetic antenna. Compared to the results obtained in the studies on "*C. aggregatum*", nearly nothing is known so far about the intraspecific interactions of the single cells of multicellular magnetotactic prokaryotes. The MMPs most likely gain its energy from sulfate reduction and use fatty acids as substrates as suggested by the results of the present study which thereby provides the basis for a successful cultivation as the premise to establish the MMP as a model system for the discovery of the molecular mechanisms that underlie the development of intraspecific multicellularity in prokaryotes. Ultrastructural analyses of the MMP from the North Sea demonstrated a highly ordered arrangement of the magnetosome chains which provides the explanation for the magnetic optimisation of these multicellular bacterial associations indicating a high degree of intraspecific interaction between the single cells via yet unknown signal exchanges. The optimisation of the magnetotactic response and the need to efficiently transfer the magnetic polarity during division of the multicellular aggregates may have been a major selective force for the formation of multicellularity among magnetotactic bacteria. Due to its complexity in structure and behaviour, the MMP represents the most highly developed type of intraspecific multicellularity known so far among prokaryotes. Future studies on the MMP therefore have the potential to elucidate the origin and evolution as well as the molecular determinants of multicellular life on the prokaryotic level.

References

Abreu, F., Silva, K.T., Martins, J.L. and Lins, U. (2006) Cell viability in multicellular magnetotactic prokaryotes. *Int Microbiol* 9: 267-271

Behrens, S., Lösekann, T., Pett-Ridge, J., Weber, P.K., Wing-On, N., Stevenson, B.S., Hutcheon, I.D., Relman, D.A., Spormann, A.M. (2008) Linking microbial phylogeny to metabolic activity at the single-cell level by using enhanced element labeling-catalyzed reporter deposition fluorescence *in situ* hybridization (EL-FISH) and NanoSIMS. *Appl Environ Microbiol* 74: 3143-3150

Bowe, F., Lipps, C.J., Tsolis, R.M., Groisman, E., Heffron, F., Kusters, J.G. (1998) At least four percent of the *Salmonella typhimurium* genome is required for fatal infection of mice. *Infect Immun* 66: 3372-3377

Burghardt, T., Näther, D.J., Junglas, B., Huber, H., Rachel, R. (2007) The dominating outer membrane protein of the hyperthermophilic Archaeum *Ignicoccus hospitalis*: a novel pore-forming complex. *Mol Microbiol* 63: 166-176

Burghardt, T., Junglas, B., Siedler, F., Wirth, R., Huber, H., Rachel, R. (2009) The interaction of *Nanoarchaeum equitans* with *Ignicoccus hospitalis*: proteins in the contact site between two cells. *Biochem Soc Trans* 37: 127-132

Czeczuga, B., Gradski, F. (1972) Relationship between extracellular and cellular production in the sulphuric green bacterium *Chlorobium limicola* Nds. compared to primary production of phytoplankton. *Hydrobiologia* 42: 85-95

Dalisay, D.S., Webb, J.S., Scheffel, A., Svenson, C., James, S., Holmström, C., Egan, S., Kjelleberg, S. (2006) A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga *Ulva australis*. *Microbiology* 152: 2875-2883

Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M. *et al*. (2002) The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc Natl Acad Sci USA* 99: 9509-9514

El-Azami-El-Idrissi, M., Bauches, C., Loucka, J., Osicka, R., Sebo, P., Ladant, D., and Leclerc, C. (2003) Interaction of *Bordetella pertussis* adenylate cyclase with CD11b/CD18. *J Biol Chem* 278: 38514-38521

Fröstl, J., and Overmann, J. (1998) Physiology and tactic response of "*Chlorochromatium aggregatum*". *Arch Microbiol* 169: 129-135

Glaeser, J. and Overmann J. (2003) The significance of organic carbon compounds for *in situ* metabolism and chemotaxis of phototrophic consortia. *Environ Microbiol* 5: 1053-1063

Glaeser, J., and Overmann, J. (2004) Biogeography, evolution, and diversity of epibionts in phototrophic consortia. *Appl Environ Microbiol* 70: 4821-4830

Gottig, N., Garavaglia, BS., Garafalo, CG., Orellano EG., Ottado, J. (2009) A filamentous hemagglutinin-like protein of *Xanthomonas axonopodis* pv. *citri*, the phytopathogen responsible for citrus canker, is involved in bacterial virulence. *PLoS One* 4: e4358

Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K.O., Rachel, R., Huber, H. (2008) *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. *J Bacteriol* 190: 1743-1750

Kaiser, D. (2001) Building a multicellular organism. *Annu Rev Genet* 35: 103-123

Kanzler, B.E.M., Pfannes, K.R., Vogl, K., Overmann, J. (2005) Molecular characterization of the non-photosynthetic partner bacterium in the consortium *"Chlorochromatium aggregatum"*. *Appl Environ Microbiol* 71: 7434-7441

Kajava, A.V., Cheng, N., Cleaver, R., Kessel, M., Simon, M.N., Willery, E., Jacob-Dubuisson, F., Locht, C., and Steven, A.C. (2001) Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. *Mol Microbiol* 42: 279-292

Keim, C.N., Abreu, F., Lins, U., Lins de Barros, H. and Farina, M. (2004) Cell organisation and ultrastructure of a magnetotactic multicellular organism. *J Struct Biol* 145: 254-262

Krämer, R. (1994) Systems and mechanisms of amino acid uptake and excretion in prokaryotes. *Arch Microbiol* 162: 1-13

Nikodinovic-Runic, J., Martin, L.B., Babu, R., Blau, W., O'Connor, K.E. (2009) Characterization of melanin-overproducing transposon mutants of *Pseudomonas putida* F6. *FEMS Microbiol Lett* 298: 174-183

Pfannes, K.R., Vogl, K., Overmann J. (2007) Heterotrophic symbionts of phototrophic consortia: members of a novel diverse cluster of betaproteobacteria characterized by tandem rrn operon structure. *Environ Microbiol* 9: 2782-2794

Pfennig, N. (1980) Synthrophic mixed cultures and symbiotic consortia with phototrophic bacteria: a review. In: *Anaerobes and anaerobic infections.* Gottschalk, G., Pfennig, N., Werner, G. (eds.). Stuttgart: Fischer, pp. 127-131

Prell, J., White, J.P., Bourdes, A., Bunnewell, S., Bongaerts, R.J., Poole, P.S. (2009) Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc Natl Acad Sci USA* 106: 12477-12482

Overmann, J., Tuschak, C., Sass, H., Fröstl, J.M. (1998) The ecological niche of the consortium "*Pelochromatium roseum*" *Arch Microbiol* 169: 120-128

Reisinger, V., Eichacker, L.A. (2006) Analysis of membrane protein complexes by blue native PAGE. *Proteomics* Suppl 2: 6-15

Relman, D.A., Domenighini, M., Tuomanan, E., Rappuoli, R., and Falkow, S. (1989) Filamentous hemagglutinin of *Bordetella pertussis*: Nucleotide sequence and crucial role in adherence. *Proc Natl Acad Sci USA* 86: 2637-2641

Risso, C., Sun, J., Zhuang, K., Mahadevan, R., DeBoy, R., Ismail, W., Shrivastava, S., Huot, H., Kothari, S., Daugherty, S., Bui, O., Schilling, C.H., Lovley, D.R., Methé, B.A. (2009) Genome-scale comparison and constraint-based metabolic reconstruction of the facultative anaerobic Fe(III)-reducer Rhodoferax ferrireducens. *BMC Genomics* 10: 447

Sirevag, R., Ormerod, J.G. (1970) Carbon dioxide-fixation in photosynthetic green sulfur bacteria. *Science* 169: 186-188

Van der Veen, S., Hain, T., Wouters, J.A., Hossain, H., de Vos, W.M., Abee, T., Chakraborty, T., Wells-Bennik, M.H. (2007) The heat-shock response of Listeria monocytogenes comprises genes involved in heat shock, cell division, cell wall synthesis, and the SOS response. *Microbiology* 153: 3593-3607

Takeichi, M. (1990) Cadherins: a molecular family important in selective cell-cell adhesion. *Annu Rev Biochem* 59: 237-52

Vogl, K., Glaeser, J., Pfannes, K.R., Wanner, G., and Overmann, J. (2006) *Chlorobium chlorochromatii* sp. nov., a symbiotic green sulfur bacterium isolated from the phototrophic consortium "*Chlorochromatium aggregatum*". *Arch Microbiol* 185: 363-372

Wanner, G., Vogl, K., Overmann, J. (2008) Ultrastructural characterization of the prokaryotic symbiosis in *"Chlorochromatium aggregatum"*. *J Bacteriol* 190: 3721-3730

Widdel, F. (1980) Anaerober Abbau von Fettsäuren und Benzoesäure durch neu isolierte Arten Sulfat-reduzierender Bakterien. Dissertation. Georg August-Universität zu Göttingen. Lindhorst/Schaumburg-Lippe, Göttingen, Germany, 443 p.

Yousef F., Espinosa-Urgel, M. (2007) In silico analysis of large microbial surface proteins. *Res Microbiol* 158: 545-550

Zhang, P., Khursigara, C.M., Hartnell, L.M., Subramaniam, S. (2007) Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy. *Proc Natl Acad Sci USA* 104: 3777-3781

Danksagung

An erster Stelle danke ich Prof. Dr. Jörg Overmann für die Bereitstellung des höchst interessanten Themas, das mich sehr für die Wissenschaft begeistert hat, sowie für die erstklassige Betreuung während meiner Doktorarbeit.

Großer Dank gilt der gesamten AGO für das freundschaftliche Arbeitsklima und die immerwährende Hilfsbereitschaft. Ein Extra-Dankeschön möchte ich meinen Kollegen Mareike Jogler, Johannes Müller und Ovidiu Rücker sowie meiner langjährigen Kommilitonin Annemarie Hütz für die super Zusammenarbeit im Labor und die vielen schönen gemeinsamen Unternehmungen aussprechen.

Ganz besonders bedanke ich mich bei Prof. Dr. Dirk Schüler für die Unterstützung bei meiner Arbeit mit den MMPs am MPI Bremen und später in München sowie für die große Hilfsbereitschaft bei allen wissenschaftlichen Fragen.

Außerdem bedanke ich mich bei Dr. Christian Jogler für den gemeinsamen Spaß an der Forschung während der Probenahmen in der Nordsee und den spannenden Mikromanipulator-Experimenten sowie für die zahlreichen Diskussionen über Mikrobiologie und darüber hinaus.

Ganz besonders herzlich danke ich meiner Freundin für ihr liebevolles Verständnis und ihre Einfühlsamkeit in den Höhen und Tiefen der Promotion.

Abschließend gilt mein größter Dank meinen geliebten Eltern Angela und Peter sowie meiner Großmutter Auguste Wenter. Ohne Euren Glauben und Rückhalt während meiner gesamten Ausbildung wäre das nun Erreichte nicht möglich gewesen.

Curriculum vitae

Personal data

Name	Roland Wenter
Date of birth	25.10.1979
Place of birth	Munich, Germany

University education

April 2006-February 2010	Ph.D. student in the group of Prof. Dr. Jörg Overmann at the Ludwig-Maximilians-University Munich; Research on the molecular basis of prokaryotic multicellularity
Oktober 2000-February 2006	Study of biology at the Ludwigs-Maximilians University Munich (Diploma); Main focus on microbiology and biochemistry

School education

September 1990-Juni 1999	Abitur at the Theresien-Gymnasium Munich; Advanced courses in Biology and English
September 1986-Juli 1990	Elementary school in Munich

Research experience

April 2006-December 2009	Graduate research assistant in the group of Prof. Dr. Jörg Overmann; PhD thesis
March 2005-February 2006	Undergraduate research assistant the group of Prof. Dr. Jörg Overmann; Diploma thesis
September 2003-February 2004	Undergraduate research assistant in the group of Prof. Dr. Reinhold Herrmann; Cytological analyses of plant chromosomes

Teaching experience

April 2006-June 2009	Supervision of research interns and Diploma students
March/April 2004	Teaching assistant in the lab course 'Microbiology for biology students'
July/August 2003	Teaching assistant in the lab course 'Microbiology and genetics for students of biochemistry and chemistry'
February/March 2003	Teaching assistant in the lab course 'Microbiology for medicine students'

Awards, grants and scholarships

August 2008	EES^{LMU} Travel Grant for the ISME meeting in Australia
Juli 2008	Travel Grant of the Munich University Society for the ISME meeting in Australia
Oktober 2001-March 2006	LMU Mentoring Program
March 1999	"Environmental Protection Award" of "Jugend forscht"

List of publications

Wenter, R., Hütz, K.A., Dibbern, D., Plöscher, M., Li, T., Reisinger, V., Plöscher, M., Eichacker, L.A., Eddie, B., Hanson, T., Bryant, D.A., Overmann, J. (2009) Expression-based identification of genetic determinants of the bacterial symbiosis "*Chlorochromatium aggregatum*". *Environ Microbiol* (published online ahead of print on April 01, 2010; doi:10.1111/j.1462-2920.2010.02206.x)

Wenter, R., Wanner, G., Schüler, D., Overmann, J. (2009) Ultrastructure, tactic behavior and potential for sulfate reduction of a novel multicellular magnetotactic prokaryote from North Sea sediments. *Environ Microbiol* 11: 1493-1505

Vogl, K., **Wenter, R.,** Dressen, M., Schlickenrieder, M., Plöscher, M., Eichacker, L.A., Overmann, J. (2008) Identification and analysis of four candidate symbiosis genes from ´*Chlorochromatium aggregatum*´, a highly developed bacterial symbiosis. *Environ Microbiol* 10: 2842-2856

Conference presentations

Wenter, R., Wanner, G., Schüler, D., Overmann, J. Ecophysiology and cell-cell interaction of a novel multicellular magnetotactic prokaryote from North Sea sediments. Association of General and Applied Microbiology (VAAM) Workshop "Symbiotic Interactions", November 19-20, 2009, Munich, Germany **(poster)**

Müller, J., **Wenter, R.**, Dibbern, D., Reisinger, V., Plöscher, M., Eichacker, L.A., Overmann, J. Molecular and physiological characterization of the close bacterial symbiosis in phototrophic consortia. Association of General and Applied Microbiology (VAAM) Workshop "Symbiotic Interactions", November 19-20, 2009, Munich, Germany **(poster)**

Overmann, J., **Wenter, R.**, Vogl, K., Müller, J. Molecular basis of the bacterial symbiosis in phototrophic consortia. 13th International Symposium on Phototrophic Prokaryotes, August 9-14, 2009, Montreal, Canada **(oral presentation)**

Wenter, R., Wanner, G., Schüler, D., Overmann, J. Ultrastructure, tactic behavior and potential for sulfate reduction of a novel multicellular magnetotactic prokaryote from North Sea sediments. Annual Conference of the Association of General and Applied Microbiology (VAAM), March 8-11, 2009, Bochum, Germany **(oral presentation)**

Wenter, R., Vogl, K., Wanner, G., Overmann, J. Molecular and ultrastructural characterization of the close bacterial symbiosis in phototrophic consortia. The 12th International Symposium on Microbial Ecology (ISME-12), August 17-22, 2008, Cairns, Australia **(oral presentation)**

Wenter, R., Müller, J., Hütz, K.A., Overmann, J. Molecular basis of symbiosis in phototrophic consortia. Annual Conference of the Association of General and Applied Microbiology (VAAM), March 9-11, 2008, Frankfurt, Germany **(poster)**

Wenter, R., Hütz, K.A., Müller, J., Pfannes, K.R., Vogl, K., Zikeli, K., Overmann, J. Molecular basis of symbiosis in phototrophic consortia. Annual Conference of the Association of General and Applied Microbiology (VAAM), April 1-4, 2007, Osnabrück, Germany **(oral presentation)**