Molecular Ecology of the NOR5/OM60 Group of *Gammaproteobacteria*

Dissertation zur Erlangung des akademischen Grades

Doktors der Naturwissenschaften

(Dr. rer. nat.)

Dem Fachbereich Biologie/Chemie der

Universität Bremen vorgelegt von

YAN Shi

Bremen
Februar 2009

Die vorliegende Arbeit wurde in der Zeit von April 2006 bis Februar 2009 am Max-Planck-Institut für marine Mikrobiologie in Bremen angefertigt.

1. Gutachter: Prof. Dr. Rudolf Amann
2. Gutachter: Prof. Dr. Ulrich Fischer

Tag des Promotionskolloquiums: 26. März 2009

Table of Contents

Summary

In this thesis, a newly discovered gammaproteobacterial group – the NOR5/OM60 clade, which includes the novel aerobic anoxygenic phototrophs (AAnPs), was studied in the aspects of phylogeny, biogeography and physiology using both molecular and culturing methods.

By means of 16S rRNA phylogenetic analysis, NOR5/OM60 group was defined as a monophyletic clade inside the class *Gammaproteobacteria*. More than 500 16S rRNA sequences of this clade were retrieved from public databases. Further studies classified the sequences into 13 subclades. My studies on the biogeography of NOR5/OM60 clade showed a cosmopolitan distribution. They were found in all oceans and at the coasts of all the continents. NOR5/OM60 seems to be most abundant in marine coastal water and sediment. However, they have also been encountered in surface water of open ocean, deep-sea sediment, freshwater, saline lakes and soil.

The abundance of members of the NOR5/OM60 clade in various marine sites was determined by fluorescence *in situ* hybridization (FISH) with a newly designed and optimized probe set. In the ocean, the common relative abundance ranges from <1% in the open ocean to about 10% in the North Sea coastal water. The existence in coastal sediment and freshwater was also proved by FISH.

More than 30 strains of the NOR5/OM60 clades had been isolated as pure cultures in several studies. Of these, the genomes of five strains (KT71, RAp1red, Ivo14, HTCC2080 and HTCC2148) were fully sequenced. In this thesis, these five genomes were compared to each other and to a closely related strain, HTCC2143 from the BD1-7 group. The photosynthesis (PS) superoperon, the key genes of 3-hydroxypropionate cycle for carbon fixation, and the *sox* operon for sulfur compound oxidation were found in four strains: KT71, RAp1red, Ivo14 and HTCC2080. Genomic comparison indicated that, like KT71, many NOR5/OM60 members may also be AAnPs and potentially be able to use light as energy source.

Zusammenfassung

In dieser Arbeit wurde eine kürzlich entdeckte Gruppe, die NOR5/OM60-Gruppe der *Gammaproteobacteria*, einige von dessen Mitglieder den aeroben anoxygenen und phototrophen Bakterien (AAnP) gehöhren, in Phylogenie, Biogeographie und Physiologie anhand molekularer und kultivierungsbasierender Methoden untersucht.

Anhand der phylogenetischen Analyse der 16S rRNA Sequenzen wurde die NOR5/OM60-Gruppe als eine Monophylie in der Klasse *Gammaproteobacteria* begrenzt. Mehr als 500 16S rRNA Sequenzen aus dieser Gruppe wurden in öffentlichen Datenbanken gefunden. Weitere Studien klassifizierten die Sequenzen in 13 Untergruppen. Biogeographische Untersuchungen zeigten eine weltweite Verteilung der NOR5/OM60-Gruppe, welche in allen Ozeanen und an Küsten aller Kontinenten vorkommt. Diese Organismen scheinen am häufigsten in marinen Küstengewässer und Sediment aufzutreten, jedoch kommen sich auch in Hochseeoberflächengewässer, Tiefseesedimenten, Süßwasser, Salzseen und im Boden vor.

Die Abundanz der NOR5/OM60 Populationen in verschiedenen marinen Habitaten wurde mit der Fluoreszenz *in situ* Hybridisierung (FISH) mit einem neu entwickelten und optimierten Sondenset untersucht. In marinen Proben reichte die relative Abundanz von <1% in der Hochsee bis über 10% im Nordsee-Küstenwasser. Auch in Küstensedimenten und im Süßwasser wurde das Vorkommen durch FISH bestätigt.

In mehreren Studien wurden bislang mehr als 30 Stämme der NOR5/OM60-Gruppe in Reinkulturen isoliert. von fünf dieser Stämme wurden die Genome (KT71, RAp1red, Ivo14, HTCC2080 und HTCC2148) vollständig sequenziert. In dieser Arbeit wurden die fünf Genome untereinander und mit einem engverwandten Stamm, HTCC2143 aus der BD1-7-Gruppe, verglichen. Das Superoperon der Photosynthese (PS), die Schlüsselgene des 3-Hydroxypropionat-Zyklus für CO_2-Assimilation, und das *sox*-Operon zu Oxidation von Schwefelverbindungen wurden in vier Stämmen – KT71, RAp1red, Ivo14 und HTCC2080 gefunden. Die Ergebnisse der vergleichenden Genomik legen nahe, dass wie KT71 viele NOR5/OM60-Mitglieder potentielle AAnP sind und auch Licht als Energiequelle verwenden können.

摘要

NOR5/OM60 演化支是一類新發現的 γ-變形菌(*Gammaproteobacteria*)類群，包含有好氧不產氧光合細菌(AAnP)。本論文利用分子生物學及純培養方法研究了它們的系統演化分類、地理分佈及生理生態學功能。

通過 16S rRNA 演化分析，NOR5/OM60 類群被詳細界定爲 γ-變形菌綱中的一個單系群。從公共數據庫中發現有 500 多個 16S rRNA 序列屬於這一分支。它們被劃分爲 13 個亞支。本研究顯示 NOR5/OM60 成員廣泛分佈於全球，包括各大洋中，及各大洲的沿岸。在海洋中，該支在沿岸的水體及沉積物中最爲豐富。此外，它們在於大洋表層、深海沉積物、淡水、咸水湖及土壤中也有存在。

利用幾個重新設計和優化的探針組，NOR5/OM60 在各海洋樣品中的豐度通過熒光原位雜交(FISH)得到測定。在海洋中，其相對豐度的範圍從大洋中的低於 1%到北海沿岸水體中的 10%左右。在沿岸沉積物及淡水中，NOR5/OM60 的存在也通過 FISH 得到確認。

從多項研究中，已有 30 餘株 NOR5/OM60 被分離爲純株。其中有 5 株（KT71、RAp1red、Ivo14、HTCC2080 及 HTCC2148）的基因組被完全測序。本研究比較了此 5 株及一近緣菌株——BD1-7 類群的 HTCC2143 株的基因組。在 4 個基因組（KT71、RAp1red、Ivo14 及 HTCC2080）中發現了若干生理代謝途徑的關鍵基因，包括光合作用(PS)的超操縱組，3-羥基丙酸固碳途徑，以及還原性硫化合物氧化途徑的 *sox* 操縱組。上述基因組學發現顯示，如同 KT71，NOR5/OM60 的很多成員也屬於 AAnP 且有可能利用光作爲能量來源。

List of Abbreviations

AAnP	aerobic anoxygenic phototroph
ATP	adenosine triphosphate
BChl	bacteriochlorophyll
CARD-FISH	fluorescence *in situ* hybridization with catalyzed reporter deposition
Chl	chlorophyll
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
FISH	fluorescence *in situ* hybridization
GOS	global ocean survey
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
LHC	light harvesting complex
PCR	polymerase chain reaction
PFGE	pulse field gel electrophoresis
PS	photosynthesis
rRNA	ribosomal ribonucleic acid
TCA cycle	tricarboxylic cycle (= citric acid cycle)

1 Introduction

1.1 Marine bacteria that utilize light

The ocean covers 71% of the earth surface, and provides roughly half of the earth's total net primary productivity, which was estimated as around 10^{11} tons of carbon per year (Field et al., 1998). Photoautotrophs, which convert carbon dioxide into organic matters with light as energy source, are either oxygenic or anoxygenic. Oxygenic photoautotrophs, including plants, eukaryotic algae and cyanobacteria, use water as electron donor, and release oxygen gas. The anoxygenic phototrophs, like sulfur green bacteria, non-sulfur green bacteria, purple bacteria and heliobacteria (Table 1), their anoxygenic photosynthesis uses other compounds than water as electron donor, typically reduced sulfur compounds, and the process occurs strictly under anaerobic conditions (Madigan and Martinko, 2006). In the marine environment, the newly recognized aerobic anoxygenic phototrophs (AAnPs) are of great interest.

AAnPs use light as an energy source, and carry out photosynthesis only under aerobic conditions. They presumably use the ATP produced to support their metabolism and growth. They appear to have an important role in marine carbon cycling (Kolber et al., 2000; Kolber et al., 2001). Although they are able to carry out photosynthesis, light is not necessary for their growth. They grow best when organic substrates are available, and strong light inhibits their growth.

AAnPs seem to be highly abundant in the oceans. Recent studies based on infrared microscopy showed abundances of 4.5 ± 2.4% with maximum of 13.5% in coastal waters, while in oceanic water frequency was lower at 1.5 ± 1.3% (Sieracki et al., 2006; Jiao et al., 2007; Yutin et al., 2007). Bacteriochlorophyll *a* (BChl *a*) is the only photosynthetic pigment in reaction center of AAnP. The ratio of marine BChl *a* to phytoplankton chlorophyll *a* can be up to 10% (Kolber et al., 2001; Jiao et al., 2003). Therefore, AAnPs consist of a considerable part of the marine biomass, as well as part of global phototrophy and productivity, and probably also an important chain in global carbon cycle.

Table 1 List of prokaryotes that carry out photosynthesis (Madigan and Martinko, 2006)

Classification	Type	Pigment in reaction center	Electron donor	Carbon source
cyanobacteria	oxygenic photosynthesis	Chl *a* or *b*	H_2O	CO_2 by Calvin Cycle
purple sulfur and non-sulfur bacteria	anaerobic anoxygenic photosynthesis	BChl *a* or *b*	reduced sulfur compounds, organic compounds, H_2	CO_2 by Calvin Cycle, organic compounds
green sulfur bacteria (*Chlorobi*)		BChl *a*	reduced sulfur compounds	CO_2 by reverse TCA cycle, organic compounds
green non-sulfur bacteria (*Chloroflexi*)		BChl *a*	organic compounds, H_2, H_2S	CO_2 by hydroxy-propionate pathway, organic compounds
heliobacteria		BChl *g*	organic compounds	organic compounds
AAnPs	aerobic anoxygenic photosynthesis	BChl *a*	organic compounds	organic compounds

Besides the phototrophs, many prokaryotes in the ocean's photic zone contain another pigment, rhodopsin. It converts light energy into proton gradient which can be used to produce chemical energy in the form of ATP (Patzelt et al., 2002). Some extreme halophilic archaea, e.g. *Halobacterium*, use membrane-bound bacteriorhodopsin to convert light into ATP. In the open ocean, its homolog – proteorhodopsin, was widely found in *Bacteria*. This is also presumably the basis of energy source for these bacteria. The most famous examples for proteorhodopsin-containing bacteria are the SAR86 group (Béjà et al., 2001) and *Pelagibacter* (Giovannoni et al., 2005).

1.2 The NOR5/OM60 group

1.2.1 Phylogeny of *Bacteria*

Microorganisms are living creatures that usually cannot be seen by naked eyes. Due to the small size and limited variance of morphology, the microorganisms were unlike higher plants and animals, not properly classified based on their phenotype. A

stable classification had to await the application of molecular markers for phylogenetic studies. Only since the mid 1970s, 16S/18S rRNA had been used as molecular marker to study the phylogeny of organisms. In 1990, Woese et al. proposed three domains of life by means of comparative sequence analysis of 16S rRNA (Woese et al., 1990): *Bacteria*, *Archaea* and *Eukarya*. In this way, all the cellular organisms, including the micro-organisms, could be arranged in on tree of life. Thirty years later, we still only know a few big branches and several leaves of the large tree. The cultivation-independent retrieval of molecular information indicates that there are still many more "leaves" and "twigs" that are not yet isolated in pure culture. So far, only few studies have been focusing on the small phylogenetic groups (order to genus level), especially those from the environment. For the phylogenetic groups of *Bacteria* without pathogenic importance, our knowledge about their phylogeny, physiology and functions is still quite poor when comparing to what we know for plants and animals.

The largest phylogenetic branch (phylum) of the domain *Bacteria* is *Proteobacteria*, which is divided into five classes: *Alpha*-, *Beta*-, *Gamma*-, *Delta*- and *Epsilonproteobacteria*. In the aquatic environments, *Alphaproteobacteria* inhabit mainly marine water, *Betaproteobacteria* mainly in fresh water, while *Gammaproteobacteria* dwell in both marine and fresh water. The most famous model organism of *Bacteria – Escherichia coli* is also a member of *Gammaproteobacteria*.

1.2.2 History, phylogeny and abundance of the NOR5/OM60 clade

In 1997, two almost full-length 16S rRNA sequences, clones OM60 (U70696) and OM241 (U70702), were retrieved from a marine coastal site off North Carolina, US (Rappé et al., 1997). They were affiliated into the class *Gammaproteobacteria*, far from other known sequences. By now, hundreds of sequences have been recognized to be closely related to these clones, including Japanese deep-sea clones BD1-7 and BD2-7 (Li et al., 1999) and a German North Sea clone KTc1119 (Eilers et al., 2000). In 1999, strain KT71 (AY007676) was isolated from marine surface water at the "Kabeltonne" station off the island of Helgoland, North Sea (Eilers et al., 2001), and the binominal name "*Congregibacter litoralis*" has been suggested for this isolate (Fuchs et al., 2007). The strain KT71 was found to be closely related to the above mentioned clones, and they were named as NOR5 clade inside the class *Gammaproteobacteria*. Several further strains

isolated by a novel high throughput culturing method, including HTCC2080, were also found to be related to strain KT71, and they were placed in the OM60/OM241 clade (Connon and Giovannoni, 2002), later referred to as OM60 clade (Cho and Giovannoni, 2004). More than 30 strains have been isolated and shown to be affiliated to this group. Since these clade names are redundant, they were renamed as the NOR5/OM60 clade (Fuchs et al., 2007). Many more related sequences are currently available in the public databases.

Based on comparative 16S rRNA sequence analysis, the NOR5/OM60 clade is most closely related to the genera *Endobugula*, *Microbulbifer*, *Teredinibacter* (all *Alteromonadales*), *Cellvibrio* (*Pseudomonadales*) and several other groups of oligotrophic marine *Gammaproteobacteria* including the clades BD1-7, KI89A, OM182 and SAR92 (Figure 1) (Cho and Giovannoni, 2004).

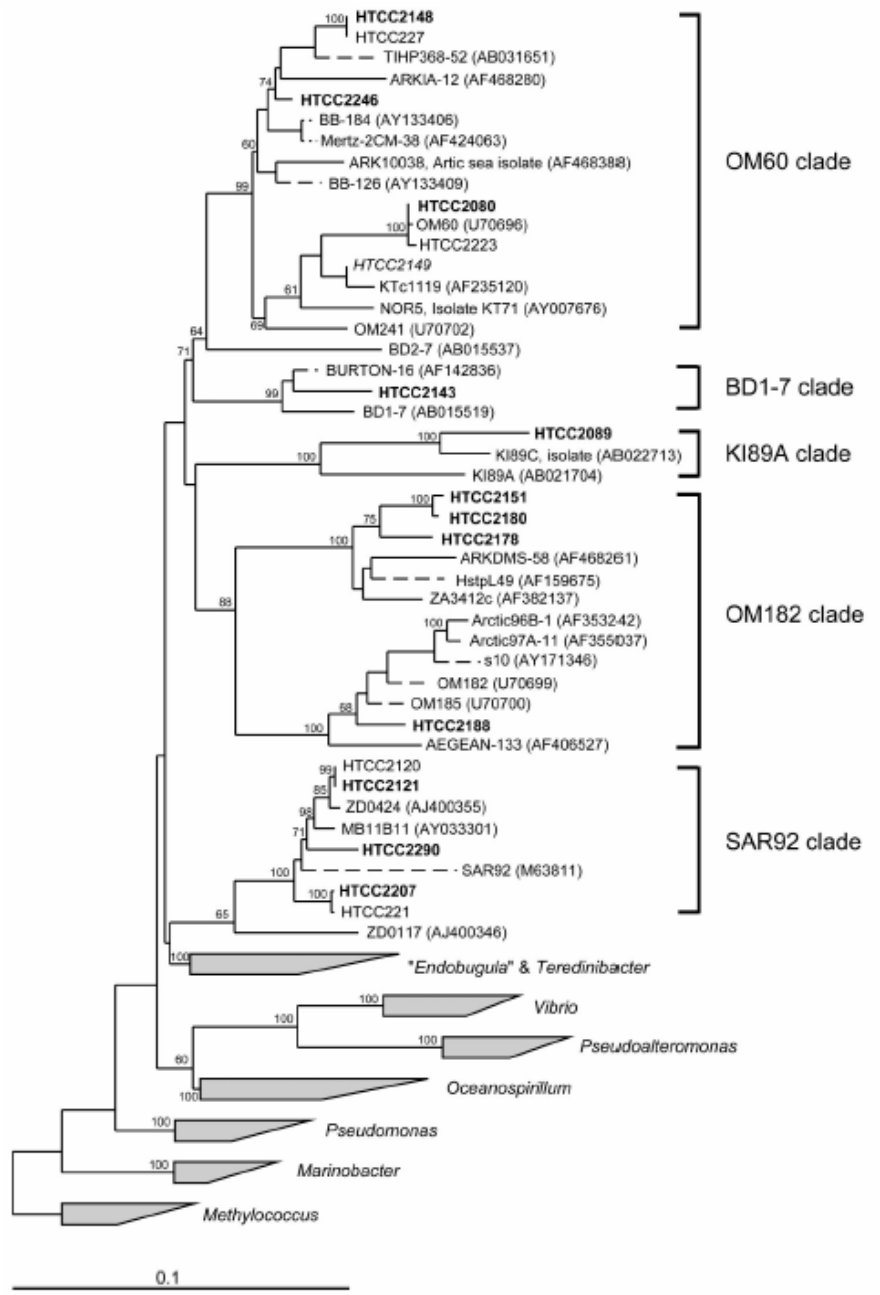

Figure 1 Neighbor-joining tree of 16S rRNA showing relationship of OM60 group and its relatives, from the work of Cho and Giovannoni (Cho and Giovannoni, 2004).

The probe NOR5-730 (Eilers et al., 2001) which targets on most members of the NOR5/OM60 group has been used to count the abundance of this group in several marine environments. In surface water off Helgoland, Germany, where the clone KTc1119 and strain KT71 were found, the yearly NOR5/OM60 percentage by DAPI counts varies

between 0.2 and 2.8% (Keller, 2003). However, in the surface water of the North Sea, 8% (Eilers et al., 2001) and even 11% (Pernthaler and Pernthaler, 2005) of DAPI counts were also reported. Therefore the NOR5/OM60 clade is considered to be an abundant group in marine coastal water.

1.2.3 Strain KT71

As the first isolated strain of NOR5/OM60 clade (Eilers et al., 2001), strain KT71 was investigated in detail. Its physiology was characterized, and its genome was fully sequenced. KT71 cells are highly pleiomorphic. The shape can be coccoid to long, bended rods, with the length ranging normally from 1 – 3 μm (Figure 2).

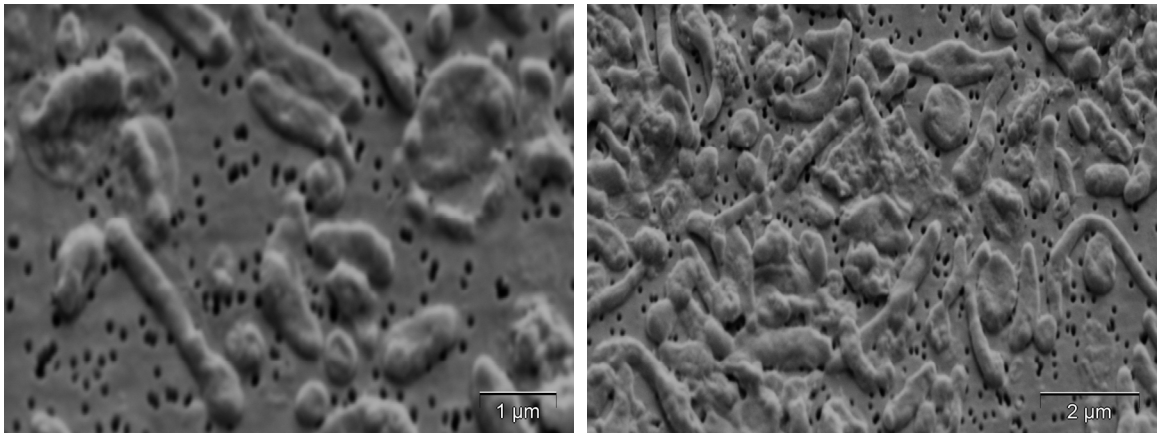

Figure 2 Electron microscopic picture of KT71 culture in SYPG medium (taken by J. Wulf)

In the nearly complete genome of KT71, a complete photosynthesis (PS) superoperon was discovered from its genome (Fuchs et al., 2007). The superoperon contains *bch* (bacteriochlorophyll synthesis), *puf* (light-harvesting complex I (LHC I) and reaction center) and *crt* (carotenoid synthesis) genes.

The strain can be grown in either complex SYPG medium or oligotrophic MPM-m medium. From cell extracts of KT71, a bacteriochlorophyll *a* (BChl *a*) peak and a carotenoid spirilloxanthin-like peak have been detected using high performance liquid chromatography (HPLC) analysis. Significant amounts of BChl *a* were only detected when growing with light on the oligotrophic MPM-m medium for an extended time. Physical tests indicated that KT71 could not grow autotrophically. The growth of KT71 seemed to be faster with light (Fuchs et al., 2007).

KT71 is an obligatory aerobic organism with preference for low-oxygen niches. In agar-shake culture, the cells actively move to a surface with about 10% oxygen saturation (30 µM O_2). In liquid culture, KT71 forms large flocs. The gene cluster *sox*H-RCDXYZA-B for sulfur compounds oxidization was found in the genome of KT71, however, supplementation of media with thiosulfate or elemental sulfur did not significantly promote growth of KT71 using different carbon sources.

1.2.4 Members of NOR5/OM60 group are AAnPs

For a long time all cultured representatives of marine AAnPs were belonging to the classes *Alpha-* and *Betaproteobacteria*. Most famous genera are *Erythrobacter*, *Roseobacter* and *Sphingomonas*.

Strain KT71 is the first discovered AAnP member of *Gammaproteobacteria* based on the existence of photosynthesis (PS) superoperon and the expression of BChl *a* (Fuchs et al., 2007). The *puf*L and *puf*M genes, which are part of the PS superoperon and encoding reaction center, are found in several NOR5/OM60 strains, including HTCC2080 and several North Sea strains (Cho et al., 2007). Closely related *puf*LM sequences as well as fosmid clones containing nearly identical PS gene arrangement have been reported, indicated that an essential part of AAnPs might be from the NOR5/OM60 clade (Béjà et al., 2002; Yutin and Béjà, 2005).

1.3 Sampling stations

The samples for this study were taken from various regions of the world. This includes the German North Sea coastal region, coastal regions in China, a transect in the North Atlantic, Namibian upwelling regions, and many other locations. The large variety of sampling ensured both a general global view as well as series of sampling for correlation of the NOR5/OM60 group with environmental parameters.

1.3.1 Yangtze River estuary

On September 6 – 8th, 2006, a small cruise was made at the estuary of Yangtze River (Figure 3). This region is highly influenced by the incoming fresh water of the Yangtze River. The salinity was between 20 – 32 psu. The water is rich in nutrients. There are also frequently algal blooms. This region is under frequent studies for the physical and microbiological characteristics (Uzuka et al., 1996; Jiao et al., 2002).

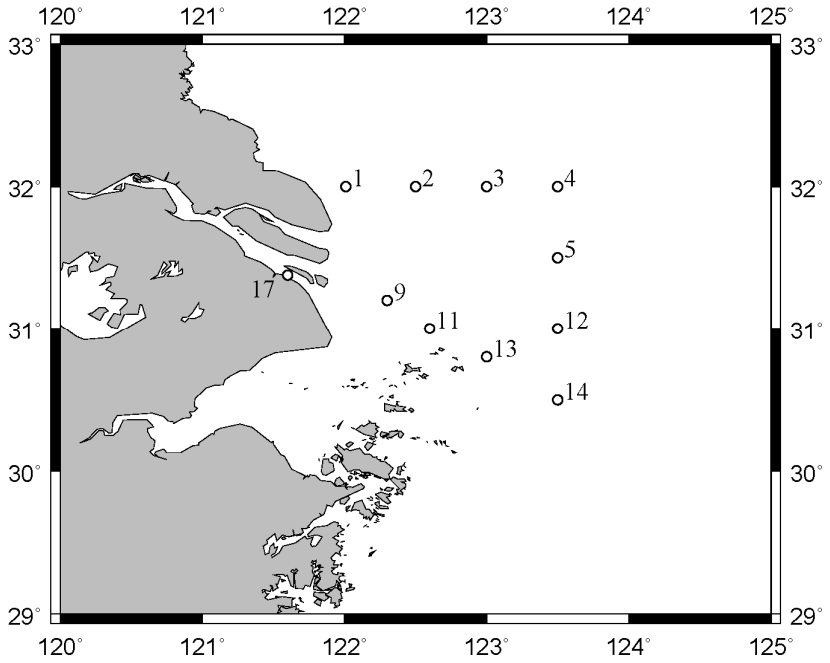

Figure 3 Sampling stations at the estuary of Yangtze river.

1.3.2 Namibian upwelling region

The Namibian coast is influenced by the Benguela Current, which causes an upwelling of cold and oxygen-depleted water from the deeper sea. The samples examined in this thesis were taken on March 22 – 23th, 2003, along 23°S near Walvis Bay, from the coast into the Atlantic Ocean (14.4°E – 12.0°E, Figure 4). Surface water samples (10 m) from all 13 stations and three depth profiles were collected.

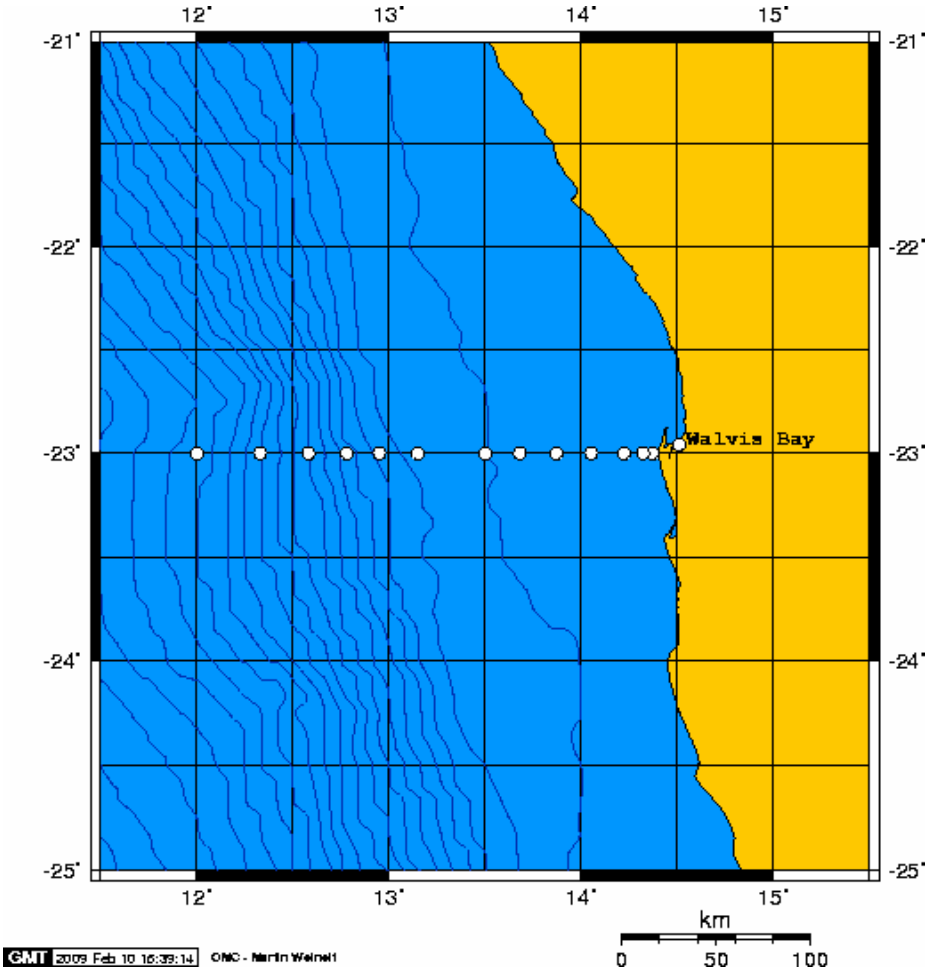

Figure 4 The sampling stations along 23°S across the Benguela Current at the Namibian coastal region. The interval of bathymetry contours is 200 m.

1.3.3 Vision cruise

The Vision cruise was conducted in the period of September 20th – October 3rd, 2006. Sampling was done along the transect 30°W, from Iceland to south of the Azores Islands (Figure 5). The cruise passed the cold Eastern Greenland Current, the warm North Atlantic drift and the warm oligotrophic Gyre.

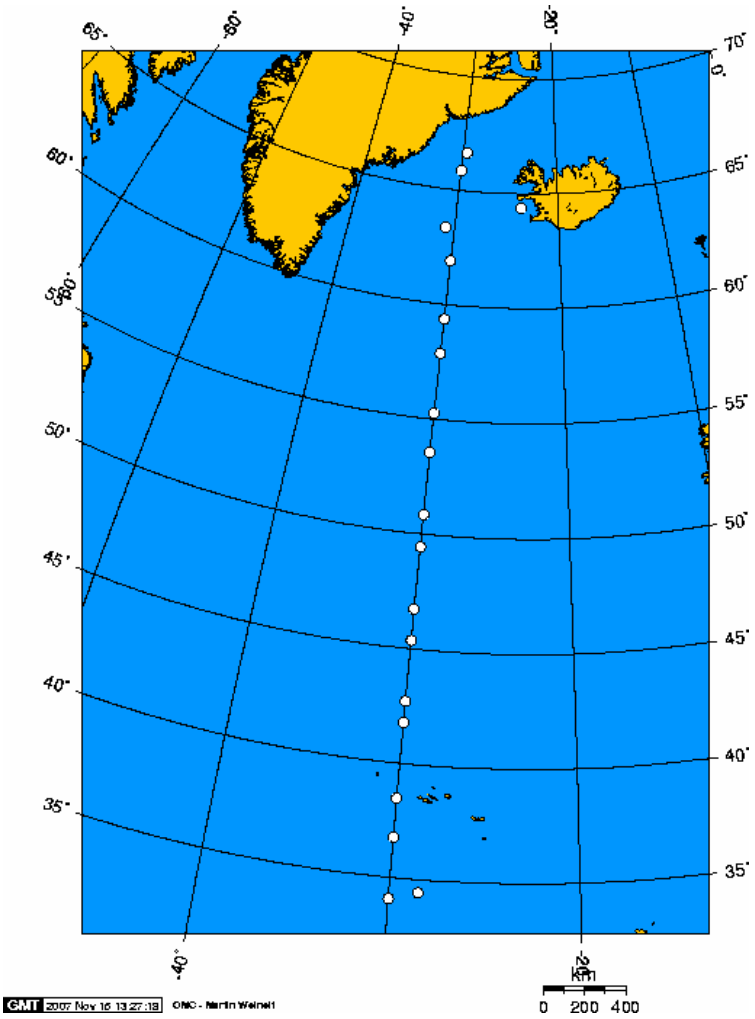

Figure 5 Sampling stations of the Vision cruise in North Atlantic, along 30°W, from Greenland down to the south of Azores Islands.

1.3.4 German Bight

Samples were taken from 1 m depth at station "Kabeltonne", Helgoland (54.18°N, 7.90°E), German Bight, on seven separate days from May to July of 2007, and at Cuxhaven in July 2007 (Figure 6).

1.3.5 North Sea sediment

The German North Sea coastal region is characterized by a huge intertidal sand flat, the Wadden Sea. This is a place with high nutrient concentrations and high bacterial mineralization rate in both sandy and muddy sediments (de Beer et al., 2005). Sediment samples were taken at Janssand (53.72°N, 7.68°E, Figure 6), in March and August of 2007.

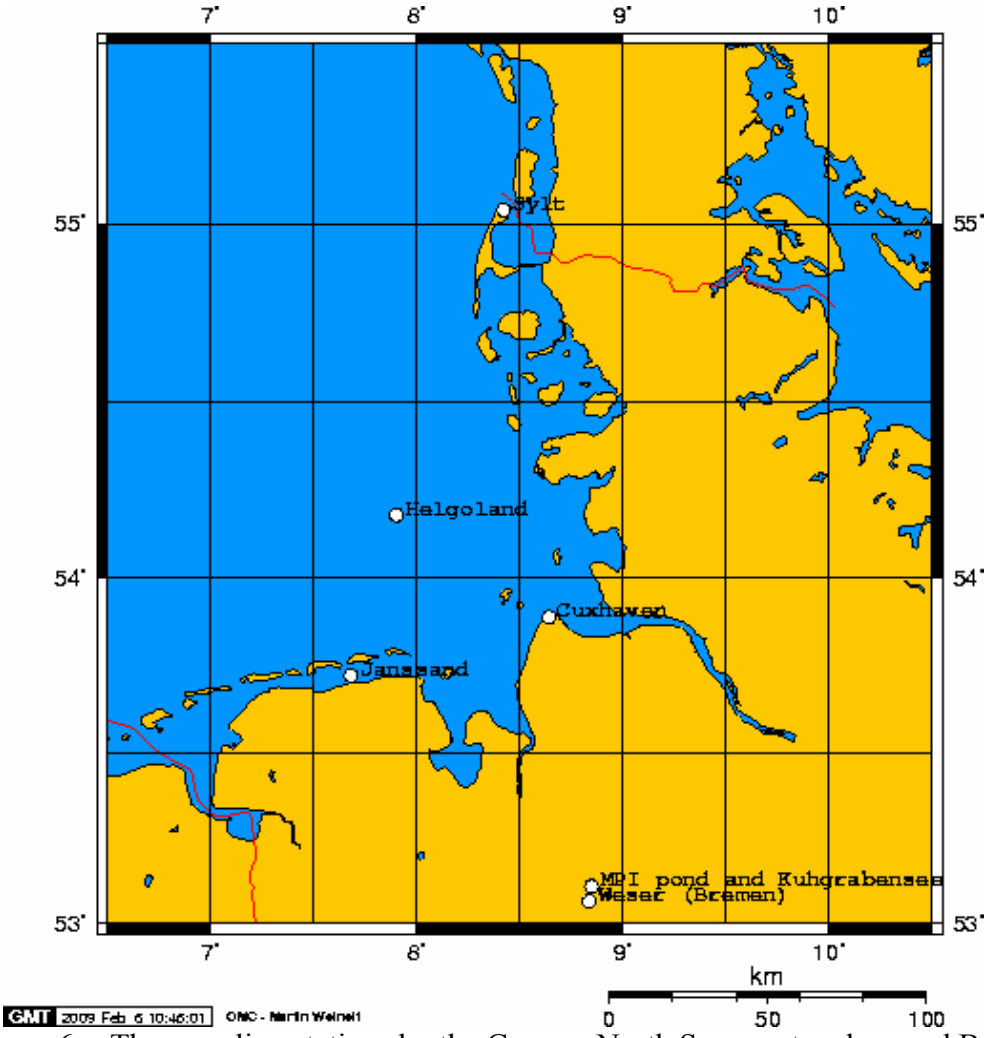

Figure 6 The sampling stations by the German North Sea coast and around Bremen.

1.3.6 Other sampling stations

A summary of all sampling stations is given in Table 2. Surface water samples from Xiamen, China were taken in September 2006 at the Xiamen ferry port, and in July 2007 from a sandy coastal area near Xiamen University. Other marine water samples were obtained from Southampton dock water, UK, coastal water near Barcelona, Spain. We also checked fresh water samples from the river Weser and freshwater ponds in Bremen for comparison. Other sediment samples were taken from intertidal sandy surface sediment from Sylt at the German North Sea coast.

1.4 Methodological aspects

The methods of this study can be categorized in two major parts: culture-dependent and culture-independent. The latter also provide data on microorganisms that

have not yet been cultured. The culture-independent methods have been used to learn more about the diversity of microorganisms (Amann et al., 1995). They include the 16S rRNA approach and metagenomics studies.

1.4.1 The 16S rRNA approach

The full cycle 16S rRNA approach was described by Amann and colleagues (Amann et al., 1995). It includes DNA extraction, PCR-based 16S rRNA amplification, cloning and sequencing followed by comparative sequence analysis. In a second phase, probe design and hybridization are carried out to identify, localize and quantify the existence of the targeted group of microorganisms.

The number of 16S rRNA sequence data is always increasing. By SILVA release 96 from Oct. 2008 (Pruesse et al., 2007), 324,342 full sequences and 756,668 partial sequences are included. These can be handled by the software package ARB (Ludwig et al., 2004). Phylogenies can be calculated, and probes can be designed for a particular clade. Probes are most often oligonucleotides that are labeled, e.g. with fluorescent dyes. After hybridization and washing, target cells can be visualized under the microscope. This method is fluorescence *in situ* hybridization (FISH). The probes then bind to the complementary stretch of 16S rRNA. The specificity for hybridization can be adjusted by hybridization temperature and formamide concentration to find the point with greatest difference between the melting curves of full match and mismatch sequences (Pernthaler et al., 2001), which can be determined by hybridization on strains, environmental samples or clones (Amann et al., 1995; Schramm et al., 2002). Competitor oligonucleotides are often designed to improve the probe specificity for differentiating full-match targets and sequences with one or a few mismatches (Manz et al., 1992). When a signal is too weak, caused e.g. by targeting a poorly accessible site, unlabeled helper nucleotides can be designed to open the secondary structure of rRNA, thereby enhancing the signal intensity (Fuchs et al., 1998; Fuchs et al., 2000). FISH with catalyzed reporter deposition (CARD-FISH) was developed to improve the sensitivity. The probes are labeled with horseradish peroxidase (HRP), and an amplified fluorescent signal created by the catalyzed deposition of fluorescently labeled tyramide. This allowed the reliable detection of micro-organisms with low rRNA content (Pernthaler et al., 2002).

1.4.2 Genomics and metagenomics

Genomics and metagenomics provide vast amount of data of DNA sequences from bacterial strains or environmental samples. They enable us to postulate functions of a strain or environmental microbes from knowledge deduced from other organisms. The method is based on the theory that closely related homologous genes might function similarly in most cases, especially the orthologs, which are supposed to be separated together with speciation of the organisms.

1.4.3 Culturing

In spite of the rapid development of culture-independent methods and genomics, enrichment, isolation and physiological tests remain to be the most reliable source for knowledge on the biochemistry and physiology of microorganisms. Not every gene in the genome is transcribed and expressed. And homologous genes found by genomics may not have the same functions. Especially there are also novel genes that no homologs have ever been found in any know strains. In these cases, physiological tests on the isolated strains are necessary.

The culture-dependent and -independent methods expand our knowledge in depth and breadth, respectively, in order to understand the vast diversity of marine micro-organisms.

1.5 Aims of this study

The goal of this thesis is to better understand the NOR5/OM60 group in several aspects:

1) to investigate its phylogeny, and based on that, to get a clear definition of the clade, and divide it into subclades;

2) to know where they occur, in what abundance, and whether there are any relationship between the habitat and the phylogeny;

3) to improve the knowledge on the physiology of the group;

4) to get insights in the living strategies of members in the NOR5/OM60 clade, and their functions to the environment;

5) to understand their evolution, how and where their ancestor inhabited, and what functions the different subgroups have gained.

2 Results and Discussion

2.1 Phylogeny

Based on an extensive comparison of trees obtained with various programs for phylogenetic reconstruction on more than 150 almost full-length NOR5/OM60 and closely related 16S rRNA gene sequences, a new consensus tree (Figure 8) was constructed. With all treeing methods, NOR5/OM60 clade was monophyletic within *Gammaproteobacteria*. In contrast to earlier trees based on less sequences (Figure 1, Figure 7) (Cho and Giovannoni, 2004; Fuchs et al., 2007) the current reconstruction of the NOR5/OM60 clade now includes the strain KT71 and the clones OM60 and OM241, a cluster of freshwater clones, and BD2-7, a clone retrieved from the deep-sea as well. Another deep-sea sequence BD1-7 was still excluded from the NOR5/OM60 clade. Sequence identities within NOR5/OM60 are typically >92%, while identities to outgroup sequences are usually below 92%, although exceptions do occur.

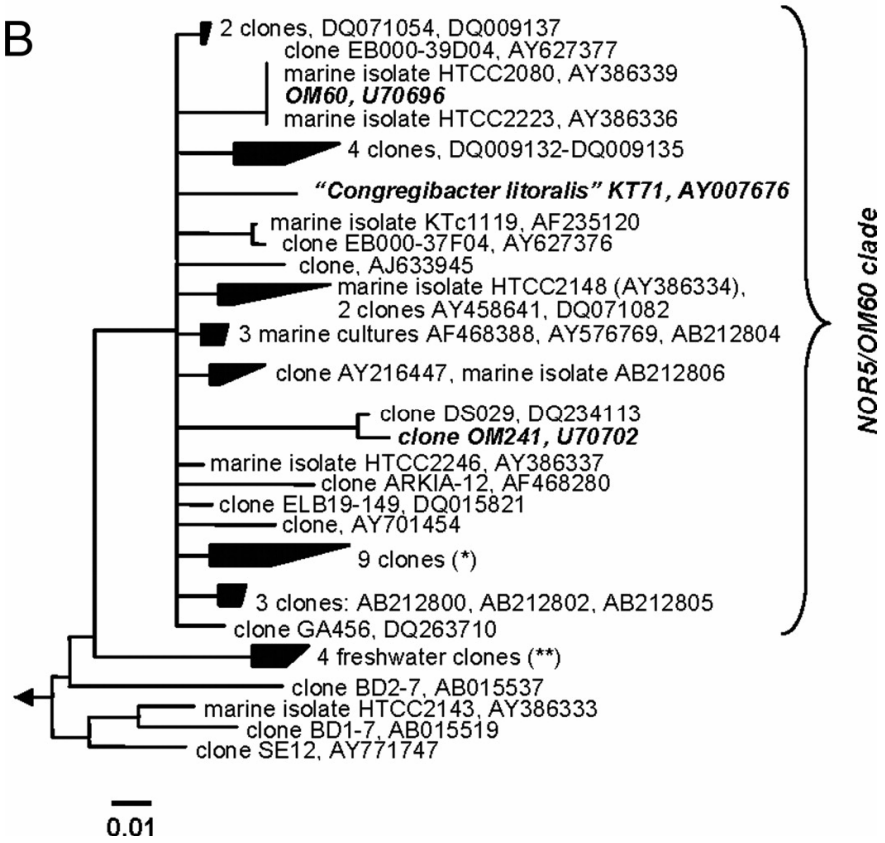

Figure 7 Consensus tree of the NOR5/OM60 clade reconstructed with 86 almost-full-length sequences (>1,350 nt), by Fuchs et al. (Fuchs et al., 2007).

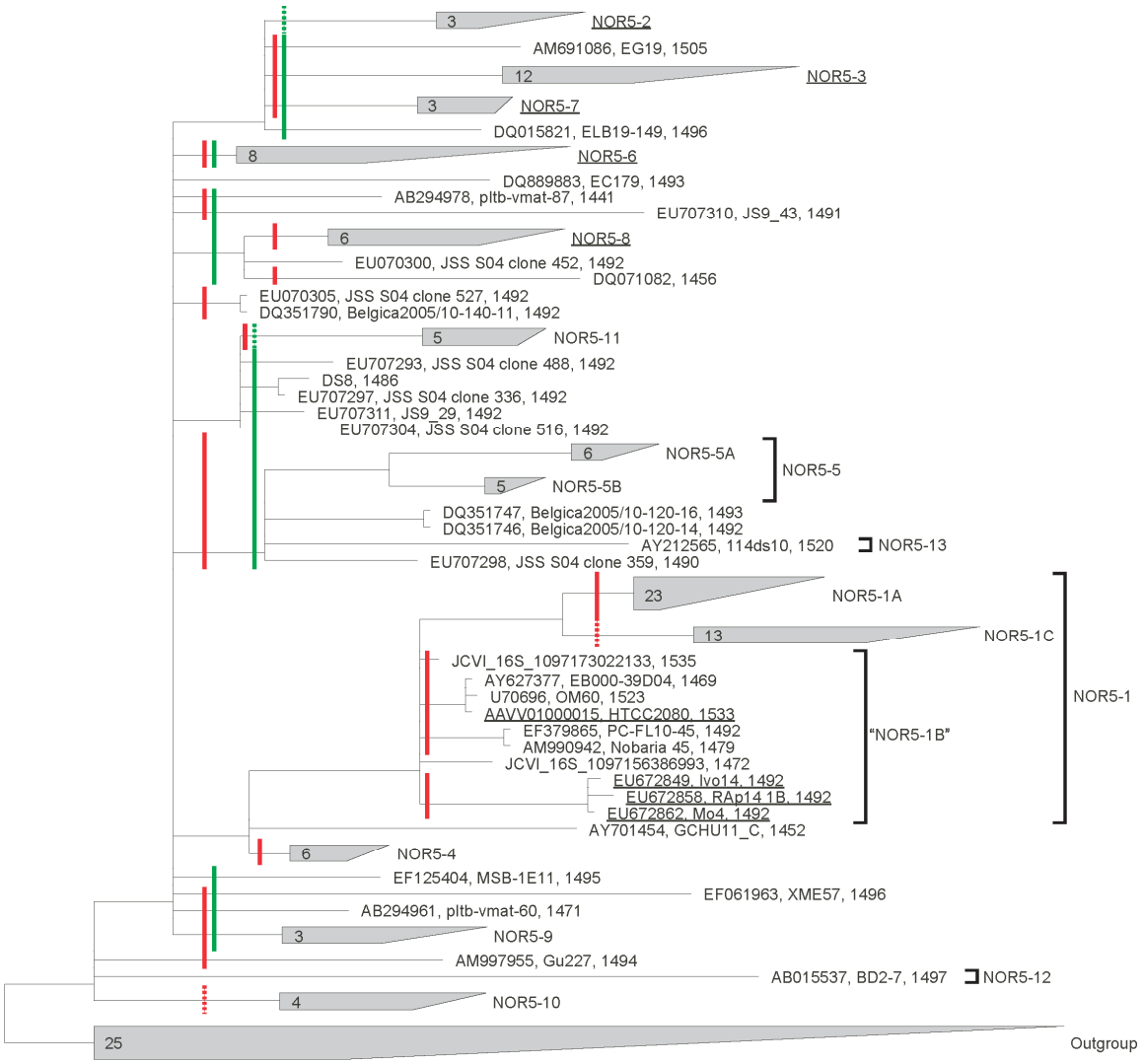

Figure 8 Consensus tree reconstructed in this study, based on almost full length (>1400 nt) 16S rRNA sequences of members of the NOR5/OM60 clade. Underlined names are cultured isolates and subclades that include cultured isolates. The red and green bars on the left of branches show the clades that can be targeted by probe NOR5-730 and NOR5-1238, respectively, and the dashed lines for partly targeted subclades.

According to the 16S rRNA phylogeny, 13 subclades can be identified. The composition of each subclade is quite stable, but the relationship between the subclades varies much, depending on the algorithms and filters used for reconstruction. The largest subclade NOR5-1 covers more than one third of all the available full-length sequences as well as many partial sequences. Two main monophyletic subgroups, NOR5-1A and NOR5-1C can be recognized. The rest, including strain HTCC2080 and North Sea strain

Ivo14, are called "NOR5-1B", which might be paraphyletic. Another stable subclade NOR5-4, including the clone OM241, is the sister group of NOR5-1 in most of the trees. Subclade NOR5-3 includes the "*Congregibacter litoralis*" KT71 as well as 17 other NOR5/OM60 strains which all have been recently isolated from the oxic layer marine surface sediment of the German island Sylt. Subclades NOR5-2 and NOR5-7 were close to NOR5-3 in most phylogenetic reconstructions. Subclades NOR5-5, NOR5-6, NOR5-8, NOR5-9 and NOR5-11 together contain one fifth of all the NOR5/OM60 sequences. Subgroups NOR5-10 and NOR5-12 were deeply branching in most of the trees. They are dominated by sequences obtained from the deep-sea. BD2-7 is the only full-length sequence of NOR5-12, and shows low identity (usually <92%) with other NOR5/OM60 sequences. Clone 114ds10 (AY212565) (Simpson et al., 2004) is the only full sequence in the terrestrial subclade NOR5-13 which includes also 13 partial sequences recovered from freshwater, freshwater sediment or soil. About 30% of NOR5/OM60 sequences, most of which are partial sequences, cannot yet be grouped into any of the subclades mentioned above.

The most cultured strains are concentrated in a few subclades, e.g. the NOR5-2/3/7 branch, NOR5-1B and NOR5-8. However few strains were isolated from the most abundant pelagic subclades, like NOR5-1A/C or NOR5-4 (see Unit 1, SI Table 4).

2.2 Biogeography

2.2.1 Worldwide existence

All the locations on the world, where NOR5/OM60 existence was identified by now, are labeled on Figure 10. This includes identifications by isolation, 16S rRNA gene libraries, metagenomic studies, and by fluorescence *in situ* hybridization (FISH). Comparing to the knowledge at the beginning of this study (Figure 9), there are much more locations at which NOR5/OM60 existence was identified, especially from the East Asian region, open ocean of North Atlantic and Pacific Ocean, and deep-sea.

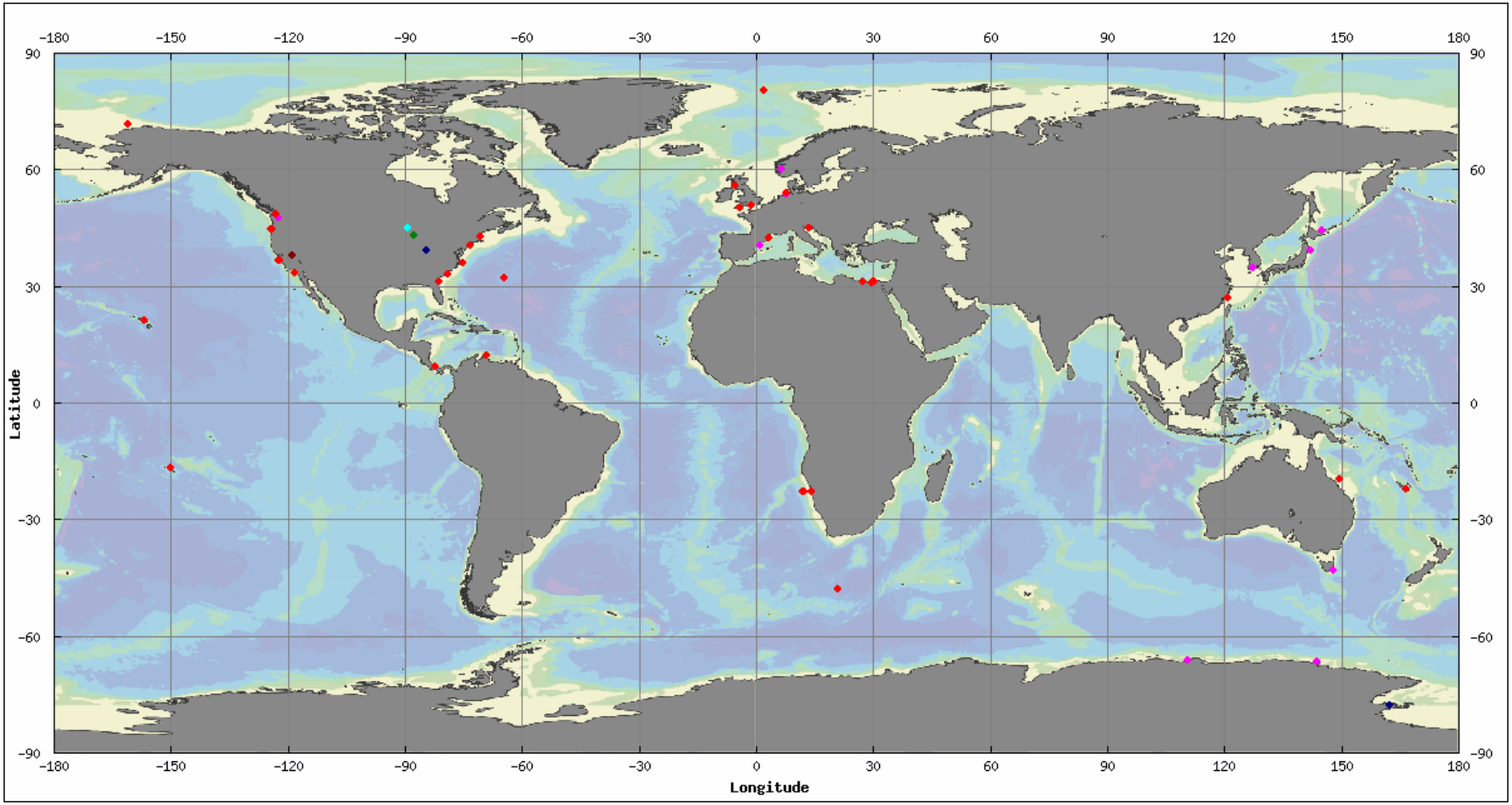

Figure 9 The world map labeling identification of the NOR5/OM60 clade and relatives by 2005. The colors of the points stand for: marine water or undefined marine samples – red; marine sediment – pink; saline lake – brown; fresh water – dark blue; fresh sediment – green; soil – light blue.

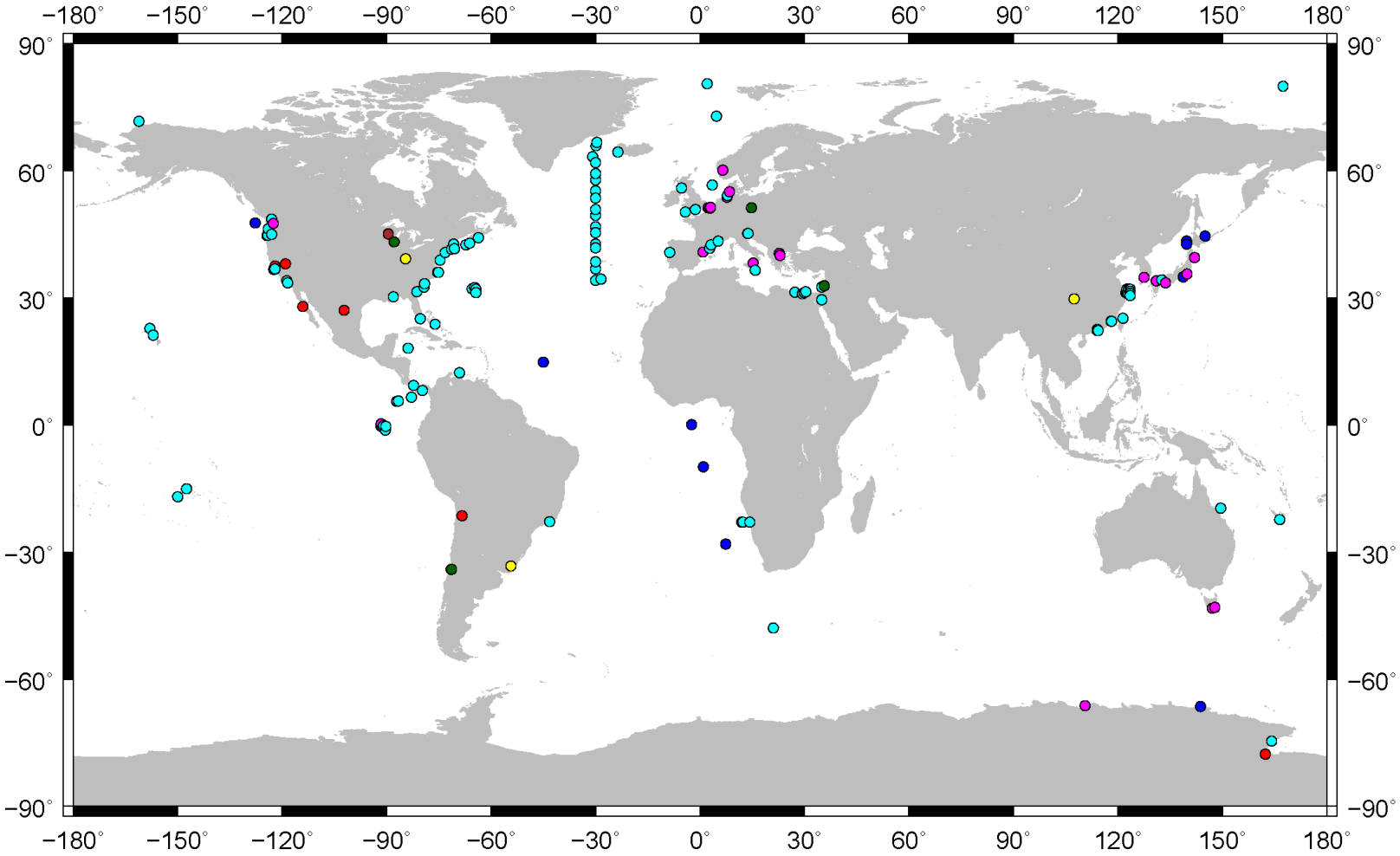

Figure 10 Biogeography of the NOR5/OM60 clade. Sequence-, isolation-, and FISH-based identifications of NOR5/OM60 were marked on the world map. Colors refer to the habitat from where the sample was retrieved: marine water or other marine habitats – cyan; marine coastal sediment – magenta; hypersaline – red; soil – brown; fresh water – yellow; fresh sediment – dark green; deep-sea – dark blue.

The NOR5/OM60 clade is cosmopolitan in the marine realm. Identifications have been reported from almost all oceans and at many coastal sites. American, European and East Asian coasts are particularly well covered with 16S rRNA gene libraries. There seems to be no latitudinal preference, since NOR5/OM60 clones were reported from mangrove (Liang et al., 2007; Liao et al., 2007), coral reefs (Frias-Lopez et al., 2002; Koren and Rosenberg, 2006; Barneah et al., 2007) as well as sea-ice habitat (Brinkmeyer et al., 2003).

NOR5/OM60 sequences were also reported in deep-sea sediments sampled near Antarctica (Bowman and McCuaig, 2003), Japan (Li et al., 1999; Inagaki et al., 2003; Arakawa et al., 2006) as well as in the northeast Pacific (Huber et al., 2006) and the Atlantic (Schauer, unpublished). Additional reports on NOR5/OM60 sequences come from environments with different salinity: freshwater rivers (Sekiguchi et al., 2002; Simpson et al., 2004), a rice paddy (DQ830363), freshwater sediments (MacGregor et al., 2001; Wobus et al., 2003), activated sludge (Klein et al., 2007), soil (Liles et al., 2003; Hartmann and Widmer, 2006), while also from hypersaline environments (Glatz et al., 2006; Ley et al., 2006; Rusch et al., 2007). There is even one sequence from human plasma (clone NF37-A2; AY886614) (Vernon et al., 2002).

The subclades of NOR5/OM60 group show clear preferences towards different environments (see Unit1, Table 3). The large subclades NOR5-1 and NOR5-4 appear nearly exclusively in marine water column. Subclades NOR5-10 and NOR5-12 contain mainly identifications reported from deep-sea samples, and NOR5-13 is a freshwater clade. Sequences of the other NOR5/OM60 subclades were retrieved from marine sediment and water column. However, geographic patterns for the various NOR5/OM60 subclades, either latitudinal or with respect to certain oceanic provinces are difficult to detect.

The GOS dataset (Rusch et al., 2007) contains 3,728 16S rRNA gene sequences of a length of >300 nt. By comparative sequence analysis 30 of these sequences (0.8%) could be unambiguously grouped within the NOR5/OM60 clade. Therein 28 belong to subclade NOR5-1 which is typical of marine surface water, and two belong to its sister subclade NOR5-4. The sequences were found in 21 out of total 44 sampling stations.

2.2.2 Design and optimization of new probes sets

Using the comprehensive set of 16S rRNA sequences of the NOR5/OM60 clade collected in this study, the old probes NOR5-730 and NOR5-130 were re-evaluated. Several new probes were designed for both the NOR5/OM60 group as well as its several subgroups. Subsequently, these probes were optimized with new designed helpers and competitors (Table 2).

The probe NOR5-730 (Eilers et al., 2001) covered 131 of 155 (84%) high-quality, almost full-length 16S rRNA sequences of the NOR5/OM60 clade in the Silva Ref dataset (Version 91) (Pruesse et al., 2007). It is not possible to design a single probe that perfectly matches all the NOR5/OM60 sequences without outgroup hits. The new designed probe NOR5-1238 targets 46% of all high-quality NOR5/OM60 sequences, excluding the two major subclades NOR5-1 and NOR5-4. A combination of the probes NOR5-730 and NOR5-1238 increases the current coverage of the NOR5/OM60 clade to 92%, without any outgroup hits (Figure 8). This combination fails to detect part of NOR5-1C, NOR5-2, NOR5-10, and sequences in the NOR5-12 subclade.

Besides the probe NOR5-130 (Eilers et al., 2001), which targets the NOR5-3 subgroup including several North Sea strains, new probes NOR5-1AC-830, NOR5-1B-840 and NOR5-4-77 were also designed. The NOR5-1 and NOR5-4 are the dominant NOR5/OM60 subclades in the marine water column. It was not possible to design a single probe for the whole NOR5-1 subclade, but probes for the NOR5-1A plus -1C together, and NOR5-1B, could be designed.

The helper oligonucleotides were designed for all above probes in attempt to improve their hybridization efficiency (Fuchs et al., 2000). Helpers are unlabeled oligonucleotides that bind in the vicinity of the probe, thereby opening the secondary structure of the rRNA. The application of two helpers per probe significantly increased the intensity of monolabeled and CARD-FISH signals. By hybridizations of probe mixture on both pure strains and environmental samples, the optimal formamide concentration for hybridization of probes NOR5-730 and NOR-1238 was determined as 50%. The combination of these two probes with helpers NOR5-659h, NOR5-709h, NOR5-1217h and NOR5-1287h was routinely used in this study for detecting cells of NOR5/OM60 group in various environments.

Table 2 Probes, helpers and competitors that were designed and revised in this study. Suffix -h stands for helper, and prefix c- stands for competitor.

Name	Targeted group	Sequence (5′ - 3′)	Target site (16S rRNA *E. Coli* numbering)	% FA (46°C)	Reference
NOR5-730	more than 2/3 of NOR5/OM60 sequences	TCG AGC CAG GAG GCC GCC	730 - 747	50	(Eilers et al., 2001)
NOR5-709h		TTC GCC ACY GGT ATT CCT CCA	709 - 729		This study
NOR5-659h		GAA TTC TAC CTC CCT CTC YCG	659 - 679		This study
NOR5-1238	more than half of NOR5/OM60 sequences, excluding NOR5-1 and -4	CCC TCT GTG CGT TCC ATT	1238-1255	50-55	This study
NOR5-1217h		GTA GCA CGT GTG TAG CCC AGG	1217-1237		This study
NOR5-1287h		ATC CGG ACT ACG AAA CGT TTT	1287-1307		This study
NOR5-130	several NOR5-3 isolates	CCC CAC TAC TGG ATA GAT	130 - 147	40	(Eilers et al., 2001)
KT71-110h		TCC TAC GCG TTA CTC ACC CG	110 - 129		This study
KT71-148h		TCG AGT TTC CCC GAG TTG TC	148 - 167		This study
KT71-210h		CTC CAA TAG CGC GAG GTC CG	210 - 229		This study
NOR5-1AC-830	NOR5-1A and NOR5-1C	TCT CAA GTA CCC CTA CAG	830-847	40-45	This study
NOR5-1AC-809h		CTA GTA GAC ATC GTT TAC GGC	809-829		This study
NOR5-1AC-848h		GCG TTA GCT GCG CTA CAA AGG	848-868		This study
cNOR5-1AC-830		TCT CAA GTA CCC CAA CAG	830-847		This study
NOR5-1B-840	NOR5-1B	GCT ACC AAG GTC TCA AGT	840-857	50	This study
NOR5-1B-819h		ACC CCA ACA GCT AGT AGA CAT	819-839		This study
NOR5-1B-858h		TCT ACT TAT TGC GTT AGC TGC	858-878		This study
cNOR5-1B-840		GCT ACA AAG GTC TCA AGT	840-860		This study
NOR5-4-77	NOR5-4	GTA CTC AGT CCG AAA ACC	77-100	45-55	This study
NOR5-4-101h		GTT ACT CAC CCG TCC GCC GCT	101-121		This study
NOR5-4-53h		TTT CTC GCT CGA CTT GCA TGT	53-73		This study
cNOR5-4-77		GTA CTC AGT CCG AAG ACC	77-100		This study

Note: Y = C or T

The differences between the targets of NOR5-1AC and NOR5-1B are only two bases: 834 and 852 by *E. coli* numbering. Therefore, competitors (cNOR5-1AC-830 and cNOR5-1B-840) had to be used in both cases in order to better separate the two groups. These probes together with the corresponding helpers and competitors were optimized for the hybridization conditions: NOR5-1B-840 with the strain Ivo14, while NOR5-1AC-830 and NOR5-4-77 with environmental samples. Due to limit of time and relatively lower abundances detected, the probes for subclades have not yet been tested for more samples.

2.2.3 Quantification of NOR5/OM60 in the environment

The cells detected by CARD-FISH with the probe mixture NOR5-730/NOR5-1238 in marine plankton and benthos samples were pleomorphic, often coccoid to rod-shaped, sometimes also bended to vibrio shape. The cell length is between 0.5 and 3 μm, and diameter between 0.5 and 1 μm. In plankton samples, we mostly detected single cells, suggesting that they are free-living. However, as described before (Fuchs et al., 2007) we also detected cells that were attached to microaggregates. In sediment samples, cells detected as NOR5/OM60 were also arranged in rosettes, suggesting that they actively grow in this environment.

The optimized NOR5-730/NOR5-1238 probe/helper mixture was used for CARD-FISH-based quantifications in various marine samples. In the brackish to marine Yangtze River estuary (salinities of 22 – 32 psu), we detected between 0 and 2.3% of all DAPI-stained cells. Absolute numbers went up to 1.2×10^5 cells mL^{-1}. Counts in surface waters obtained from an open ocean North Atlantic transect in September 2006 (Vision cruise) were usually between 0.1% and 0.5% (3×10^3 – 1×10^4 cells mL^{-1}). NOR5/OM60 cells were present in all the samples, with no obvious trend from high to low latitude. In a transect in the Namibian coastal upwelling region along 23.0°S, the NOR5/OM60 counts at 10 – 15 m depth decreased with fluctuation from 3.0% (2.0×10^5 cells mL^{-1}) near the coast to 0.5% (1.3×10^4 cells mL^{-1}) in the open ocean. Three depth profiles made at coastal (14.36°E), mid-shelf (13.15°E) and open ocean station (12.00°E) all clearly showed a steep decrease of the NOR5/OM60 abundances with depth. The highest relative abundance of NOR5/OM60 cells encountered in this study was recorded as 1.7 – 6.6% (8.2×10^3 – 1.2×10^5 cells mL^{-1}) in the surface water samples near the North Sea island Helgoland at station “Kabeltonne” (54.18°N 7.90°E).

Counts were also high in sandy intertidal sediments taken at Janssand (53.72°N, 7.68°E), with 2.5 – 4.0% in the top 3 cm of the sediment, and 1.4 – 3.1% at 3 – 12 cm depth. Counts in March 2007 were generally lower than in August. The absolute number of NOR5/OM60 was at the order of 10^7 cells cm^{-3}, in the surface sediments as high as 1.5×10^8 cells cm^{-3}. A preliminary quantification of NOR5/OM60 was done in freshwater samples taken in Bremen, Germany. Abundances were less than 0.1% in the River Weser and two ponds, one freshwater, the other with a salinity of 2 psu. Each time the negative control using NON338 were tested in order to exclude unspecific bindings. The results are summarized in Unit 1, Table 2.

In general, the NOR5/OM60 clade members are more abundant in coastal areas than in open ocean settings, as shown in the Namibian sample. Amongst all the samples in this study, CARD-FISH counts in coastal surface waters (N = 30) showed an average of 2.1 ± 1.5 %, whereas open ocean surface water samples (N = 36) had an average of 0.5 ± 0.4%. The same trend is shown in the GOS dataset (Rusch et al., 2007) in which the frequency of NOR5/OM60 16S rRNA gene sequences was significantly higher in coastal stations (1.4%) than in open ocean stations (0.3%).

The large CARD-FISH dataset also shows a clear preference of members of the NOR5/OM60 clade for the euphotic zone. This is supported by a study from the coastal Pacific Newport Hydroline station (Cho et al., 2007) as well as fosmid libraries constructed from bacterioplankton samples at Aloha Station, Hawaii (DeLong et al., 2006).

There are indications for strong seasonal fluctuation of the NOR5/OM60 abundance in coastal waters. The samples from Xiamen, Barcelona and Helgoland showed high counts of NOR5/OM60 co-occurring with algal blooms. Linear regression analysis revealed correlation between NOR5/OM60 abundance and other parameters. The NOR5/OM60 abundance was highly correlated to turbidity ($R^2 = 0.79$, Namibian transect) and chlorophyll fluorescence ($R^2 = 0.73$, Namibian transect and $R^2 = 0.74$, Yangtze River estuary). Algae are a source of fresh organic material, which in turn could serve as substrate for NOR5/OM60. However, it is still too early to speculate on a specific link to particular algal species.

The NOR5/OM60 group and AAnPs showed some common features in distribution, at least in some regions: they both occur at higher percentage in coastal water than in open ocean; they are more abundant in summer or autumn than in winter or spring; most of them both appear in euphotic zone in the marine water column; and they are positively related to high chlorophyll concentration (Schwalbach and Fuhrman, 2005; Cottrell et al., 2006; Sieracki et al., 2006; Jiao et al., 2007; Yutin et al., 2007).

2.3 North Sea strains of NOR5/OM60

2.3.1 Isolation sources and growth features

In addition to KT71, which was isolated in 1999 by H. Eilers at station Kabeltonne from North Sea surface water, 22 strains were isolated by MarMic Class 2009, in the year 2005. The sources for these isolates were several sediment samples taken in the north of the island Sylt, around the town List and the bay Königshafen (Table 3). The medium used was “PLA-rich”, which based on artificial sea water with complex carbon sources and complemented with cycloheximide and ampicillin in order to isolate marine *Planctomycetes* strains.

Table 3 List of the strains isolated from the North Sea

	Source	Subclade	Color
KT71	Helgoland, surface water	NOR5-3	white
RAp1red	Sylt, aerobic sediment	NOR5-3	dark red
RAp2	Sylt, aerobic sediment	NOR5-3	dark red
RAp5	Sylt, aerobic sediment	NOR5-3	dark red
RAp6	Sylt, aerobic sediment	NOR5-3	dark red
RAp7	Sylt, aerobic sediment	NOR5-3	dark red
RAp8	Sylt, aerobic sediment	NOR5-3	dark red
RAp9	Sylt, aerobic sediment	NOR5-3	dark red
RAp11	Sylt, aerobic sediment	NOR5-3	dark red
RAp13red	Sylt, aerobic sediment	NOR5-1B	pink
RAp14	Sylt, aerobic sediment	NOR5-3 / NOR5-1B [a]	dark red
Ivo10red	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red
Ivo11	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red
Ivo14	Sylt, top oxic layer of muddy sediment	NOR5-1B	pink
Ivo19	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red
Pao12	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red
Mo4	Sylt, oxic layer of sandy sediment	NOR5-1B	pink
Mo5	Sylt, oxic layer of sandy sediment	NOR5-1B	pink
Mo10red	Sylt, oxic layer of sandy sediment	NOR5-3	dark red
Mo12red	Sylt, oxic layer of sandy sediment	NOR5-3	dark red
Mel5	Sylt, 15 cm depth of muddy sediment	NOR5-3	dark red
Mel6	Sylt, 15 cm depth of muddy sediment	NOR5-3	dark red
Mel7	Sylt, 15 cm depth of muddy sediment	NOR5-3	dark red

[a] From strain RAp14, two different 16S rRNA sequences were retrieved, therefore it might be a mixture from two strains.

The strains can be grown in SYPG rich medium, either in liquid or on agar plate. After a transfer, the colonies are usually visible only after 7 – 15 days on the agar plate. Besides the strain KT71, whose colonies are transparent to opal, colony pigmentation of the other strains are clearly separated according to the 16S rRNA phylogeny: NOR5-3 colonies turn from white to dark red until dark brown. The bigger the colonies grow, the darker the color is. Colonies of many NOR5-3 strains are sticky. On the other hand, the colonies of NOR5-1B strains appear later than the NOR5-3 strains. The colonies are smaller, not sticky, and the color is usually transparent to light pink. The suspected mixture of two strains – RAp14 has dark red color, and the agar medium is also stained brownish.

In the liquid medium, the cells of NOR5-3 strains usually grow attaching to the bottom of the plastic cell culturing bottles into a fluffy layer. After moderate shaking, the cells can be suspended in the liquid. The NOR5-1B cultures do not attach to the bottom, but well separated.

The growth curve is difficult to measure. Shaking seems to influence the growth of the strains, since aggregation could be necessary for growth. Also, opening the bottle may change of oxygen concentration in the medium and therefore influence the growth rate.

2.3.2 Pigments

The pigments of the cells were extracted using a mixture of aceton:methanol = 7:2, and the samples went through the high performance liquid chromatography (HPLC) by the washing solvent acetonitril:methanol:tetrahydrofurane = 15:3:2. The results showed that a pigment composition (with 3 highest peaks at the 363.7, 753.1 and 589.3 nm, see Unit 3, Table 3) that is possibly the bacteriochlorophyll a (BChl *a*), was found in the extract of strains Ivo10red and Mo12red (both NOR5-3), but not in the strains KT71 and Mo10red (both NOR5-3) and RAp14red (NOR5-3/1B), Mo4 and Ivo14 (both NOR5-1B). However, the BChl *a* has been proved from the culture of KT71 (Fuchs et al., 2007), since the expression can be influenced by the culturing conditions.

2.3.3 General genomic features

According to the 16S rRNA sequences, 18 of the North Sea strains belong to the NOR5-3 subclade, and 5 belong to NOR5-1B (see Unit 2, Fig. 5, left). From the strain RAp14, two types of 16S rRNA can be acquired. The 16S rRNA sequences of NOR5-1B subclade are nearly identical. The strains of NOR5-3 subclade can be divided into four groups: 12 sequences including RAp1red, 5 sequences including RAp7 and one sequence of RAp14, and the sequences of strains KT71 and Mo10red are different to all the others. Inside the first two groups respectively, the 16S rRNA sequences are nearly identical. The 16S rRNA similarity between the groups of NOR5-3 is 98.9 – 99.5%.

In order to see whether these strains are identical beyond the 16S rRNA level, the pulse field gel electrophoresis (PFGE) was used to test on the 12 strains, whose 16S rRNA sequences are identical, including the strain RAp1red. The genomes were cut with endonuclease *Swa*I, which recognizes the signature 5′-ATTT|AAAT-3′. The results (see Unit 3, Figure 1) showed that each genome of the 12 strains is different with each other. Therefore, the strains are not identical.

2.3.4 *puf*M genes

The *puf*M gene coding for reaction center M chain was amplified from the North Sea strain using the primer set pufL_WW_F (5′-Y TAV TGG TGG VVN TGG TGG-3′, designed in this study) and pufM_uni_R (5′-YC CAT NGT CCA NCK CCA RAA-3′ reverse (Yutin et al., 2005)). The *puf*M genes are amplified from all the strains except KT71 and Ivo19, and the sequences were retrieved from most of them. The phylogenetic tree of *puf*M (see Unit 2, Fig. 5, right) showed high parallelism to the 16S rRNA tree. Interestingly, from the strain RAp14red, in which 2 types of 16S rRNA sequences were amplified, also 2 types of *puf*M clones are obtained, each one resembling the sequences from the NOR5-1B and NOR5-3 clade. Therefore, the genes for the reaction center of light-harvesting complex I (LHC I) exist in all or most of the NOR5/OM60 strains of the North Sea, and there is no trace for a gene lateral gene transfer.

2.4 Genomics

2.4.1 General comparison of the genomes

The genomes of six strains: KT71, RAp1red, Ivo14, HTCC2080, HTCC2148 (all NOR5/OM60) and HTCC2143 (BD1-7) were fully sequenced. The general information of the genomes is listed in Table 4.

Table 4 Basic information of the six genomes in this study

	KT71	RAp1red	Ivo14	HTCC2080	HTCC2148	HTCC2143
Clade	NOR5-3	NOR5-3	NOR5-1B	NOR5-1B	NOR5-8	BD1-7
Scaffolds	2	6	1	2	31	4
Total length (bp)	4,344,414	4,208,084	3,261,541	3,582,105	4,326,936	3,940,784
Percentage of N*	0.40%	0.10%	0.43%	0.17%	0.39%	0.38%
G+C content	57.68%	56.34%	56.74%	51.82%	52.96%	47.16%
rRNA operons	2	2 – 4	1	1	1 – 3	1

*N indicates undetermined nucleotides (besides A, G, C and T)

The genomes are pairwise compared using the PROmer program (Figure 11). In the figure, the homologous regions are labeled on the plot (red for the same direction, and blue for reverse). In this way, we can estimate the genome-wide relationship between each genome pair including synteny of genes. It also indicates which genes are orthologous, and whether the scaffolds are correctly arranged. This analysis showed that KT71 and RAp1red are very closely related, and the 6 scaffolds of RAp1red were not arranged in order (Figure 11a). After reverse and rearranging the RAp1red scaffolds

(Figure 11b), we can see that most of the genes are aligned on the diagonal line. Therefore, it indicates that most of the genes have not gone through a gene lateral transfer. The relationship of KT71 to Ivo14 and HTCC2080 are more distant (Figure 11c, d). The HTCC2148 genome was split in 31 scaffolds, so it is too hard to rearrange them as for RAp1red. Nevertheless, the plot from the larger scaffolds of HTCC2148 to KT71 shows a more distant relationship of the two genomes (Figure 11e). Between KT71 and HTCC2143, only scattered regions are well aligned (Figure 11f). This indicated a much more distant relationship. In the time since the last common ancestor, many events of genome rearrangement and lateral gene transfer must have occurred.

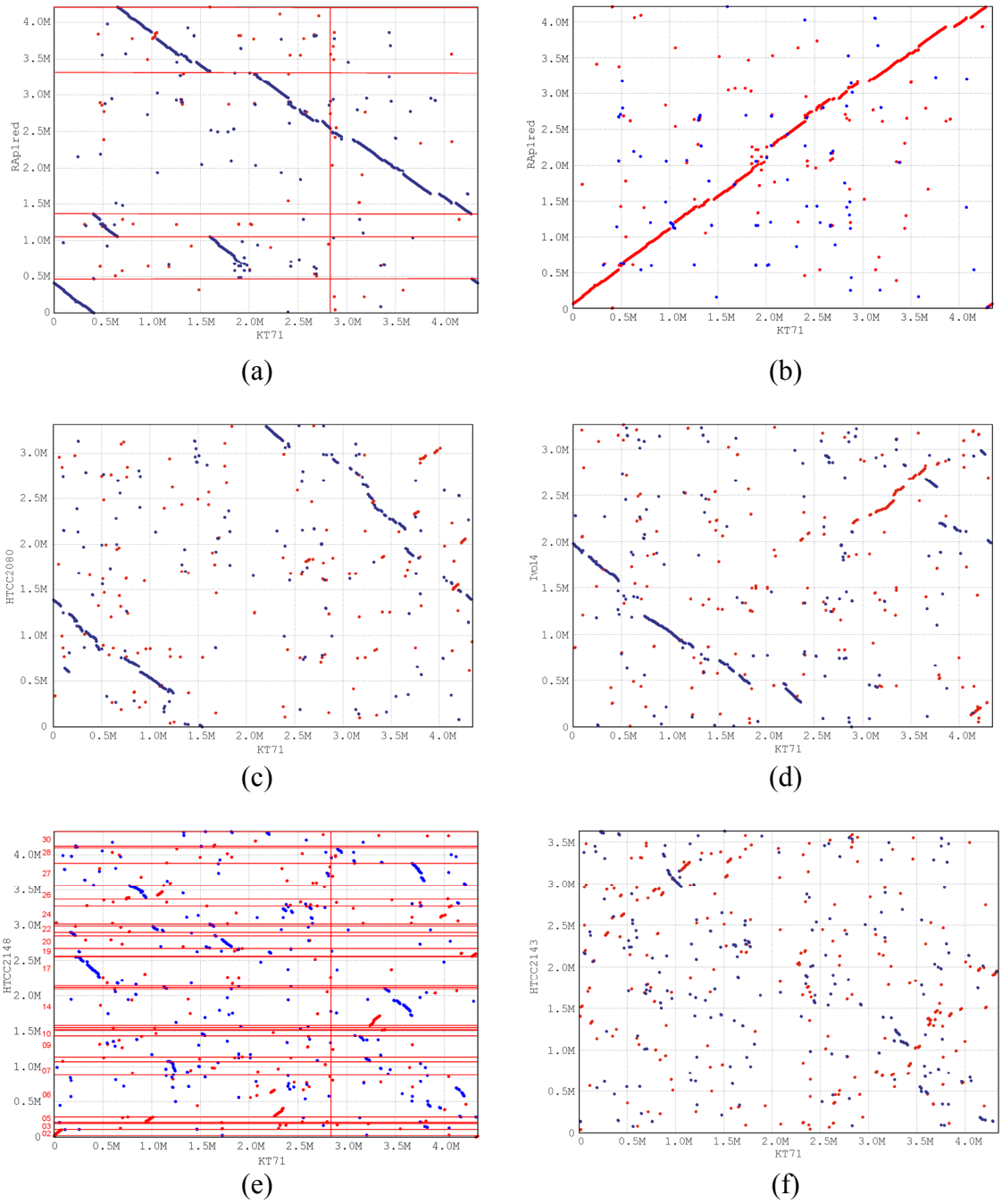

Figure 11 Pairwise alignment of the genomes. The aligned region in the same direction was plotted as red, and reverse complement as blue. (a) KT71 (x-axis) – RAp1red (y-axis) with the published scaffold order; (b) KT71 – RAp1red, with scaffolds of RAp1red rearranged in the order 1-3-5-2-4-6 and reverse-complemented; (c) KT71 – HTCC2080,; (d) KT71 – Ivo14; (e) KT71 – HTCC2148; (f) KT71 – HTCC2143. In (a) and (e), the red lines label the border of scaffolds. In (c) and (d), the green circles label the locations of PS super-operons.

For each pairwise alignment, the orthologous regions were summed up using a self-made java script, and the percentage of total orthologous regions were calculated for each pair of genomes. The similarity matrix was made. The neighbor-joining trees for genome homology and 16S rRNA showed the same topology (Figure 12). This once again indicates that 16S rRNA-based phylogeny is a good proxy for genome evolution.

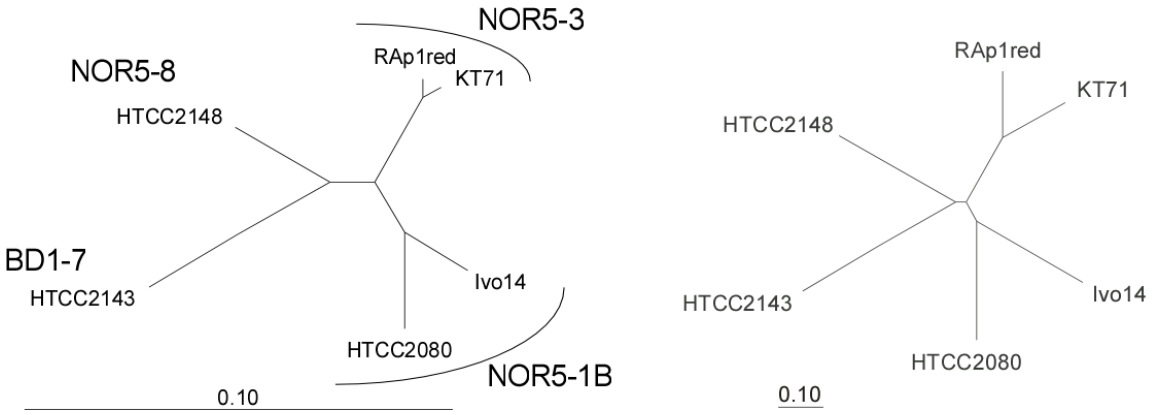

Figure 12 Neighbor joining trees based on 16S rRNA similarity (left) and genome homology (right).

2.4.2 Functional genes

2.4.2.1 PS superoperon

The photosynthesis (PS) superoperon was found in four genomes: KT71, RAp1red, Ivo14 and HTCC2080 (Figure 13), but was fully absent in the genomes of HTCC2148 and HTCC2143. The superoperon contains *bch* (bacteriochlorophyll synthesis), *puf* (light-harvesting complex I (LHC I) and reaction center) and *crt* (carotenoid synthesis) genes. The organization of the PS superoperon is highly similar, at a length of 40 – 45 kbp, but not exactly identical. The *puf*LMCBA arrangement is the same and unique for *Gammaproteobacteria* (Yutin and Béjà, 2005).

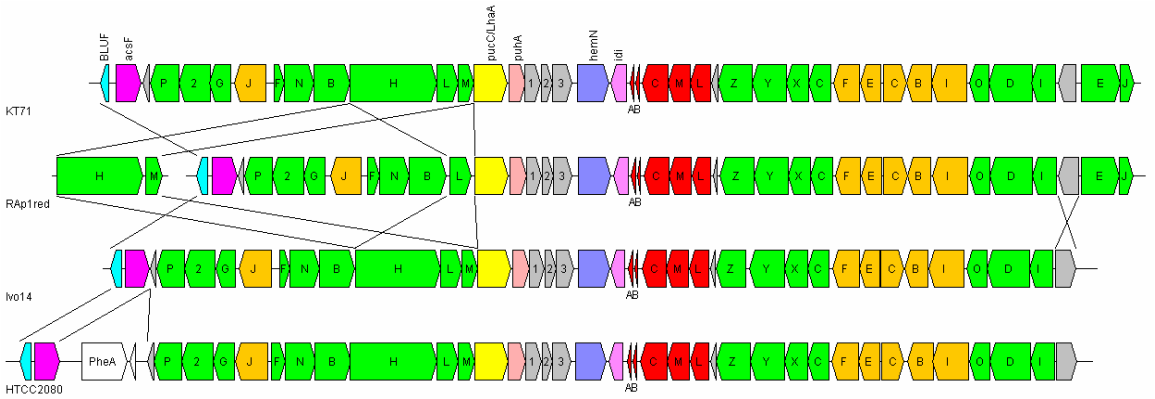

Figure 13 Comparison of PS operons of KT71, RAp1red, Ivo14 and HTCC2080. Green, *bch* genes; red, *puf* genes; orange, *crt* genes; light grey, unknown conserved genes. The *bch*HM genes of RAp1red locate on ~24 kbp upstream of PS superoperon on the same scaffold (scaffold 4).

According to the former study (Yutin et al., 2007), the *puf*M genes from the NOR5-1B group are located inside the Group K, while the gene from KT71 is the closest relative of Group K. However, the *puf*L and *puf*M sequences of HTCC2148 and HTCC2246 (also a NOR5/OM60 member, which could not be grouped into any subclades), which were acquired using PCR amplification, did not group with other NOR5/OM60 members, but rather with two different groups of *Alphaproteobacteria* (Cho et al., 2007). This puzzle was not solved in this study, since the *puf*LM as well the whole PS superoperon were not found in the genome of HTCC2148. Since the probability that the whole PS superoperon was missing from the genome sequencing is not high, it is most possible that contamination of other photosynthetic bacteria happened during the PCR of the *puf* genes of HTCC2148.

2.4.2.2 Proteorhodopsin

The proteorhodopsin genes (*pop*) were found only in the genomes of HTCC2148 and HTCC2143, but not in the four genomes in which the PS superoperon was present. The *pop* gene of HTCC2148 is located at the beginning of a very short scaffold (scaffold 18, 4490 bp) and the sequence is not complete (540 bp), while the *pop* gene in HTCC2143 is complete (690 bp).

The *pop* gene of HTCC2143 is the closest relative of the SAR92 group, one of the most closely related groups to the NOR5/OM60 clade and HTCC2143 (16S rRNA sequence identities between the groups are 88 – 92%), while that of HTCC2148 also

cluster with other *Alpha-* and *Gammaproteobacteria*, although the exact position cannot be determined due to incompleteness of its sequence.

Downstream of the HTCC2143 proteorhodopsin gene are the genes for retinal synthesis, in the order *pop-crt*EIBY-*blh-fni* (*crt*E = *ids*A), all translated in the same direction. This gene arrangement is exactly the same as in HTCC2207 (Stingl et al., 2007). Therefore, the existence of proteorhodopsin in HTCC2143 is convincing. However, the genes for retinal synthesis are not found in the genome of HTCC2148, and the downstream of *pop* are functionally unrelated genes. Since retinal is the chromophore for rhodopsin, the functionality of *pop* gene in the genome of HTCC2148 is therefore quite questionable.

2.4.2.3 Carbon fixation

The key genes of Calvin Cycle, reverse citric acid cycle and reductive acetyl-CoA pathway were not found in any of the six genomes. However, in this study, several genes of the 3-hydroxypropionate cycle were identified in the four genomes of NOR5/OM60 strains. This includes the malonyl-CoA reductase gene (*mcr*) and the the propionyl-CoA synthase gene (*pcs*). These are two key genes which have not been found to be involved in any pathway other than the carbon-fixing 3-hydroxypropionate cycle (Hügler et al., 2002). The two genes were found in the tandem arrangement as *pcs-mcr* in the genomes of RAp1red, Ivo14 and HTCC2080. We have found only *pcs* in KT71, while *mcr* is missing as reported before (Friedmann et al., 2007).

Until now, these large genes (for HTCC2080, *mcr* 3651 bp and *pcs* 5421 bp) can be found only in a few strains: *Chloroflexus spp.*, *Roseiflexus spp.* (both *Chloroflexi*) and *Erythrobacter sp.* NAP1 (*Alphaproteobacteria*); a single *pcs* gene was found in *Chloroherpeton thalassium* ATCC35110 (*Chlorobi*). This is the first time that these genes are found in *Gammaproteobacteria* and the second time in *Proteobacteria*.

A comparative sequence analysis for all the available genomic *pcs* genes to date (Figure 14) shows clustering of the NOR5/OM60 sequences. The *pcs* sequence of the strain Ivo14 is closer to that of the other North Sea strains than to HTCC2080, which means that the *pcs* phylogeny is not parallel to 16S rRNA phylogeny. The similarity of *pcs* from all the sources is high (e.g. 46 – 54% amino acid identity between the

NOR5/OM60 and *Chloroflexi* sequences). Therefore it is highly possible that the *pcs* genes in NOR5/OM60 have the same function as in *Chloroflexi*.

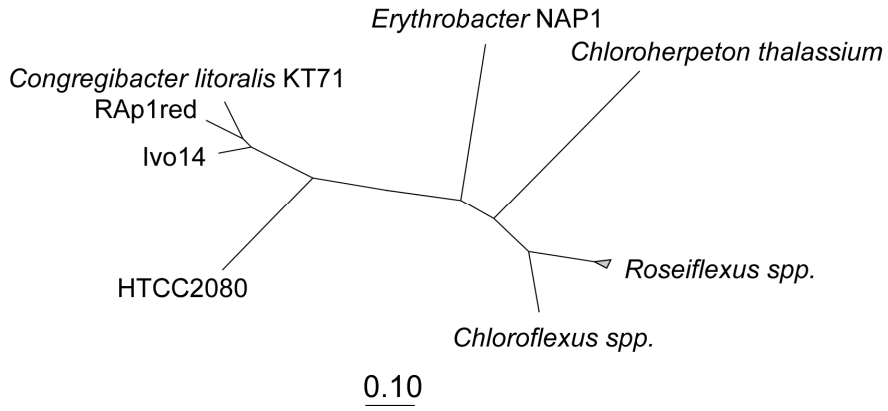

Figure 14 Maximum likelihood tree of genes for all the known propionyl-CoA synthase (*pcs*) genes from genomes. Both *Chloroflexus* and *Roseiflexus* belong to the phylum *Chloroflexi*, while *Chloroherpeton* belongs to *Chlorobi* and *Erythrobacter* belongs to *Alphaproteobacteria*.

The enzymes for the first step of 3-hydroxypropionate pathway, *acc*A, *acc*BC and *acc*D for acetyl-CoA carboxylase were found in all the five strains of the NOR5/OM60 clade, all separated at three isolated locations on the genomes. Genes for propionyl-CoA carboxylase (*pcc*BA), methylmalonyl-CoA epimerase (*mce*), methylmalonyl-CoA mutase (*mcm*) and a putative arginine/ornithine transport system ATPase occur tandemly in all the six genomes.

The last steps of 3-hydroxypropionate cycle in *Chloroflexus* are more complicated than previously thought and are still under investigation (Friedmann et al., 2007). For the supposed succinyl-CoA:L-malate CoA transferase and L-malyl-CoA lyase, homologs with relatively low similarity to those in *Chloroflexus* can be found in the NOR5/OM60 genomes. It is hard to judge whether the NOR5/OM60 strains use these enzymes to close the cycle. On the other hand it is possible that the NOR5/OM60 strains may use a different pathway to recycle succinyl-CoA and to regenerate acetyl-CoA.

The absence of the *mcr* gene in KT71 is in accordance with the fact that KT71 was not able to grow autotrophically in physiological tests (Fuchs et al., 2007). The reason why it still keeps the huge *pcs* gene is not clear yet. The only other reported strain from *Proteobacteria*, the alphaproteobacterial AAnP *Erythrobacter sp.* NAP1, was

proved to be able to assimilate CO_2 (Kolber et al., 2001). The daily cellular CO_2 fixation rate was 3% of the cellular carbon content and contributed to about 1% of total carbon anabolism.

Since it is the first time that the *pcs* gene is found in *Gammaproteobacteria*, we searched for its homologous sequence using BLAST against metagenomic databases. Hundreds of homologous sequences were found from the Global Ocean Survey (GOS) project (http://camera.calit2.net/index.php) (Rusch et al., 2007), and many of them are obviously more similar to the sequences of the NOR5/OM60 strains than to the other groups (e-values differentiate more than 10^{30} times). The sampling locations at which *pcs* genes were sequenced are also widely distributed. Therefore, the 3-hydroxypropionate pathway might be a common route for carbon fixation in the marine surface layer, and more studies in detail are expected to determine if they belong to the NOR5/OM60 group.

2.4.2.4 Sulfur compound oxidation genes

The *sox* operon encoding enzymes for the oxidation of sulfur compounds is present in all genomes containing the PS-superoperon, i.e. KT71, RAp1red, Ivo14 and HTCC2080 (Figure 15), but not in HTCC2148 and HTCC2143. Among all the *sox* genes, *sox*CDXYZAB are the core genes for reducing thiosulfate (Friedrich et al., 2005).

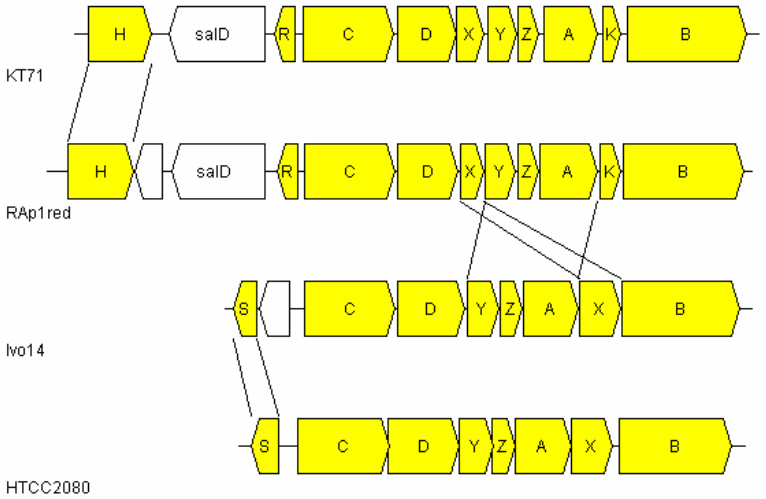

Figure 15 Arrangement of *sox* operon in KT71, RAp1red, Ivo14 and HTCC2080 genomes. The *sox*X of KT71 and RAp1red show low similarity and different length with those of Ivo14 and HTCC2080.

The operon arrangement *sox*CDXYZAKB in KT71 and RAp1red is the same as in several *Gamma*- and *Alphaproteobacteria*, like *Neptuniibacter caesariensis* MED92 and

Methylobacterium sp. 4-46. However, the *sox* operon of Ivo14 and HTCC2080 has the arrangement *sox*CDYZAXB, the same as in several *Betaproteobacteria*, like *Dechloromonas aromatica* RCB. The closest relatives of several genes of *sox* genes of Ivo14 and HTCC2080 also fall in *Beta-* or *Deltaproteobacteria*, e.g. as shown for the *soxB* gene. All these features suggest a lateral gene transfer of the whole *sox* operon from *Beta-* or *Deltaproteobacteria*, while the *sox* operons of NOR5-3 strains are closer to other *Gammaproteobacteria*.

The distribution of several functional gene groups is summarized in Unit 2, Table 3.

2.5 Evolution and functions of NOR5/OM60 group in the ocean

Despite the wide spectrum of the environments where the sequences were discovered, the NOR5/OM60 clade still seems to have originated from the ocean, since all the sequences from freshwater-related environments (fresh water, sediment and soil) form one subclade NOR5-13, indicating a relatively recent adaptation. The deep-branching subclades, e.g. NOR5-10 and NOR5-12, are mainly consisted of deep-sea sequences. These sequences usually have long branches. It might be a hint that the NOR5/OM60 group evolved from the deep-sea. However, it can also be due to a higher mutation rate for the deep-sea members.

The branch including the NOR5-1 and NOR5-4 subclades is the largest of the NOR5/OM60 group. It is composed of nearly exclusively marine surface water sequences, whereas most other subclades contain also sequences from sediments. A hypothesis is that the ancestor of NOR5-1 and NOR5-4 has lost genes required for life in sediment and became specialized in marine surface water. This needs to be proved by further genomic investigations on more strains.

Since the PS superoperon exists in four strains and in two subclades of the NOR5/OM60 group, and both the gene arrangement and gene sequences indicate no sign for a main lateral gene transfer event, photosynthesis might be an intrinsic common trait for the NOR5/OM60 group inherited from their common ancestor. This would differentiate the NOR5/OM60 clade from the other *Gammaproteobacteria*. However, considering the rather high 16S rRNA sequence diversity within the NOR5/OM60 clade

and its broad habitat range, it cannot be taken for granted that all members of NOR5/OM60 are AAnPs. Especially it is interesting to know whether some NOR5/OM60 members from deep-sea sediment can utilize light energy, since some deep-sea AAnP strains have been reported (Yurkov et al., 1999).

The complementary distribution of the PS superoperon and proteorhodopsin in the six sequenced genomes of NOR5/OM60 is intriguing. Since the PS superoperon is large, and therefore expensive for the bacteria to maintain, some bacteria may have acquired rhodopsin during evolution, as an alternative for utilizing light energy, in order to afford the loss of the PS superoperon. The hypothesis that the PS superoperon and rhodopsin can substitute each other is supported by the fact that the *puf* and rhodopsin genes rarely co-exist in the same genome. Only three prokaryote genomes known to-date contain both *puf* and rhodopsin genes: cyanobacterial *Nostoc* sp. PCC7120, alphaproteobacterial *Methylobacterium* sp. 4-46 and chloroflexal *Roseiflexus* sp. RS-1. All the three rhodopsin sequences are distantly related to the proteobacterial rhodopsins and probably have different functions.

The same distribution of PS superoperon, 3-hydroxypropionate pathway as well as *sox* operon provides the possibility that some members of the NOR5/OM60 group (including NOR5-3 and NOR5-1B subclades) might be able to oxidize sulfur compounds and reduce and fix CO_2 using light energy. However, none of the three points were yet proved in the first physiological tests of KT71 (Fuchs et al., 2007). If these points can be experimentally proved in the other strains of the NOR5/OM60 group, it might lead to the discovery of an important style of photoautotrophy in the ocean. Since the phylogenies of the *pcs* and *sox* are not parallel to 16S rRNA, these three sets of genes do not seem to have been acquired at the same time during the evolution and might not apply for all the marine NOR5/OM60 bacteria.

Many other important functional genes, such as *sox* operon and flagella super-operon, since either they are not generally existing in the NOR5/OM60 group or the phylogenies are not congruent with the 16S rRNA, it seems they are not always inherited from the common ancestor of the NOR5/OM60 group, and so they are not common traits for the members of this group.

3 Outlook

3.1 Phylogeny and biogeography

This study included the first major effort to investigate the phylogeny and distribution of the NOR5/OM60 group. The coverage of potential sampling sites is still far from comprehensive. Further studies are needed, for example, from the deep-sea water column, deep sediment, or in Indian Ocean region. A more systematic sampling effort could also unveil the relationship of the NOR5/OM60 group and its subclades with environmental parameters and provide hints on how the bacteria are living.

Probe sets for several subclades (NOR5-1AC, NOR5-1B, NOR5-3 and NOR5-4) were designed and optimized in this study. They have not yet been applied onto many environmental samples, partially due to the low counts of the subclades in the ocean. Recently, a new technique was developed for counting specific cells with low percentage in the environments using CARD-FISH onto densely made filters (Gomez, personal communication). This will bring more detailed data for distribution of separate subclades.

3.2 Comparative genomics

Based on discrepancies discovered in this study, there is an urgent need to re-sequence strain HTCC2148, which belongs to the subclade NOR5-8. This study has contributed mostly on NOR5-3 and NOR5-1B subclades. Therefore, an elaborate study on HTCC2148 will enable us to better understand the common features of the NOR5/OM60 clade. However, the incompleteness of its genome sequence left several questions open: does HTCC2148 have the *puf*LM genes and the PS superoperon as earlier indicated (Cho et al., 2007)? And does the proteorhodopsin gene exist in the genome? Is it functional? Therefore, a re-sequencing of HTCC2148 and functional studies are necessary to address these questions. Amplification of the respective genes from the strain could be a first step in this direction.

The sequences from more available isolates will be also helpful, for example, HTCC2146, which does not belong to a known subclade, strain 3X/A02/235 (NOR5-2), and NEP-1 (NOR5-7).

3.3 Gene searching in metagenomic libraries

Metagenomics will help us to find out important or abundant genes from environments. The results in this study indicate that it can give semi-quantitative results of gene frequency and distribution. In this study, the 16S rRNA and *pcs* gene have been searched for their relatives in the GOS dataset. The results supported our hypothesis and data on the distribution of both genes that are related to the NOR5/OM60 group. Further novel gene sequences or characteristic gene arrangement of an operon can be searched in metagenomic databases, in order to get a first estimation of their prevalence.

3.4 Combining detection of FISH and functions

One possibility to directly determine the relationship of a phylogenetic group and its function is a combination of FISH and *in situ* tools for measuring cell functions. The functional identification can be on the gene level, expression level, or metabolite level.

Gene-FISH is conceivable to detect the presence of a single gene in the genomic DNA of a cell. A polynucleotide probe can be used for hybridization on genomic DNA. After washing and amplification steps, the signal is then amplified with fluorescent dye, which can be observed under the microscope (Moraru, unpublished). CARD-FISH can be done consequently with a dye with another color.

On the expression level, for example, the infrared autofluorescence signal of BChl *a* can be used for directly observation of AAnPs (Schwalbach and Fuhrman, 2005; Jiao et al., 2006). In this way, we could directly determine how many NOR5/OM60 are AAnP, and *vice versa*, how many of the AAnP are NOR5/OM60.

On the metabolite level, nano-SIMS (secondary ion mass spectroscopy) is a new technique to trace the isotopes of interest in the cells. Musat *et al.* have recently developed halogen *in situ* hybridization-secondary ion mass spectroscopy (HISH-SIMS), which uses halogen-labeled probes for identification instead of fluorescence-labeled, in order to trace the metabolites and to identify the phylogeny of the cells *in situ* at the same time (Musat et al., 2008).

3.5 Physiological tests for the model strains of NOR5/OM60

On the isolate strains, physiology tests should be done to determine whether and when the functional genes are expressed. Biochemical assays are required to understand

the important metabolic pathways, in the first place the photosynthesis, sulfur compound oxidization and the 3-hydroxypropionate pathway. It is important to know what genes are involved and what intermediates are produced in these pathways, in order to reconstruct the whole picture of these pathways.

4 References

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation. *Microbiol Rev* **59**: 143-169.

Arakawa, S., Sato, T., Sato, R., Zhang, J., Gamo, T., Tsunogai, U. et al. (2006) Molecular phylogenetic and chemical analyses of the microbial mats in deep-sea cold seep sediments at the northeastern Japan Sea. *Extremophiles* **10**: 311-319.

Barneah, O., Ben-Dov, E., Kramarsky-Winter, E., and Kushmaro, A. (2007) Characterization of black band disease in Red Sea stony corals. *Environ Microbiol* **9**: 1995-2006.

Béjà, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411**: 786-789.

Béjà, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T. et al. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**: 630-633.

Bowman, J.P., and McCuaig, R.D. (2003) Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment. *Appl Environ Microbiol* **69**: 2463-2483.

Brinkmeyer, R., Knittel, K., Jurgens, J., Weyland, H., Amann, R., and Helmke, E. (2003) Diversity and structure of bacterial communities in arctic versus antarctic pack ice. *Appl Environ Microbiol* **69**: 6610-6619.

Cho, J.C., and Giovannoni, S.J. (2004) Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria. *Appl Environ Microbiol* **70**: 432-440.

Cho, J.C., Stapels, M.D., Morris, R.M., Vergin, K.L., Schwalbach, M.S., Givan, S.A. et al. (2007) Polyphyletic photosynthetic reaction centre genes in oligotrophic marine Gammaproteobacteria. *Environ Microbiol* **9**: 1456-1463.

Connon, S.A., and Giovannoni, S.J. (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* **68**: 3878-3885.

Cottrell, M.T., Mannino, A., and Kirchman, D.L. (2006) Aerobic anoxygenic phototrophic bacteria in the Mid-Atlantic Bight and the North Pacific Gyre. *Appl Environ Microbiol* **72**: 557-564.

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

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U. et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496-503.

Eilers, H., Pernthaler, J., Glöckner, F.O., and Amann, R. (2000) Culturability and in situ abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol* **66**: 3044-3051.

Eilers, H., Pernthaler, J., Peplies, J., Glöckner, F.O., Gerdts, G., and Amann, R. (2001) Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton. *Appl Environ Microbiol* **67**: 5134-5142.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**: 237-240.

Frias-Lopez, J., Zerkle, A.L., Bonheyo, G.T., and Fouke, B.W. (2002) Partitioning of bacterial communities between seawater and healthy, black band diseased, and dead coral surfaces. *Appl Environ Microbiol* **68**: 2214-2228.

Friedmann, S., Alber, B.E., and Fuchs, G. (2007) Properties of R-citramalyl-coenzyme A lyase and its role in the autotrophic 3-hydroxypropionate cycle of Chloroflexus aurantiacus. *J Bacteriol* **189**: 2906-2914.

Friedrich, C.G., Bardischewsky, F., Rother, D., Quentmeier, A., and Fischer, J. (2005) Prokaryotic sulfur oxidation. *Current Opinion in Microbiology* **8**: 253-259.

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

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.

Fuchs, B.M., Spring, S., Teeling, H., Quast, C., Wulf, J., Schattenhofer, M. et al. (2007) Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *P Natl Acad Sci USA* **104**: 2891-2896.

Giovannoni, S.J., Bibbs, L., Cho, J.C., Stapels, M.D., Desiderio, R., Vergin, K.L. et al. (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**: 82-85.

Glatz, R.E., Lepp, P.W., Ward, B.B., and Francis, C.A. (2006) Planktonic microbial community composition across steep physical/chemical gradients in permanently ice-covered Lake Bonney, Antarctica. *Geobiology* **4**: 53-67.

Hartmann, M., and Widmer, F. (2006) Community structure analyses are more sensitive to differences in soil bacterial communities than anonymous diversity indices. *Appl Environ Microbiol* **72**: 7804-7812.

Huber, J.A., Johnson, H.P., Butterfield, D.A., and Baross, J.A. (2006) Microbial life in ridge flank crustal fluids. *Environ Microbiol* **8**: 88-99.

Hügler, M., Menendez, C., Schägger, H., and Fuchs, G. (2002) Malonyl-coenzyme A reductase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation. *J Bacteriol* **184**: 2404-2410.

Inagaki, F., Suzuki, M., Takai, K., Oida, H., Sakamoto, T., Aoki, K. et al. (2003) Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk. *Appl Environ Microbiol* **69**: 7224-7235.

Jiao, N.Z., Zhang, Y., and Chen, Y. (2006) Time series observation based InfraRed Epifluorescence Microscopic (TIREM) approach for accurate enumeration of bacteriochlorophyll-containing microbes in marine environments. *J Microbiol Meth* **65**: 442-452.

Jiao, N.Z., Yang, Y.H., Koshikawa, H., and Watanabe, M. (2002) Influence of hydrographic conditions on picoplankton distribution in the East China Sea. *Aquat Microb Ecol* **30**: 37-48.

Jiao, N.Z., Sieracki, M.E., Zhang, Y., and Du, H.L. (2003) Aerobic anoxygenic phototrophic bacteria and their roles in marine ecosystems. *Chinese Science Bulletin* **48**: 1064-1068.

Jiao, N.Z., Zhang, Y., Zeng, Y.H., Hong, N., Liu, R.L., Chen, F., and Wang, P.X. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* **9**: 3091-3099.

Keller, L. (2003) Herbstsukzessionen und Aktivität der pelagischen Bakteriengemeinschaft in der Deutschen Bucht. In *Max-Planck-Institut für Marine Mikrobiologie*. Bremen, p. 89.

Klein, A.N., Frigon, D., and Raskin, L. (2007) Populations related to Alkanindiges, a novel genus containing obligate alkane degraders, are implicated in biological foaming in activated sludge systems. *Environ Microbiol* **9**: 1898-1912.

Kolber, Z.S., van Dover, C.L., Niederman, R.A., and Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**: 177-179.

Kolber, Z.S., Plumley, F.G., Lang, A.S., Beatty, J.T., Blankenship, R.E., van Dover, C.L. et al. (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* **292**: 2492-2495.

Koren, O., and Rosenberg, E. (2006) Bacteria associated with mucus and tissues of the coral Oculina patagonica in summer and winter. *Appl Environ Microbiol* **72**: 5254-5259.

Ley, R.E., Harris, J.K., Wilcox, J., Spear, J.R., Miller, S.R., Bebout, B.M. et al. (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72**: 3685-3695.

Li, L.N., Kato, C., and Horikoshi, K. (1999) Bacterial diversity in deep-sea sediments from different depths. *Biodivers Conserv* **8**: 659-677.

Liang, J.B., Chen, Y.Q., Lan, C.Y., Tam, N.F.Y., Zan, Q.J., and Huang, L.N. (2007) Recovery of novel bacterial diversity from mangrove sediment. *Mar Biol* **150**: 739-747.

Liao, P.C., Huang, B.H., and Huang, S. (2007) Microbial community composition of the Danshui river estuary of northern Taiwan and the practicality of the phylogenetic method in microbial barcoding. *Microb Ecol* **54**: 497-507.

Liles, M.R., Manske, B.F., Bintrim, S.B., Handelsman, J., and Goodman, R.M. (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl Environ Microbiol* **69**: 2684-2691.

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

MacGregor, B.J., Toze, S., Alm, E.W., Sharp, R., Ziemer, C.J., and Stahl, D.A. (2001) Distribution and abundance of Gram-positive bacteria in the environment: development of a group-specific probe. *J Microbiol Meth* **44**: 193-203.

Madigan, M.T., and Martinko, J.M. (2006) *Brock Biology of Microorganisms, 11th Edition*: Pearson Prentice Hall.

Manz, W., Amann, R., Ludwig, W., Wagner, M., and Schleifer, K.H. (1992) Phylogenetic Oligodeoxynucleotide Probes for the Major Subclasses of Proteobacteria - Problems and Solutions. *Syst Appl Microbiol* **15**: 593-600.

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F. et al. (2008) A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *P Natl Acad Sci USA* **105**: 17861-17866.

Patzelt, H., Simon, B., terLaak, A., Kessler, B., Kuhne, R., Schmieder, P. et al. (2002) The structures of the active center in dark-adapted bacteriorhodopsin by solution-state NMR spectroscopy. *P Natl Acad Sci USA* **99**: 9765-9770.

Pernthaler, A., and Pernthaler, J. (2005) Diurnal variation of cell proliferation in three bacterial taxa from coastal North Sea waters. *Appl Environ Microbiol* **71**: 4638-4644.

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

Pernthaler, J., Glöckner, F.O., Schönhuber, W., and Amann, R. (2001) Fluorescence in situ hybridization (FISH) with rRNA-targeted oligonucleotide probes. In *Methods in Microbiology, Vol 30*. San Diego: ACADEMIC PRESS INC, pp. 207-226.

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

Rappé, M.S., Kemp, P.F., and Giovannoni, S.J. (1997) Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. *Limnol Oceanogr* **42**: 811-826.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S. et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: 398-431.

Schramm, A., Fuchs, B.M., Nielsen, J.L., Tonolla, M., and Stahl, D.A. (2002) Fluorescence in situ hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries. *Environ Microbiol* **4**: 713-720.

Schwalbach, M.S., and Fuhrman, J.A. (2005) Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR. *Limnol Oceanogr* **50**: 620-628.

Sekiguchi, H., Watanabe, M., Nakahara, T., Xu, B.H., and Uchiyama, H. (2002) Succession of bacterial community structure along the Changjiang River determined by denaturing gradient gel electrophoresis and clone library analysis. *Appl Environ Microbiol* **68**: 5142-5150.

Sieracki, M.E., Gilg, I.C., Thier, E.C., Poulton, N.J., and Goericke, R. (2006) Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnol Oceanogr* **51**: 38-46.

Simpson, J.M., Domingo, J.W.S., and Reasoner, D.J. (2004) Assessment of equine fecal contamination: the search for alternative bacterial source-tracking targets. *FEMS Microbiol Ecol* **47**: 65-75.

Stingl, U., Desiderio, R.A., Cho, J.C., Vergin, K.L., and Giovannoni, S.J. (2007) The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73**: 2290-2296.

Uzuka, N., Watanabe, S., and Tsunogai, S. (1996) Dimethylsulfide in coastal zone of the East China Sea. *Journal of Oceanography* **52**: 313-321.

Vernon, S.D., Shukla, S.K., Conradt, J., Unger, E.R., and Reeves, W.C. (2002) Analysis of 16S rRNA gene sequences and circulating cell-free DNA from plasma of chronic fatigue syndrome and non-fatigued subjects. *BMC Microbiology* **2**.

Wobus, A., Bleul, C., Maassen, S., Scheerer, C., Schuppler, M., Jacobs, E., and Roske, I. (2003) Microbial diversity and functional characterization of sediments from reservoirs of different trophic state. *FEMS Microbiol Ecol* **46**: 331-347.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a Natural System of Organisms - Proposal for the Domains Archaea, Bacteria, and Eucarya. *P Natl Acad Sci USA* **87**: 4576-4579.

Yurkov, V.V., Krieger, S., Stackebrandt, E., and Beatty, J.T. (1999) Citromicrobium bathyomarinum, a novel aerobic bacterium isolated from deep-sea hydrothermal vent plume waters that contains photosynthetic pigment-protein complexes. *J Bacteriol* **181**: 4517-4525.

Yutin, N., and Béjà, O. (2005) Putative novel photosynthetic reaction centre organizations in marine aerobic anoxygenic photosynthetic bacteria: insights from metagenomics and environmental genomics. *Environ Microbiol* **7**: 2027-2033.

Yutin, N., Suzuki, M.T., and Béjà, O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl Environ Microbiol* **71**: 8958-8962.

Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., and Béjà, O. (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ Microbiol* **9**: 1464-1475.

List of Publications and Manuscripts

Publications presented in this thesis

Yan, S., Fuchs B. M., Lenk S., Harder J., Wulf J., Jiao N. Z., and Amann R., 2009. Biogeography and phylogeny of the NOR5/OM60 clade of Gammaproteobacteria. Systematic and Applied Microbiology **32:**124-139.

Manuscripts presented in this thesis

Yan, S., Fuchs B. M., Harder J., and Amann R., Potential novel photoautotrophy in the NOR5/OM60 clade of *Gammaproteobacteria* discovered by genome comparison.

Yan, S. Characterization of the NOR5/OM60 strains from the North Sea (in preparation)

Further publications

Fuchs, B. M., Spring S., Teeling H., Quast C., Wulf J., Schattenhofer M., Yan S., Ferriera S., Johnson J., Glöckner F. O., and Amann R., 2007. Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. Proceedings of the National Academy of Sciences of the United States of America **104:**2891-2896.

Unit 1

Biogeography and phylogeny of the NOR5/OM60 clade of *Gammaproteobacteria*

Available online at www.sciencedirect.com
ScienceDirect
Systematic and Applied Microbiology 32 (2009) 124–139
SYSTEMATIC AND APPLIED MICROBIOLOGY
www.elsevier.de/syapm

Biogeography and phylogeny of the NOR5/OM60 clade of *Gammaproteobacteria*

Shi Yan[a], Bernhard M. Fuchs[a,*], Sabine Lenk[a], Jens Harder[b], Jörg Wulf[a], Nian-Zhi Jiao[c], Rudolf Amann[a]

[a]*Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Bremen D-28359, Germany*
[b]*Department of Microbiology, Max Planck Institute for Marine Microbiology, Bremen D-28359, Germany*
[c]*National Key Laboratory for Marine Environmental Sciences, University of Xiamen, Xiamen, Fujian 361005, China*

Abstract

The phylogeny, abundance, and biogeography of the NOR5/OM60 clade was investigated. This clade includes "*Congregibacter litoralis*" strain KT71, the first cultured representative of marine aerobic anoxygenic phototrophic *Gammaproteobacteria*. More than 500 16S rRNA sequences affiliated with this clade were retrieved from public databases. By comparative sequence analysis, 13 subclades could be identified, some of which are currently restricted to discrete habitat types. Almost all sequences in the largest subclade NOR5-1 and related subclade NOR5-4 originated from marine surface water samples. Overall, most of the NOR5/OM60 sequences were retrieved from marine coastal settings, whereas there were fewer from open-ocean surface waters, deep-sea sediment, freshwater, saline lakes and soil.

The abundance of members of the NOR5/OM60 clade in various marine sites was determined by fluorescence *in situ* hybridization using a newly designed and optimized probe set. Relative abundances in coastal marine waters off Namibia and the Yangtze estuary were up to 3% of the total 4′,6-diamidino-2-phenylindole (DAPI) counts, and in the German Bight off Helgoland the abundance was even up to 7%. In an open-ocean North Atlantic transect, between Iceland and the Azores, the NOR5/OM60 group was much less abundant (0.1–0.5%). Interestingly, the surface layer of North Sea intertidal sediments was very rich in NOR5/OM60, with absolute numbers $>10^8$ cells cm^{-3} (or 4% of the total DAPI). An analysis of the frequencies of NOR5/OM60 16S rRNA genes in the Global Ocean Survey datasets provided further support for a marine cosmopolitan occurrence of NOR5/OM60, and a clear preference for coastal marine waters.

Keywords: Microbial ecology; Bacteria; Fluorescence *in situ* hybridization; Phylogeny; Biogeography

Introduction

Aerobic anoxygenic phototrophic bacteria (AAnP) use light as an energy source, and appear to have an important role in marine carbon cycling [33,34]. They also seem to be highly abundant in the oceans. Recent studies based on infrared microscopy showed abundances of $4.5 \pm 2.4\%$ with a maximum of 13.5% in coastal waters, while in oceanic water the frequency was lower at $1.5 \pm 1.3\%$ [30,56,62]. For a long time, all cultured representatives of marine AAnP belonged to

*Corresponding author.
E-mail address: bfuchs@mpi-bremen.de (B.M. Fuchs).

0723-2020/$ - see front matter © 2008 Elsevier GmbH. All rights reserved.
doi:10.1016/j.syapm.2008.12.001

the class *Alphaproteobacteria*. This recently changed when genome analyses of marine gammaproteobacterial isolates revealed the presence of complete photosynthesis superoperons [22]. The North Sea isolate KT71 was identified as AAnP based on the expression of photosynthetic pigments [22]. In parallel, HTCC2080, one of three gammaproteobacterial strains obtained by high-throughput cultivation from coastal Pacific surface water off Oregon, was shown to grow faster and to have higher cell yield with light rather than in the dark [12]. All four isolates are from one monophyletic gammaproteobacterial clade that had been predicted based on metagenomics [8,58].

The history of this clade dates back to 1995, when a 284 bp gammaproteobacterial 16S rRNA sequence retrieved from Sargasso Sea surface water was referred to as SAR125 (L35466) [47]. In 1997, two almost full-length sequences were found to be closely related to SAR125; clones OM60 (U70696) and OM241 (U70702) were from a marine coastal site off North Carolina, USA [51]. In 1999, strain KT71 (AY007676) was isolated from marine surface water at the "Kabeltonne" station off the island of Helgoland, North Sea [19]. Strain KT71 was placed in a group named NOR5, and recently the binominal name "*Congregibacter litoralis*" has been suggested for this isolate [22]. Several related strains isolated by a novel high-throughput culturing method, including HTCC2080, were placed in the OM60/OM241 clade [14], which was later referred to as the OM60 clade [11]. Since all these clade names are now largely redundant, we refer to it here as the NOR5/OM60 clade [22] and, currently, many more related sequences can be found in various databases.

Based on comparative 16S rRNA sequence analysis, the NOR5/OM60 clade is most closely related to the genera *Endobugula*, *Microbulbifer*, *Teredinibacter* (all *Alteromonadales*), *Cellvibrio* (*Pseudomonadales*) and several other groups of oligotrophic marine *Gammaproteobacteria*, including the clades BD1-7, KI89A, OM182 and SAR92 [11]. A sequence retrieved from deep-sea sediment, BD2-7 (AB015537) [37], was considered to represent a sister clade to the NOR5/OM60 clade.

Data on the biogeography of the NOR5/OM60 clade are so far sparse. Besides the mostly qualitative evidence from sequence retrieval there is also some quantification by fluorescence *in situ* hybridization (FISH). The probe NOR5-730 has yielded counts in North Sea surface water of up to 8% [19], and even 11% [48] of all DAPI-stained cells. With the same probe, an abundance of $3.4 \pm 1.1\%$ was detected in a Pacific coastal transect along the Newport Hydroline [12], where maximum numbers of up to 1×10^5 cells ml^{-1} were linked to the chlorophyll maximum.

The goals of this study were to provide a more detailed description of the phylogeny of the NOR5/OM60 clade and to analyze its geographic distribution, as well as abundance, in the marine environment. Therefore, all available NOR5/OM60 16S rRNA sequences from public databases were mapped. This information originated mostly from PCR-based clone libraries, but also encompassed isolates and the metagenomic Global Ocean Survey (GOS) [52]. Based on this comprehensive dataset the specificity of published 16S rRNA-targeted oligonucleotide probes was checked, and new ones were designed. A probe mixture was optimized for a catalyzed reporter deposition FISH (CARD-FISH) assay. With this, the abundance of members of the NOR5/OM60 clade was quantified in coastal marine sites off China, Namibia, Spain, and Germany, as well as in an open-ocean transect in the North Atlantic.

Materials and methods

Sequence retrieval

NOR5/OM60-related 16S rRNA sequences were initially retrieved with the ARB program (http://www.arb-home.de/) [43] from the SILVA database (Version 91) [50] (http://www.arb-silva.de/) by targeting group-specific signatures (e.g. complement to probe NOR5-730 with 0–2 mismatches). Highly related sequences were also identified using Blastn (http://www.ncbi.nlm.nih.gov/BLAST/), and imported into the database. Additionally, the 16S rRNA subset of the Camera GOS Cruise, a metagenomic project that sampled various environments, including pelagic and coastal seawater, fresh water or hypersalinic environments [52], was downloaded (http://web.camera.calit2.net/) and analyzed in a similar way.

In total, more than 500 16S rRNA sequences including 147 almost full-length sequences with a length of >1400 nucleotides were collected. Sequences were manually checked for sequence quality and chimera, using the pintail value [5] provided with the SILVA 16S rRNA gene database and using the program "chimera check" at the website of the ribosomal database project (http://rdp8.cme.msu.edu/cgis/chimera.cgi?su=SSU) [13]. Twenty-eight probably chimeric NOR5/OM60 sequences were used only for biogeographic studies, but not for phylogenetic analysis or probe design. The sequences were aligned using the ARB aligner, and added to the universal parsimony tree using "ARB parsimony" with a "positional variability by parsimony" filter for *Bacteria* [50].

The 16S rRNA sequences from 22 newly isolated strains from sediment off the North Sea island Sylt (55.01°N, 8.26°E) were determined by standard molecular techniques (PCR, sequencing) and submitted

to GenBank under the accession numbers EU672847–EU672869. A detailed description of the strains will be published elsewhere (Harder, unpublished).

Phylogeny

All the qualified, almost full-length sequences (>1400 nt) that belonged to the NOR5/OM60 clade, as well as several closely related outgroup sequences (in total around 150 sequences), were selected for phylogenetic reconstructions. Three column filters were made in ARB and were used for selecting the columns with certain conservative levels for calculation. Each filter kept 1493, 1450 and 1393 bases, respectively. The sequences were filtered into different datasets, and each was then used as input for four different algorithms: "ARB neighbor joining" (Felsenstein correction), "ARB parsimony interactive", Maximum likelihood using AxML, and the posterior possibility algorithm using MrBayes (Version 3.1, http://mrbayes.csit.fsu.edu/) [27]. The MrBayes trees were built according to the manual (http://mrbayes.csit.fsu.edu/manual.php) with settings of a likelihood model in two parallel runs, each containing four chains. The program ran until the "average standard deviation of split frequencies" became less than 0.1. Then, the first thousand trees of the unstable generations were "burnt in" using "half-compat", and MrBayes consensus trees were constructed. Subsequently, the trees obtained using the three filtered sets and four methods were compared manually. Groups that were stable in all or most of the trees were named as subclades. Whenever the branching patterns varied in many of the trees, a multifurcation was introduced at that position [42].

Probe design, check and optimization

The "probe check" module in the ARB program was used to check probes NOR5-730 and NOR5-130 [19] for coverage and specificity, but also to design new probes for NOR5/OM60 subclades. Only sequences with >1400 nt were used for probe design. Candidate probes were also checked against the 16S rRNA sequence databases, including the partial sequences. Helper and competitor nucleotides were designed and tested as described before [21,46] (Table 1). The probes were optimized by performing hybridization with all the relevant helpers and competitors at varying formamide concentrations [23]. Optimizations were carried out at 46 °C on pure cultures with a fully complementary 16S rRNA.

Biogeographic analysis

For each of the NOR5/OM60 sequences retrieved from the databases, the geographic and environmental data were manually collected, either directly from

Table 1. Probes, helpers and competitors that were used in this study.

Name	Targeted group	Sequence (5′–3′)	Target site (16S rRNA *Escherichia coli* numbering)	Formamide	Reference
NOR5-730	NOR5/OM60 clade	TCG AGC CAG GAG GCC GCC	730–747	50%	[19]
NOR5-709h	n.a.	TTC GCC ACY GGT ATT CCT CCA	709–729	n.a.	This study
NOR5-659h	n.a.	GAA TTC TAC CTC CCT CTC YCG	659–679	n.a.	This study
NOR5-1238	NOR5/OM60 clade, excluding NOR5-1 and -4	CCC TCT GTG CGT TCC ATT	1238–1255	50–55%	This study
NOR5-1217h	n.a.	GTA GCA CGT GTG TAG CCC AGG	1217–1237	n.a.	This study
NOR5-1287h	n.a.	ATC CGG ACT ACG AAA CGT TTT	1287–1307	n.a.	This study
EUB I–III	*Bacteria*	GCT GCC TCC CGT AGG AGT, GCA GCC ACC CGT AGG TGT, GCT GCC ACC CGT AGG TGT	338–355	35%	[16]
NON	Negative control	ACT CCT ACG GGA GGC AGC	Reverse complement of EUB I	35%	[60]

Suffix "h" stands for helper oligonucleotide.
n.a.: not applicable.

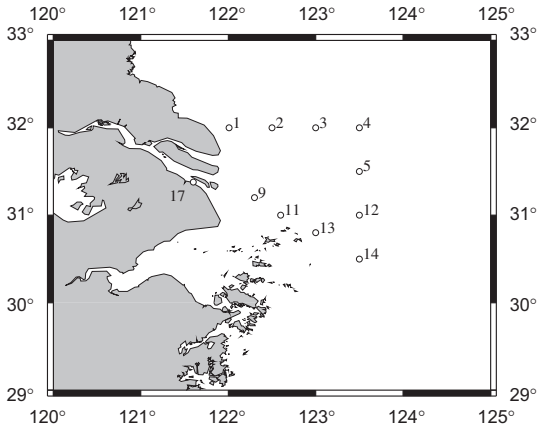

Fig. 1. Sampling stations off the Yangtze River estuary.

GenBank entries or from corresponding publications. Data included longitude, latitude, depth or altitude, and habitat information in categories such as marine water, hypersaline water or fresh water. The resulting table (SI Table 3) was the basis for creating the biogeographic map shown in Fig. 3. Results of the CARD-FISH with probe NOR5-730 and NOR5-1238 were also included in the map.

Sampling sites and procedures

Yangtze River estuary

On September 6–8, 2006, a small cruise was undertaken in the Yangtze River estuary (Fig. 1). Samples were taken from surface water, and immediately fixed with 1% paraformaldehyde (PFA) for 1 h, filtered onto polycarbonate filters (Millipore, 47 mm in diameter, 0.2 μm pore size) and stored frozen at −20 °C.

Namibian upwelling region

The cruise took place on March 22–23, 2003, along 23°S near Walvis Bay, from the coast into the Atlantic Ocean (14.4–12.0°E), through the Benguela Current. Surface water samples (10 m) from 13 stations, as well as three depth profiles, were collected and immediately fixed in 1% PFA for 1 h at room temperature or for 24 h at 4 °C. Subsequently, samples were kept frozen at −80 °C. For FISH, samples were carefully thawed, filtered onto polycarbonate filters (Millipore, 47 mm in diameter, 0.2 μm pore size) and further processed for FISH (see below).

Vision cruise

The Vision cruise was conducted in the period September 20–October 3, 2006. Sampling was carried out along the transect 30°W, from Iceland to the south of the Azores Islands, from surface waters (mostly at a depth of 10 m, SI Table 1). All water samples were fixed, filtered and stored as described for the Yangtze estuary.

German Bight

Samples were taken from a depth of 1 m at station "Kabeltonne", Helgoland (54.18°N, 7.90°E), German Bight, on seven separate days from May to July 2007 (see Table 2 for details) and from Cuxhaven in July 2007. The water samples from Helgoland were first pre-filtered at 10 μm to remove large particles and then fixed, filtered and stored as described for the Yangtze estuary. The water samples from Cuxhaven were treated identically but they were not pre-filtered.

North Sea sediment

The sediment samples were sampled from a sandy intertidal flat at Janssand (53.72°N, 7.68°E), German North Sea coast, in March and August 2007. Each time two adjacent cores were sampled for duplicates, and subsampled for each 1 cm range. The subsamples were fixed and sonicated, then the supernatant was filtered onto polycarbonate filters, as described previously [41,49].

Small sampling campaigns

A summary of all additional sampling stations is given in Table 2. Surface water samples from Xiamen, China were taken in September 2006 at the Xiamen ferry port, and in July 2007 from a sandy coast near Xiamen University. Other marine water samples were obtained from Southampton dock water, UK, and coastal water near Barcelona, Spain. Fresh water samples from the river Weser and freshwater ponds in Bremen were also checked for comparison. All the water samples were treated as described above for the Yangtze River sample. Other sediment samples were taken from intertidal sandy surface sediment from Sylt on the German North Sea coast. The samples were fixed, sonicated and processed as described above for North Sea sediment samples.

CARD-FISH

CARD-FISH was undertaken according to Pernthaler et al. [49] with the following modifications: agarose-embedded filters were permeabilized with 10 mg ml^{-1}

Table 2. Small sampling campaigns for determining the NOR5/OM60 distribution.

Location	Coordinates	NOR5/OM60 count	Method	Note[a]	Reference
Germany, North Sea, Helgoland, surface water	54.18°N 7.90°E	Up to 6–8% in early June and late July, 1998	FISH	NOR5-730, 30% FA	[19]
		6–7% in May, 11–13% in August, 2002	CARD-FISH	NOR5-730, 55% FA, 35 °C	[48]
		3–5%, unaffected in incubation after 0.8 μm pre-filtration, 2000	FISH	NOR5-730	[7]
		0.2% in February, 1.5–1.9% in summer not pre-filtered, 2007	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study
Germany, North Sea, Cuxhaven, surface water	53.887°N 8.641°E	0.9 ± 0.3% in July, 2007	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study
Spain, Blanes Bay, coastal surface water	41.67°N 2.80°E	0.6% in January, 1.3% in July, 2.6% in October, 2005	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study
		Detectable year round, low in winter, up to 5% in July, 2003–2004	CARD-FISH	NOR5-730, 50% FA, 35 °C	[3]
UK, Southampton dock water	50.9°N, 1.4°W	0.7%	CARD-FISH	NOR5-730, 30% FA	This study
China, Xiamen coastal surface water	24.450°N 118.074°E, 24.435°N 118.095°E	1.0–2.0% in summer, 2006 and 2007	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study
Pacific Ocean, Newport Hydroline, marine water, euphotic zone	44.65°N (124.42°W, 124.88°W, 125.60°W, 127.00°W)	3.4 ± 1.1%, only in euphotic zone	FISH	NOR5-730, 35% FA	[12]
Germany, North Sea, Sylt, intertidal sediment	55.04°N, 8.42°E	3% in 0–1 cm depth, 0.2% in 7–8 cm depth, 2007	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study
Germany, River Weser, in Bremen, fresh water	53.066°N 8.836°E	0.06%, in January 2008	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study
Germany, Bremen, MPI-Pond, fresh surface water	53.110°N 8.847°E	0.03%, in January 2008	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study
Germany, Bremen, Kuhgrabensee, salinity 2 psu, surface water	53.118°N, 8.852°E	< 0.03%, in February 2008	CARD-FISH	NOR5-730 + NOR5-1238 with 4 helpers, 50% FA	This study

[a]All CARD-FISH experiments were performed at 46 °C unless stated otherwise.

lysozyme for 20 min at 35 °C. Hybridization was performed at 46 °C for 3 h, and washing was carried out at 48 °C for 15 min. For the quantification of most of the members of the NOR5/OM60 groups, a combination of NOR5-730 and NOR5-1238 with all four helpers was used (Table 1). Signal amplification was carried out for 15 min with a fluorescein-labeled tyramide. All CARD-FISH preparations were counter-stained with DAPI. The relative abundance of hybridized cells was estimated as a ratio of hybridized cell counts to counts of DAPI-stained cells using epifluorescence microscopy. At least 500 DAPI-stained cells were counted. To check for autofluorescence or unspecific binding of the probe or tyramide, all samples were checked with the non-binding probe NON. The specificity of the NOR5-specific probes was checked with CARD-FISH on PFA-fixed cultures of *Congregibacter litoralis* KT71.

Results

Phylogeny

Based on an extensive comparison of trees obtained with various programs for phylogenetic reconstruction

Fig. 2. Consensus tree reconstructed based on almost full-length (>1400 nt) 16S rRNA sequences of members of the NOR5/OM60 clade. Underlined names are cultured isolates and subclades that include cultured isolates. The black and grey bars on the left of the branches show the clades that can be targeted by probes NOR5-730 and NOR5-1238, respectively, and the dashed lines for partly targeted subclades.

on more than 150 almost full-length NOR5/OM60 and closely related 16S rRNA gene sequences, a new consensus tree was calculated (Fig. 2). With all treeing methods, the NOR5/OM60 clade formed a monophyletic group within *Gammaproteobacteria*. In contrast to earlier trees based on less sequences [11,22] the current reconstruction of the NOR5/OM60 clade now also includes, besides the sequences from strain KT71 and clones OM60 and OM241, a cluster of freshwater clones and BD2-7, a clone retrieved from the deep sea. Another deep-sea sequence, BD1-7, was still excluded from the NOR5/OM60 clade. Sequence identities within NOR5/OM60 were typically >92%, whereas identities to outgroup sequences were usually below 92%. However, exceptions did occur (e.g. due to imperfect sequence quality) and therefore sequence identity alone was insufficient to include or exclude a new sequence from the NOR5/OM60 clade.

The exact branching within the NOR5/OM60 clade depended on the algorithms and filters used for reconstruction. There were stable subclades in which the same sequences always clustered together. However,

the relationship between the subclades was unstable. The largest subclade was labeled as NOR5-1, and it comprised more than one-third of all the available full-length sequences, as well as many partial sequences. NOR5-1 showed two stable subgroups. The largest was NOR5-1A with more than 90 full and partial sequences (50% of NOR5-1). NOR5-1C also seemed to be monophyletic, which did not apply to the other sequences of NOR5-1 (“NOR5-1B”), which included strain HTCC2080 and several North Sea strains, such as Ivo14. Another stable subclade was NOR5-4, which was the sister group of NOR5-1 in most of the trees.

Subclade NOR5-3 included the 16S rRNA sequence of “*Congregibacter litoralis*” KT71, as well as that of 17 other NOR5/OM60 strains which have all been recently isolated from the oxic layer of marine surface sediment of the German island Sylt. However, only a few environmental clone sequences fell into this subclade. Subclades NOR5-2 and NOR5-7 were close to NOR5-3 in most phylogenetic reconstructions. They currently comprise only a few sequences, including those of the NEP isolates obtained from Japanese marine coastal sediments [45], and an isolate from coastal marine water sampled off Banyuls-sur-Mer [1,2].

Subclades NOR5-5, NOR5-6, NOR5-8, NOR5-9 and NOR5-11 together contained one-fifth of all the NOR5/OM60 sequences. Subgroups NOR5-10 and NOR5-12 were deeply branched in most of the trees and they were dominated by sequences obtained from the deep sea. BD2-7 was the only full-length sequence of NOR5-12, and showed low identity (usually $<92\%$) with other NOR5/OM60 sequences.

Clone 114ds10 (AY212565) [57] was the only full sequence in the terrestrial subclade NOR5-13, which also included 13 partial sequences recovered from fresh water, fresh water sediment or soil. Five more partial sequences retrieved from fresh water studies (EF192914, EF192886, EF192904, AY214643 and AY214720) [18,40] did not group in the NOR5-13 subclade.

About 30% of NOR5/OM60 sequences could not be grouped into any of the above-mentioned subclades, and most of these were partial sequences. The representatives of each subclade and closely related outgroups are listed in SI Table 2 as hallmarks for categorizing further sequences. The full list of all NOR5/OM60 sequences retrieved in this study, either full-length or partial, is shown in SI Table 3.

Biogeography

The geographic information was compiled for all identifications of members of the NOR5/OM60 clade (Fig. 3). This included identification by isolation, 16S rRNA gene libraries, metagenomic studies, and by fluorescence *in situ* hybridization (FISH). So far, only 14 isolates have been reported from marine water or coastal marine sediment (SI Table 4). Here, we report 22 more NOR5/OM60 strains isolated from surface sediments of an intertidal sandflat from the North Sea island

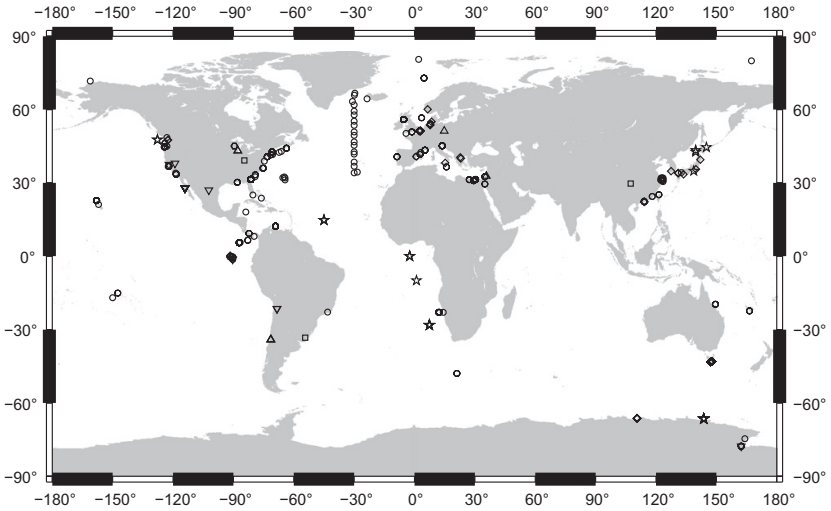

Fig. 3. Biogeography of the NOR5/OM60 clade. Sequence-, isolation-, and FISH-based identifications of NOR5/OM60 were marked on the world map. Signs refer to the habitat from where the sample was retrieved: marine water or other marine habitats – circle; marine coastal sediment – diamond; hypersaline – inverted triangle; soil – hexagon; fresh water – square; fresh sediment – triangle; deep sea – star. The map was created using the GMT (generic mapping tools) software package.

Table 3. Distribution of NOR5/OM60 subclades in different environments.

	Marine water and other marine habitats	Marine sediment	Hypersaline	Soil	Fresh water	Fresh sediment	Deep sea	Total
NOR5-1A	105	0	0	0	0	0	1	107
NOR5-1C	53	1	1	0	0	0	0	55
NOR5-1B	35	0	3	0	0	0	0	44
NOR5-4	21	0	0	0	0	0	0	23
NOR5-3	5	7	10	0	0	2	0	39
NOR5-2	2	2	0	0	0	0	0	4
NOR5-7	0	3	0	0	0	0	0	4
NOR5-5	13	7	2	2	0	0	1	30
NOR5-6	7	24	0	0	0	1	2	35
NOR5-8	12	8	0	0	0	0	0	20
NOR5-9	6	12	0	0	0	0	0	19
NOR5-11	1	6	0	0	0	0	0	11
NOR5-10	1	0	0	0	0	0	10	11
NOR5-12	0	1	0	0	0	0	2	3
NOR5-13	0	0	0	5	5	3	0	13
Other NOR5/OM60	91	56	7	1	0	1	14	179
Total	352	127	23	8	5	7	30	588

The numbers in the table give the number of 16S rRNA sequences retrieved from the public databases for a certain environment. The environmental conditions of several sequences could not be categorized; therefore the total number of a subclade can be higher than the sum of the listed numbers from different environments.

of Sylt, Germany, which belonged either to subclades NOR5-3 (17 strains) or to NOR5-1B (5 strains).

Table 3 lists the habitat preferences for each subclade according to source materials. The large subclades NOR5-1 and NOR5-4 appeared almost exclusively in the marine water column. Subclades NOR5-10 and NOR5-12 contained mainly identifications reported from deep-sea samples, and NOR5-13 was a freshwater clade. Sequences of the other NOR5/OM60 subclades were retrieved from marine sediment and the water column.

The NOR5/OM60 clade is cosmopolitan in the marine realm. Identifications have been reported from almost all oceans and at many coastal sites. In this respect, the American, European and East Asian coasts are particularly well covered with 16S rRNA gene libraries. There seems to be no latitudinal preference since NOR5/OM60 clones have been reported from mangrove [38,39] and coral reef [6,20,35], as well as sea-ice habitats [10].

NOR5/OM60 sequences were also reported in deep-sea sediments sampled near Antarctica [9] and Japan [4,28,37], as well as in the northeast Pacific [26] and the Atlantic (Schauer, unpublished). Additional reports on NOR5/OM60 sequences come from environments with different salinity: freshwater rivers [54,57], a rice paddy (DQ830363), freshwater sediments [44,61], activated sludge [32], soil [25,40], and also from hypersaline environments [24,36,52]. In addition, there are some sequences that cannot be placed on a world map since they are, for instance, from human plasma (clone NF37-A2; AY886614) [59]. Unlike the alphaproteobacterial RCA-1 cluster [55], it was still not possible in this study to detect biogeographic patterns for the various NOR5/OM60 subclades, neither latitudinal nor with respect to certain oceanic provinces.

NOR5/OM60 affiliated 16S rRNA genes in the GOS metagenomics dataset

The GOS dataset [52] contains 3728 16S rRNA gene sequences with lengths >300 nt. By comparative sequence analysis, 30 of these sequences (0.8%) could be unambiguously grouped within the NOR5/OM60 clade. Therein, 28 belonged to subclade NOR5-1, which is typical of marine surface water, and two belonged to its sister subclade NOR5-4. The sequences were found in 21 out of a total of 44 sampling stations. Except for station GS-04 (Atlantic, Canadian coast, salinity 28.3 psu), in which four out of a total of 31 16S rRNA gene sequences affiliated with the NOR5/OM60 clade, all the other libraries contained at most two NOR5/OM60 sequences. No sequences were found at low salinity stations (GS-12, salinity 3.5 psu, and GS-20,

freshwater). NOR5/OM60 16S rRNA sequences were clearly more frequent in coastal (20/1471; 1.4%) than in open-ocean samples (5/1516; 0.3%).

Design and optimization of probes for the quantification of the NOR5/OM60 clade

Our comprehensive collection of 16S rRNA sequences of the NOR5/OM60 clade facilitated a re-evaluation of oligonucleotide probes targeting this group. Probe NOR5-730 (Table 1) has been used in several studies for FISH-based quantification of members of the NOR5/OM60 clade [19]. This probe covered 131 out of 155 (84%) high-quality, almost full-length 16S rRNA sequences of the NOR5/OM60 clade included in the SILVA Ref dataset (Version 91) [50]. The probe design function of ARB [43] was used in an attempt to design new probes with an improved coverage of the NOR5/OM50 clade. However, it was not possible to design a single probe that perfectly matched all the NOR5/OM60 sequences without targeting outgroup sequences. The new probe NOR5-1238 (Table 1) targeted 46% of all high-quality NOR5/OM60 sequences, excluding the two major subclades NOR5-1 and NOR5-4. A combination of the probes NOR5-730 and NOR5-1238 increased the current coverage of the NOR5/OM60 clade to 92%, without any outgroup hits (Fig. 2). The combination failed to detect part of NOR5-1C, NOR5-2, NOR5-10, and all sequences in the NOR5-12 subclade.

Helper oligonucleotides were designed for all the above probes in an attempt to improve their hybridization efficiency (Table 1) [21]. Helpers are unlabeled oligonucleotides that bind in the vicinity of the probe, thereby presumably opening the secondary structure of the rRNA. The application of two helpers per probe significantly increased the intensity of monolabeled and CARD-FISH signals. Fixed cells of *Congregibacter litoralis* KT71 were used to determine the optimal formamide concentration for hybridization of probes NOR5-730 and NOR-1238 as 50%. These two probes were subsequently used at this formamide concentration in combination with helpers NOR5-659h, NOR5-709h, NOR5-1217h and NOR5-1287h for a specific and sensitive identification of members of the NOR5/OM60 clade in environmental samples.

Quantification of members of the NOR5/OM60 clade by CARD-FISH

The cells detected by CARD-FISH with the probe mixture NOR5-730/NOR5-1238 in marine plankton and benthos samples were pleomorphic (Fig. 4), often from coccoid to rod-shaped, sometimes were also bent in a vibrio shape. The length of the cells was between 0.5 and 3 μm, with a diameter between 0.5 and 1 μm. In plankton samples, single cells were mostly detected, suggesting that they were free-living. However, as described before [22] cells were also detected that were attached to microaggregates. In sediment samples, cells detected as NOR5/OM60 were also arranged in rosettes (Fig. 4f).

The optimized NOR5-730/NOR5-1238 probe/helper mixture was subsequently used for CARD-FISH-based quantifications in various marine samples. In the brackish to marine Yangtze River estuary (salinities of 22–32 psu), between 0% and 2.3% of all DAPI-stained cells were detected. Absolute numbers went up to 1.2×10^5 cells ml^{-1} (Table 4). No NOR5/OM60 cells were detected further up the Yangtze River at a freshwater reference site with 0.2 psu salinity.

Counts in surface waters obtained during an open-ocean North Atlantic transect in September 2006 (Vision cruise) were between 0.1% and 0.5% (3×10^3–1×10^4 cells ml^{-1}). NOR5/OM60 cells were present in all the samples, with no obvious trend from high to low latitude. The counts were only higher, at 0.9%, in a coastal sample taken during the same cruise off Iceland (SI Table 1).

In a transect in the Namibian coastal upwelling region along 23.00°S, the NOR5/OM60 counts at depths 10–15 m decreased with fluctuation from 3.0% (2.0×10^5 cells ml^{-1}) near the coast to 0.5% (1.3×10^4 cells ml^{-1}) in the open ocean (Fig. 5). Three depth profiles made at coastal (14.36°E), mid-shelf (13.15°E) and open-ocean stations (12.00°E) all clearly showed a steep decrease of the NOR5/OM60 abundances with depth. Below 70 m, the counts became marginal (<0.05%), and NOR5/OM60 cells were not detected below 300 m. With a number of 3.3×10^5 cells ml^{-1} (3.3% of DAPI) the counts were highest at the station closest to the coast (14.36°E) at a water depth of 5 m.

The highest relative abundance of NOR5/OM60 cells encountered in this study was recorded near the North Sea island Helgoland at station "Kabeltonne" (54.18°N 7.90°E). In surface water samples retrieved during May–July 2007, which were prefiltered through a 10 μm filter, members of the NOR5/OM60 clade ranged from 1.7% to 6.6% (8.2×10^3–1.2×10^5 cells ml^{-1}).

Counts were also high in sandy intertidal sediments taken at Janssand (53.72°N, 7.68°E), in the back-barrier region of the island of Spiekeroog close to the German North Sea coast. At this sample point, a range of 2.5–4.0% was detected in the top 3 cm of the sediment, and 1.4–3.1% at a depth of 3–12 cm (Fig. 6). Counts in March 2007 were generally lower than in August. The absolute number of NOR5/OM60 was in the order of

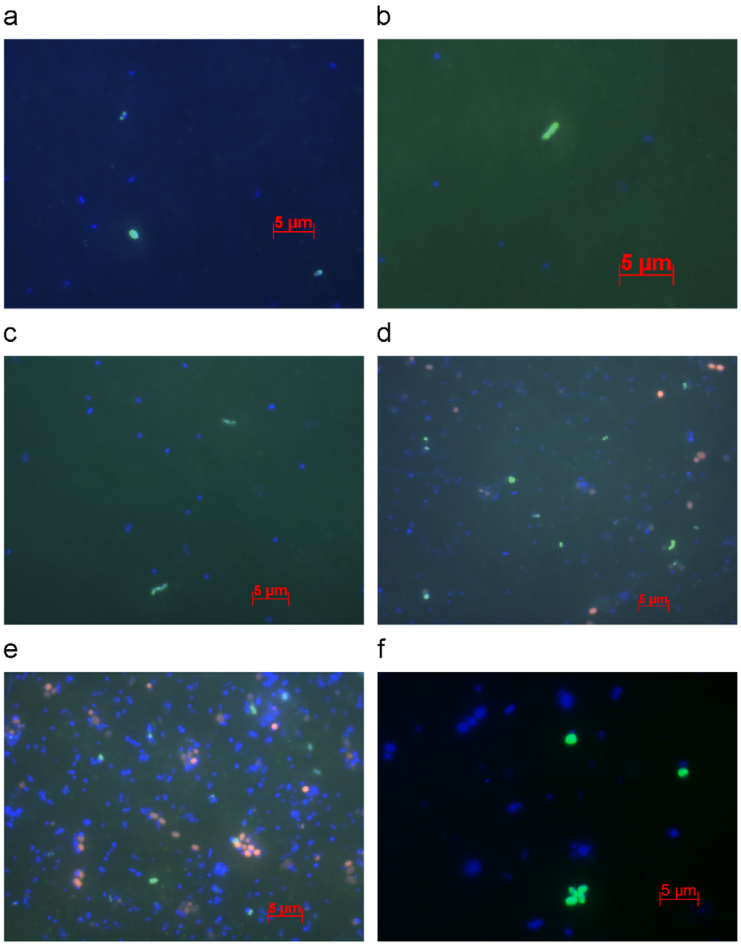

Fig. 4. Microscopic pictures of NOR5/OM60 cells in the environment: (a)–(c) North Atlantic open ocean, (d) and (e) from the Yangtze River estuary, (f) from Janssand sediment, North Sea. For all the pictures, blue: DAPI-stained DNA; green: fluorescein conferred probe signal for the NOR5/OM60 group; red: autofluorescence of cyanobacteria.

Table 4. NOR5/OM60 counts from the Yangtze River estuary cruise on September 6–8, 2006.

Station	Latitude (°N)	Longitude (°E)	Salinity (psu)	Temperature (°C)	NOR5/ OM60 count	DAPI cell count (10^6 cells ml^{-1})	Chlorophyll ($\mu g\,l^{-1}$)
CJ-1	32.00	122.01	25.1	27.8	0	n.d.	1.3
CJ-2	32.00	122.50	29.6	26.5	0.6±0.2%	4.1	0.4
CJ-3	32.00	123.00	31.2	25.2	1.6±0.3%	n.d.	2.2
CJ-4	32.00	123.50	31.2	25.6	0.9±0.2%	7.8	1.3
CJ-5	31.50	123.50	31.2	26.3	0.4±0.1%	10.8	0.9
CJ-12	31.00	123.50	31.8	28.5	0.10±0.05%	n.d.	0.5
CJ-14	30.50	123.50	32.3	28.0	0.5±0.1%	3.5	1.1
CJ-13	30.80	123.00	30.1	26.7	2.3±0.6%	4.1	3.9
CJ-11	31.00	122.60	28.1	27.1	1.4±0.3%	8.7	2.3
CJ-9	31.20	122.30	22.5	27.2	1.2±0.3%	3.7	1.0
CJ-17	31.38	121.60	0.2	28.9	0	n.d.	n.d.

n.d.: not determined.

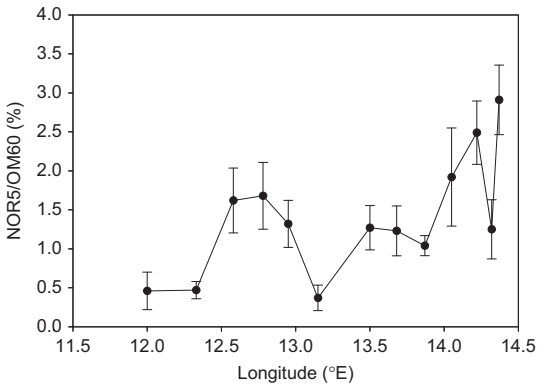

Fig. 5. NOR5/OM60 counts by CARD-FISH in the Namibian upwelling region on March 22–23, 2003, along 23°S near Walvis Bay. The counts were mostly at a depth of 10 m, with the exception of the two easternmost points that were at a depth of 15 m.

10^7 cells cm^{-3}, and in the surface sediments it was as high as 1.5×10^8 cells cm^{-3}.

Further quantifications of the NOR5/OM60 clade, either from this study or from former studies, are summarized in Table 3. A few preliminary quantifications of NOR5/OM60 in freshwater samples were also taken from Bremen, Germany. Abundances were less than 0.1% in the River Weser and from two ponds, one freshwater and the other with a salinity of 2 psu.

Discussion

The interest in the NOR5/OM60 clade has significantly increased with the discovery that this clade encompasses the elusive marine gammaproteobacterial branch of AAnPs [12,22]. The data obtained in this study clearly support the hypothesis of a cosmopolitan

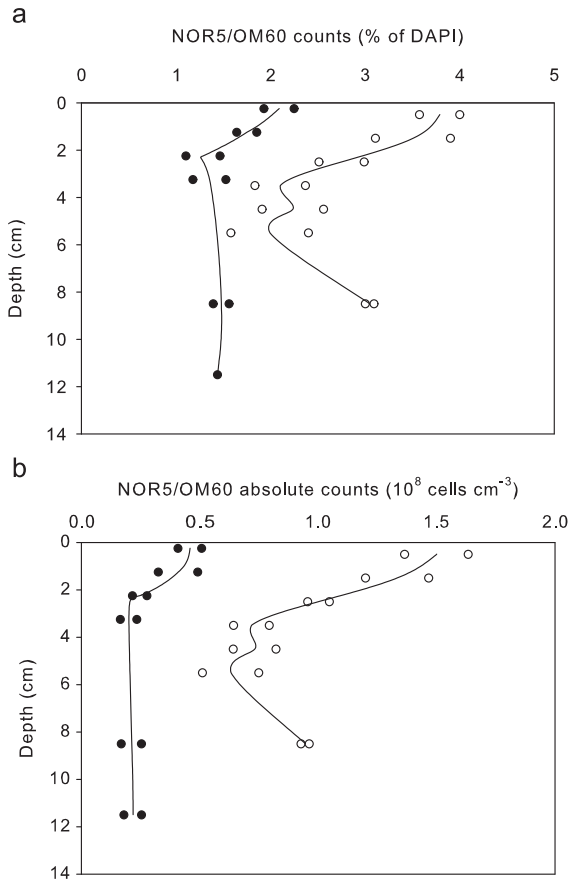

Fig. 6. NOR5/OM60 counts from Janssand sediment in March (black circles) and August (white circles) 2007. At each time point, two adjacent cores were sampled for duplication.

distribution of the NOR5/OM60 clade in marine surface waters. The presence of NOR5/OM60 is now confirmed for all oceans except the Indian Ocean, which has still not yet been examined in this regard. Several lines of evidence suggest that members of the NOR5/OM60 clade are generally more abundant in coastal areas than in open-ocean settings. CARD-FISH counts in coastal surface waters ($N = 30$) showed an average of $2.1 \pm 1.5\%$, whereas open-ocean surface water samples ($N = 36$) had an average of $0.5 \pm 0.4\%$. Fully independent support for this finding comes from the GOS dataset [52] in which the frequency of NOR5/OM60 16S rRNA gene sequences was significantly higher in coastal stations (1.4%) than in open-ocean stations (0.3%). Nevertheless, these data cannot be directly translated into cell frequencies since the number of rRNA operons per genome is quite different in marine bacteria [31]. The two fully sequenced strains of the NOR5/OM60 clade, "*Congregibacter litoralis*" KT71 and HTCC2080 contain two and one ribosomal operon, respectively. When the copy number of the NOR5/OM60 group members is less than the average, the frequency of 16S rRNA genes in the metagenomic library might underestimate the relative abundance of cells, and *vice versa*.

Our large CARD-FISH dataset also provides convincing support for a preference of members of the NOR5/OM60 clade for the euphotic zone, which had been reported for the coastal Pacific Newport Hydroline station [12]. Once again, support for this conspicuous depth distribution comes from metagenomics. In fosmid libraries constructed from bacterioplankton samples at Aloha Station, Hawaii, 16S rRNA genes of the NOR5/OM60 clade were detected at depths of 10 m and 70 m, but not at 130 m or deeper [17]. However, since the absence of NOR5/OM60 sequences is not valid proof of their absence *per se*, other methods, such as quantitative PCR, need to be applied to quantify NOR5/OM60 in these deep water layers.

There are indications for strong seasonal fluctuation of the NOR5/OM60 abundance in coastal waters. Three studies on North Sea surface water had reported NOR5/OM60 blooms up to 13% co-occurring with, for example, a dinoflagellate bloom [7,19,48]. Similar observations were recently reported for northwest Mediterranean coastal waters [3]. In our study, the samples from Xiamen, Barcelona and Helgoland showed the same trend, with high counts of NOR5/OM60 co-occurring with algal blooms. The photoheterotrophic members of the NOR5/OM60 clade benefit from algal photosynthesis, yet it is too early to speculate on a specific link to particular algal species. We hope that the more sensitive and specific probes and protocols developed in this study will facilitate more detailed investigations on the seasonality of NOR5/OM60 abundance in the future.

Besides season, water depth, and distance to the coast, we also searched for other parameters that might influence the distribution of NOR5/OM60 in the water column. Consequently, linear regression analysis was used to check for correlation between NOR5/OM60 abundance and other parameters. In the Namibian transect, NOR5/OM60 abundance was highly correlated to turbidity ($r^2 = 0.79$) (Fig. 7a). The NOR5/OM60 isolate *Congregibacter litoralis* KT71 is known to form aggregates in pure culture. Also, *in situ* attachment of NOR5/OM60 cells to aggregates has been demonstrated before [22]. NOR5/OM60 was positively correlated to chlorophyll fluorescence ($r^2 = 0.73$) (Fig. 7b). Similarly, in the surface waters of the Yangtze River estuary, the NOR5/OM60 abundance showed a strong positive correlation with chlorophyll concentration ($r^2 = 0.74$) (Fig. 7c). Algae are a source of fresh organic material, which in turn could serve as a substrate for NOR5/OM60. Indeed, the cultured strain KT71 prefers short oligomers and amino acids as substrates [22], which are generally rapidly consumed in the water column [63]. So, it might be advantageous to stay close to the site of production of these substrates. No other significant correlation of NOR5/OM60 abundance was detected with temperature, salinity or total bacterio-

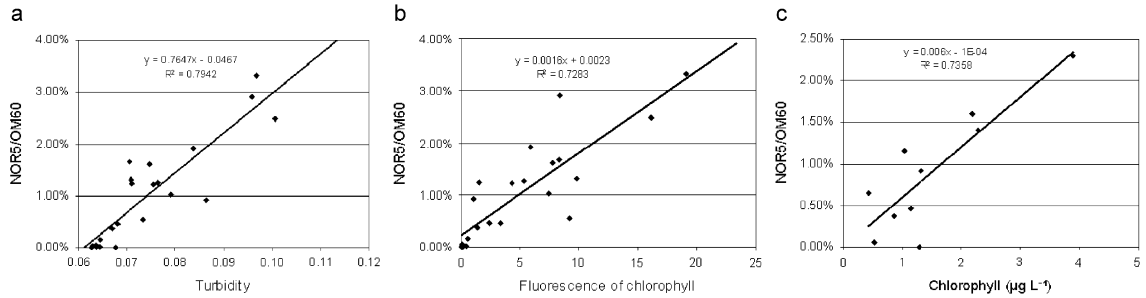

Fig. 7. Correlation of NOR5/OM60 to other environmental parameters: (a) turbidity in the Namibian transect, (b) chlorophyll fluorescence in the Namibian transect, (c) chlorophyll in the Yangtze River estuary.

plankton cell counts at either site. Also, during the North Atlantic Vision cruise, the NOR5/OM60 proportion in the surface water (10 m) showed no obvious correlation with any detected parameters (latitude, temperature, total DAPI cell count or chlorophyll fluorescence).

The NOR5/OM60 group and the AAnPs showed some common features for distribution, at least in some regions: they both occurred at a higher percentage in coastal water compared to the open ocean, and they were more abundant in summer or autumn than in winter or spring. Most of them appear in the euphotic zone in the marine water column and they are positively related to high chlorophyll concentrations [15,30,53,56,62]. It seems that many members of NOR5/OM60 may indeed be AAnPs. However, considering the rather high 16S rRNA sequence diversity within the NOR5/OM60 clade and its broad habitat range, it cannot be taken for granted that all members of NOR5/OM60 are AAnPs. The four strains isolated from marine surface waters KT71 (NOR5-3), HTCC2080 (NOR5-1B), HTCC2246 (not grouped) and HTCC2148 (NOR5-8) were shown to contain genes *puf*L and *puf*M coding for light-harvesting complex I (LHC I). However, *puf*L and *puf*M trees lacked congruence with the 16S rRNA phylogeny [12]. The *puf*M gene of HTCC2080 belongs to *puf*M Group K, the sister group of KT71, while the *puf*L and *puf*M genes of the HTCC2246 and HTCC2148 groups belong to different clusters of *Alphaproteobacteria* [62]. This suggests that the loss and gain of photosynthesis operons might be quite frequent in the NOR5/OM60 clade. For bacteria with a dominantly heterotrophic metabolism, photosynthesis might just be an accessory energy source that may not be required by all members of the NOR5/OM60 clade. Further studies that might combine the *in situ* identification of NOR5/OM60 cells by CARD-FISH with the direct identification of AAnP using infrared fluorescence microscopy [29,53] are therefore needed. Only then could we directly determine how many NOR5/OM60 are AAnPs, and, *vice versa*, how many of the AAnP are NOR5/OM60. In this respect, Yutin and colleagues [62] have recently reported on compositional changes within the AAnP between coastal and open-ocean sites. They have assessed the diversity of marine AAnPs based on comparative analysis of *puf*M sequences retrieved from the metagenomic libraries of the GOS [52]. Only six out of 85 scaffolds identified to include *puf*M in the oxic samples of the GOS dataset were highly similar to the *puf*M of *Congregibacter litoralis* KT71 and therefore likely to be from the NOR5/OM60 clade. These were from coastal samples, whereas the *puf*M of the alphaproteobacterial *Roseobacter*-type was found throughout the samples, and, additionally, as yet unidentified groups of AAnP seemed to dominate pelagic marine waters [62].

Acknowledgements

We thank Julia Arnds, Regina Schauer, Marc Mußmann, and Alaa Bakr for providing sequences. Pep Gasol, Paola Gomez and Ilaria Pizetti provided marine surface water samples from Barcelona, the North Atlantic (Vision cruise), and sampling station Kabeltonne off Helgoland, respectively. We acknowledge the staff of the State Key Laboratory at Xiamen University, especially Yao Zhang for organizing the cruise to the Yangtze River estuary, and Fan Zhang for providing chlorophyll data. The work was funded by the Max Planck Society, the German Academic Exchange Service (DAAD) and the China Scholarship Council (CSC).

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.syapm.2008.12.001.

References

[1] H. Agogué, E.O. Casamayor, M. Bourrain, I. Obernosterer, F. Joux, G.J. Herndl, P. Lebaron, A survey on bacteria inhabiting the sea surface microlayer of coastal ecosystems, FEMS Microbiol. Ecol. 54 (2005) 269–280.
[2] H. Agogué, F. Joux, I. Obernosterer, P. Lebaron, Resistance of marine bacterioneuston to solar radiation, Appl. Environ. Microbiol. 71 (2005) 5282–5289.
[3] L. Alonso-Sáez, V. Balagué, E.L. Sà, O. Sánchez, J.M. González, J. Pinhassi, R. Massana, J. Pernthaler, C. Pedrós-Alió, J.M. Gasol, Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH, FEMS Microbiol. Ecol. 60 (2007) 98–112.
[4] S. Arakawa, T. Sato, R. Sato, J. Zhang, T. Gamo, U. Tsunogai, A. Hirota, Y. Yoshida, R. Usami, F. Inagaki, C. Kato, Molecular phylogenetic and chemical analyses of the microbial mats in deep-sea cold seep sediments at the northeastern Japan Sea, Extremophiles 10 (2006) 311–319.
[5] K.E. Ashelford, N.A. Chuzhanova, J.C. Fry, A.J. Jones, A.J. Weightman, At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies, Appl. Environ. Microbiol. 71 (2005) 7724–7736.
[6] O. Barneah, E. Ben-Dov, E. Kramarsky-Winter, A. Kushmaro, Characterization of black band disease in Red Sea stony corals, Environ. Microbiol. 9 (2007) 1995–2006.
[7] C. Beardsley, J. Pernthaler, W. Wosniok, R. Amann, Are readily culturable bacteria in coastal North Sea waters suppressed by selective grazing mortality?, Appl. Environ. Microbiol. 69 (2003) 2624–2630.

[8] O. Béjà, M.T. Suzuki, J.F. Heidelberg, W.C. Nelson, C.M. Preston, T. Hamada, J.A. Eisen, C.M. Fraser, E.F. DeLong, Unsuspected diversity among marine aerobic anoxygenic phototrophs, Nature 415 (2002) 630–633.
[9] J.P. Bowman, R.D. McCuaig, Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment, Appl. Environ. Microbiol. 69 (2003) 2463–2483.
[10] R. Brinkmeyer, K. Knittel, J. Jurgens, H. Weyland, R. Amann, E. Helmke, Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice, Appl. Environ. Microbiol. 69 (2003) 6610–6619.
[11] J.C. Cho, S.J. Giovannoni, Cultivation and growth characteristics of a diverse group of oligotrophic marine *Gammaproteobacteria*, Appl. Environ. Microbiol. 70 (2004) 432–440.
[12] J.C. Cho, M.D. Stapels, R.M. Morris, K.L. Vergin, M.S. Schwalbach, S.A. Givan, D.F. Barofsky, S.J. Giovannoni, Polyphyletic photosynthetic reaction centre genes in oligotrophic marine *Gammaproteobacteria*, Environ. Microbiol. 9 (2007) 1456–1463.
[13] J.R. Cole, B. Chai, T.L. Marsh, R.J. Farris, Q. Wang, S.A. Kulam, S. Chandra, D.M. McGarrell, T.M. Schmidt, G.M. Garrity, J.M. Tiedje, The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy, Nucleic Acids Res. 31 (2003) 442–443.
[14] S.A. Connon, S.J. Giovannoni, High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates, Appl. Environ. Microbiol. 68 (2002) 3878–3885.
[15] M.T. Cottrell, A. Mannino, D.L. Kirchman, Aerobic anoxygenic phototrophic bacteria in the Mid-Atlantic Bight and the North Pacific Gyre, Appl. Environ. Microbiol. 72 (2006) 557–564.
[16] H. Daims, A. Bruhl, R. Amann, K.H. Schleifer, M. Wagner, 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 (1999) 434–444.
[17] E.F. DeLong, C.M. Preston, T. Mincer, V. Rich, S.J. Hallam, N.U. Frigaard, A. Martinez, M.B. Sullivan, R. Edwards, B.R. Brito, S.W. Chisholm, D.M. Karl, Community genomics among stratified microbial assemblages in the ocean's interior, Science 311 (2006) 496–503.
[18] C. Dorador, G. Castillo, K.P. Witzel, I. Vila, Bacterial diversity in the sediments of a temperate artificial lake, Rapel reservoir, Rev. Chil. Hist. Nat. 80 (2007) 213–224.
[19] H. Eilers, J. Pernthaler, J. Peplies, F.O. Glöckner, G. Gerdts, R. Amann, Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton, Appl. Environ. Microbiol. 67 (2001) 5134–5142.
[20] J. Frias-Lopez, A.L. Zerkle, G.T. Bonheyo, B.W. Fouke, Partitioning of bacterial communities between seawater and healthy black band diseased, and dead coral surfaces, Appl. Environ. Microbiol. 68 (2002) 2214–2228.
[21] B.M. Fuchs, F.O. Glöckner, J. Wulf, R. Amann, Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes, Appl. Environ. Microbiol. 66 (2000) 3603–3607.
[22] B.M. Fuchs, S. Spring, H. Teeling, C. Quast, J. Wulf, M. Schattenhofer, S. Yan, S. Ferriera, J. Johnson, F.O. Glöckner, R. Amann, Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis, Proc. Natl. Acad. Sci. USA 104 (2007) 2891–2896.
[23] B.M. Fuchs, G. Wallner, W. Beisker, I. Schwippl, W. Ludwig, R. Amann, Flow cytometric analysis of the in situ accessibility of *Escherichia coli* 16S rRNA for fluorescently labeled oligonucleotide probes, Appl. Environ. Microbiol. 64 (1998) 4973–4982.
[24] R.E. Glatz, P.W. Lepp, B.B. Ward, C.A. Francis, Planktonic microbial community composition across steep physical/chemical gradients in permanently ice-covered Lake Bonney, Antarctica, Geobiology 4 (2006) 53–67.
[25] M. Hartmann, F. Widmer, Community structure analyses are more sensitive to differences in soil bacterial communities than anonymous diversity indices, Appl. Environ. Microbiol. 72 (2006) 7804–7812.
[26] J.A. Huber, H.P. Johnson, D.A. Butterfield, J.A. Baross, Microbial life in ridge flank crustal fluids, Environ. Microbiol. 8 (2006) 88–99.
[27] J.P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees, Bioinformatics 17 (2001) 754–755.
[28] F. Inagaki, M. Suzuki, K. Takai, H. Oida, T. Sakamoto, K. Aoki, K.H. Nealson, K. Horikoshi, Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk, Appl. Environ. Microbiol. 69 (2003) 7224–7235.
[29] N.Z. Jiao, Y. Zhang, Y. Chen, Time series observation based InfraRed Epifluorescence Microscopic (TIREM) approach for accurate enumeration of bacteriochlorophyll-containing microbes in marine environments, J. Microbiol. Methods 65 (2006) 442–452.
[30] N.Z. Jiao, Y. Zhang, Y.H. Zeng, N. Hong, R.L. Liu, F. Chen, P.X. Wang, Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean, Environ. Microbiol. 9 (2007) 3091–3099.
[31] J.A. Klappenbach, J.M. Dunbar, T.M. Schmidt, RRNA operon copy number reflects ecological strategies of bacteria, Appl. Environ. Microbiol. 66 (2000) 1328–1333.
[32] A.N. Klein, D. Frigon, L. Raskin, Populations related to Alkanindiges, a novel genus containing obligate alkane degraders, are implicated in biological foaming in activated sludge systems, Environ. Microbiol. 9 (2007) 1898–1912.
[33] Z.S. Kolber, F.G. Plumley, A.S. Lang, J.T. Beatty, R.E. Blankenship, C.L. van Dover, C. Vetriani, M. Koblizek, C. Rathgeber, P.G. Falkowski, Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean, Science 292 (2001) 2492–2495.
[34] Z.S. Kolber, C.L. van Dover, R.A. Niederman, P.G. Falkowski, Bacterial photosynthesis in surface waters of the open ocean, Nature 407 (2000) 177–179.

[35] O. Koren, E. Rosenberg, Bacteria associated with mucus and tissues of the coral *Oculina patagonica* in summer and winter, Appl. Environ. Microbiol. 72 (2006) 5254–5259.
[36] R.E. Ley, J.K. Harris, J. Wilcox, J.R. Spear, S.R. Miller, B.M. Bebout, J.A. Maresca, D.A. Bryant, M.L. Sogin, N.R. Pace, Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat, Appl. Environ. Microbiol. 72 (2006) 3685–3695.
[37] L.N. Li, C. Kato, K. Horikoshi, Bacterial diversity in deep-sea sediments from different depths, Biodivers. Conserv. 8 (1999) 659–677.
[38] J.B. Liang, Y.Q. Chen, C.Y. Lan, N.F.Y. Tam, Q.J. Zan, L.N. Huang, Recovery of novel bacterial diversity from mangrove sediment, Mar. Biol. 150 (2007) 739–747.
[39] P.C. Liao, B.H. Huang, S. Huang, Microbial community composition of the Danshui river estuary of northern Taiwan and the practicality of the phylogenetic method in microbial barcoding, Microb. Ecol. 54 (2007) 497–507.
[40] M.R. Liles, B.F. Manske, S.B. Bintrim, J. Handelsman, R.M. Goodman, A census of rRNA genes and linked genomic sequences within a soil metagenomic library, Appl. Environ. Microbiol. 69 (2003) 2684–2691.
[41] E. Llobet-Brossa, R. Rosselló-Mora, R. Amann, Microbial community composition of Wadden Sea sediments as revealed by fluorescence in situ hybridization, Appl. Environ. Microbiol. 64 (1998) 2691–2696.
[42] W. Ludwig, O. Strunk, S. Klugbauer, N. Klugbauer, M. Weizenegger, J. Neumaier, M. Bachleitner, K.H. Schleifer, Bacterial phylogeny based on comparative sequence analysis, Electrophoresis 19 (1998) 554–568.
[43] W. Ludwig, O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A.W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, K.H. Schleifer, ARB: a software environment for sequence data, Nucleic Acids Res. 32 (2004) 1363–1371.
[44] B.J. MacGregor, S. Toze, E.W. Alm, R. Sharp, C.J. Ziemer, D.A. Stahl, Distribution and abundance of Gram-positive bacteria in the environment: development of a group-specific probe, J. Microbiol. Methods 44 (2001) 193–203.
[45] T. Maeda, K. Hayakawa, M. You, M. Sasaki, Y. Yamaji, M. Furushita, T. Shiba, Characteristics of nonylphenol polyethoxylate-degrading bacteria isolated from coastal sediments, Microbes Environ. 20 (2005) 253–257.
[46] W. Manz, R. Amann, W. Ludwig, M. Wagner, K.H. Schleifer, Phylogenetic oligodeoxynucleotide probes for the major subclasses of *Proteobacteria*-problems and solutions, Syst. Appl. Microbiol. 15 (1992) 593–600.
[47] T.D. Mullins, T.B. Britschgi, R.L. Krest, S.J. Giovannoni, Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities, Limnol. Oceanogr. 40 (1995) 148–158.
[48] A. Pernthaler, J. Pernthaler, Diurnal variation of cell proliferation in three bacterial taxa from coastal North Sea waters, Appl. Environ. Microbiol. 71 (2005) 4638–4644.
[49] A. Pernthaler, J. Pernthaler, R. Amann, Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria, Appl. Environ. Microbiol. 68 (2002) 3094–3101.
[50] E. Pruesse, C. Quast, K. Knittel, B.M. Fuchs, W. Ludwig, J. Peplies, F.O. Glöckner, SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB, Nucleic Acids Res. (2007).
[51] M.S. Rappé, P.F. Kemp, S.J. Giovannoni, Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina, Limnol. Oceanogr. 42 (1997) 811–826.
[52] D.B. Rusch, A.L. Halpern, G. Sutton, K.B. Heidelberg, S. Williamson, S. Yooseph, D.Y. Wu, J.A. Eisen, J.M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J.E. Venter, K. Li, S. Kravitz, J.F. Heidelberg, T. Utterback, Y.H. Rogers, L.I. Falcon, V. Souza, G. Bonilla-Rosso, L.E. Eguiarte, D.M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M.R. Ferrari, R.L. Strausberg, K. Nealson, R. Friedman, M. Frazier, J.C. Venter, The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific, PLoS Biol. 5 (2007) 398–431.
[53] M.S. Schwalbach, J.A. Fuhrman, Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR, Limnol. Oceanogr. 50 (2005) 620–628.
[54] H. Sekiguchi, M. Watanabe, T. Nakahara, B.H. Xu, H. Uchiyama, Succession of bacterial community structure along the Changjiang River determined by denaturing gradient gel electrophoresis and clone library analysis, Appl. Environ. Microbiol. 68 (2002) 5142–5150.
[55] N. Selje, M. Simon, T. Brinkhoff, A newly discovered *Roseobacter* cluster in temperate and polar oceans, Nature 427 (2004) 445–448.
[56] M.E. Sieracki, I.C. Gilg, E.C. Thier, N.J. Poulton, R. Goericke, Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic, Limnol. Oceanogr. 51 (2006) 38–46.
[57] J.M. Simpson, J.W.S. Domingo, D.J. Reasoner, Assessment of equine fecal contamination: the search for alternative bacterial source-tracking targets, FEMS Microbiol. Ecol. 47 (2004) 65–75.
[58] M.T. Suzuki, C.M. Preston, O. Béjà, J.R. de la Torre, G.F. Steward, E.F. DeLong, Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay, Microb. Ecol. 48 (2004) 473–488.
[59] S.D. Vernon, S.K. Shukla, J. Conradt, E.R. Unger, W.C. Reeves, Analysis of 16S rRNA gene sequences and circulating cell-free DNA from plasma of chronic fatigue syndrome and non-fatigued subjects, BMC Microbiol. 2 (2002).
[60] G. Wallner, R. Amann, W. Beisker, Optimizing fluorescent in situ hybridization with ribosomal-RNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms, Cytometry 14 (1993) 136–143.

[61] A. Wobus, C. Bleul, S. Maassen, C. Scheerer, M. Schuppler, E. Jacobs, I. Roske, Microbial diversity and functional characterization of sediments from reservoirs of different trophic state, FEMS Microbiol. Ecol. 46 (2003) 331–347.
[62] N. Yutin, M.T. Suzuki, H. Teeling, M. Weber, J.C. Venter, D.B. Rusch, O. Béjà, Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes, Environ. Microbiol. 9 (2007) 1464–1475.
[63] M.V. Zubkov, B.M. Fuchs, P.H. Burkill, R. Amann, Comparison of cellular and biomass specific activities of dominant bacterioplankton groups in stratified waters of the Celtic Sea, Appl. Environ. Microbiol. 67 (2001) 5210–5218.

SI Table 1 Sampling stations of Vision cruise; sampling was done at 10 m depth, except for Station 10 (20 m).

Station	Latitude (°N)	Longitude (°W)	NOR5/OM60 counts	DAPI cell count (10^6 cells mL^{-1})
1	64.385	23.667	0.9±0.5%	0.7
2	66.655	29.611	0.2±0.2%	2.5
3	65.877	29.942	0.4±0.1%	0.9
4	63.332	30.880	0.5±0.2%	2.1
5	61.911	30.005	0.3±0.1%	2.3
6	59.348	29.998	0.2±0.1%	2.3
7	57.860	30.002	0.1±0.05%	2.2
8	55.253	30.003	0.3±0.1%	2.1
9	53.558	30.002	0.3±0.2%	1.1
10	50.845	29.997	0.2±0.1%	2.4
11	49.439	30.002	0.2±0.1%	1.4
12	46.741	30.003	0.3±0.2%	0.8
13	45.363	30.002	0.5±0.1%	0.7
14	42.701	30.003	0.3±0.1%	1.1
15	41.793	29.998	0.4±0.2%	0.9
16	38.476	29.998	0.3±0.1%	1.1
17	36.777	29.999	0.5±0.1%	1.2
18	34.074	30.002	0.4±0.1%	1.1
19	34.415	28.482	0.3±0.1%	1.0

SI Table 2 Subclades of NOR5/OM60 clade and representative sequences.

Subclade	Representative sequences
NOR5-1A	KTc1119 (AF235120), SAR125 (L35466)
NOR5-1C	PC-PA11-51 (EF379800)
NOR5-1B	OM60 (U70696), HTCC2080 (AAVV01000015), Ivo14 (EU672849)
NOR5-2	3X/A02/235 (AY576769), ARK10038 (AF468388)
NOR5-3	KT71 (AAOA01000004), RAp1 red (EU672854)
NOR5-4	OM241 (U70702)
NOR5-5A	ELB16-085 (DQ015807)
NOR5-5B	ctg_NISAA05 (DQ396099)
NOR5-6	NEP68 (AB212806), Belgica2005/10-ZG-25 (DQ351816)
NOR5-7	NEP1 (AB212800)
NOR5-8	HTCC2148 (AY386334), Belgica2005/10-130-4 (DQ351753)
NOR5-9	Belgica2005/10-130-14 (DQ351759)
NOR5-10	Ka015 (AM997517), FS142-64B-02 (DQ513020)
NOR5-11	JSS S04 557 (EU707309)
NOR5-12	BD2-7 (AB015537)
NOR5-13	114ds10 (AY212565)
Not yet grouped NOR5/OM60	HTCC2246 (AY386337), Belgica2005/10-140-11 (DQ351790)
Closely related outgroup sequences	ctg_CGODA22 (DQ395580), ctg_CGOF078 (DQ395878), EUB3 (AY693815), BD1-7 (AB015519), HTCC2143 (AAVT01000006), NEP4 (AB212803), 51X/A01/131 (AY576729), SE12 (AY771747)

SI Table 3 Long list of public available NOR5/OM60 sequences in this study.

Accession number	Strain/clone	Clade	Length	Pub-lished	Longi-tude	Lati-tude	Environ-metal category [a]
AACY 01081804	IBEA_CTG_2122378	NOR5-1A	587	+			1
AACY 01758788	IBEA_CTG_UEALJ18TR	NOR5-1A	587	+			1
AAVV 01000015	HTCC2080	NOR5-1B	1533		-124.18	44.82	1
AB015537	BD2-7	NOR5-12	1497	+	138.65	34.92	7
AB031601	TIHP302-15 [b]	NOR5	652				2
AB031651	TIHP368-52	NOR5-8	794				2
AB041306	ECS36	NOR5-1A	582	+			1
AB041307	ECS37	NOR5-1A	467	+			1
AB041308	ECS38	NOR5-4	519	+			1
AB094860	OHKB4.92	NOR5-6	884	+	145.00	44.53	7
AB107215	CFPB-D1	NOR5-1B	389	+	132.52	34.22	1
AB116511	Y223	NOR5-6	999	+	141.97	39.46	2
AB212800	NEP1 [c]	NOR5-7	1456	+	131.00	34.00	2
AB212802	NEP3 [c]	NOR5-7	1455	+	131.00	34.00	2
AB212804	NEP17 [c]	NOR5-2	1456	+	131.00	34.00	2
AB212805	NEP65 [c]	NOR5-7	1456	+	133.60	33.50	2

AB212806	NEP68 [c]	NOR5-6	1456	+	139.80	35.60	2
AB238981	BJS72-032	NOR5-10	929	+	139.66	43.34	7
AB239061	BJS8R-025	NOR5-10	928	+	139.67	42.70	7
AB240711	BNT20-13	NOR5-10	931				7
AB266014	SIS03-24	NOR5-1A	586				1
AB266016	SIS03-26	NOR5-1A	586				1
AB294894	pItb-HW-12	NOR5-1A	1060				1
AB294932	pItb-vmat-15 [b]	NOR5	1478				1
AB294961	pItb-vmat-60	NOR5	1471				1
AB294978	pItb-vmat-87	NOR5	1441				1
AF058347	pM-54	NOR5-1C	497	+	-118.68	34.03	1
AF141510	CRE-PA25	NOR5	486	+	-123.94	46.25	1
AF141546	CRE-PA78	NOR5	483	+	-123.94	46.25	1
AF194404	kpj53f [b]	NOR5-1B	785	+	-73.23	40.71	1
AF195458	kpj432f [b]	NOR5-4	677	+	-73.23	40.71	1
AF195459	kpj282f	NOR5-1B	531	+	-73.23	40.71	1
AF195461	kpj02f	NOR5	771	+	-73.23	40.71	1
AF211276	MT33	NOR5	347	+	-122.10	37.50	3
AF235120	KTc1119	NOR5-1A	1530	+	7.87	54.15	1
AF268225	EBAC30	NOR5-1B	761	+	-122.40	36.70	1
AF327033	AWS98-7e	NOR5-5B	699	+	-161.13	71.63	1
AF354615	Arctic95D-6	NOR5-5B	617	+	167.34	79.88	1
AF365492	BM89MF5BA6	NOR5-1C	488				1
AF365559	BT49MF4BC5	NOR5	495		-82.22	9.35	1
AF365587	BT60MF2BA9	NOR5	496		-82.22	9.35	1
AF365597	BT60MF2BH10	NOR5	496		-82.22	9.35	1
AF367401	D069	NOR5	585	+	-79.11	33.33	1
AF367404	D130	NOR5-8	585	+	-79.11	33.33	1
AF369720	123	NOR5	678		127.44	34.83	2
AF424063	MERTZ_2CM_38	NOR5	996	+	143.64	-66.53	7
AF424096	MERTZ_2CM_175	NOR5	996	+	143.64	-66.53	7
AF424102	MERTZ_2CM_218 [b]	NOR5	995	+	143.64	-66.53	7
AF424105	MERTZ_2CM_227	NOR5	738	+	143.64	-66.53	7
AF424111	MERTZ_2CM_259	NOR5	996	+	143.64	-66.53	7
AF424117	MERTZ_2CM_277 [b]	NOR5-5A	996	+	143.64	-66.53	7
AF424129	MERTZ_2CM_325	NOR5-12	998	+	143.64	-66.53	7
AF424147	MERTZ_0CM_226	NOR5	996	+	143.64	-66.53	7
AF424157	MERTZ_0CM_369 [b]	NOR5	993	+	143.64	-66.53	7
AF429187	CR99-2-01	NOR5-13	600	+	107.40	29.72	5
AF441972	CD12A7	NOR5	472	+	-69.05	12.31	1
AF453551	ML516-J10	NOR5	700		-119.00	38.00	3
AF468280	ARKIA-12 [b]	NOR5	1398	+	2.00	80.52	1
AF468388	ARK10038	NOR5-2	1436	+			1
AJ252652	SC-I-73	NOR5-13	1028				4
AJ518497	Qui2P2-13	NOR5-13	500	+	14.76	51.28	6
AJ561160	RB_OIL 219	NOR5	474	+	164.12	-74.70	1
AJ627990	1D3	NOR5-3	673		0.83	40.74	1
AJ627998	1F10	NOR5-3	656		0.83	40.74	2
AJ633945	57ANG5	NOR5	1364				1

AJ633962	T63ANG230 [b]	NOR5	1354				1
AJ810625	MZ-26.ONR	NOR5-8	377	+	15.25	38.22	2
AM040145	Sylt 54	NOR5-6	890	+	8.43	55.03	2
AM086110	c5LKS13 [b]	NOR5	1455	+	35.60	32.80	6
AM176873	SZB37	NOR5-6	1455	+	114.08	22.52	1
AM229451	B1-6	NOR5-6	697		5.17	43.42	1
AM229460	B1-19	NOR5-3	1009		5.17	43.42	1
AM229480	B1-54	NOR5-6	1014		5.17	43.42	1
AM229486	B1-68	NOR5-6	735		5.17	43.42	1
AM259730	TAU-7-100	NOR5-1A	1509	+	13.73	45.13	1
AM259783	TAI-8-75	NOR5-1A	864	+	13.73	45.13	1
AM259784	TAI-8-65	NOR5-1A	833	+	13.73	45.13	1
AM259786	TAI-8-76	NOR5-1C	1370	+	13.73	45.13	1
AM259787	TAI-8-06	NOR5-1C	838	+	13.73	45.13	1
AM259789	TAI-8-07	NOR5-1C	816	+	13.73	45.13	1
AM259790	TAI-8-20	NOR5-1B	1388	+	13.73	45.13	1
AM691086	EG19	NOR5	1505				
AM990929	Nobaria 48	NOR5-1C	1492		29.42	30.95	1
AM990930	Nobaria 44	NOR5-1A	1374		29.42	30.95	1
AM990931	Nobaria 49	NOR5-4	1479		29.42	30.95	1
AM990932	Mersa Matrouh 28	NOR5-1A	1492		27.24	31.37	1
AM990933	Nobaria 46	NOR5-1C	1487		29.42	30.95	1
AM990934	Abou Qir East 157	NOR5-4	1492		30.35	31.44	1
AM990935	Mersa Matrouh 30	NOR5-1C	1492		27.24	31.37	1
AM990936	Abou Qir East 58	NOR5-4	1499		30.35	31.44	1
AM990937	Nobaria 47	NOR5-1C	1480		29.42	30.95	1
AM990938	Alexandaria Eastern harbour 89	NOR5-1A	1275		29.90	31.20	1
AM990939	Nobaria 50	NOR5-8	1475		29.42	30.95	1
AM990940	Mersa Matrouh 29	NOR5-3	1492		27.24	31.37	1
AM990941	Alexandaria Eastern harbour 140	NOR5-1C	1492		29.90	31.20	1
AM990942	Nobaria 45	NOR5-1B	1479		29.42	30.95	1
AM990943	Alexandaria Eastern harbour 122	NOR5-1C	1492		29.90	31.20	1
AM990944	Abou Qir East 156	NOR5-1C	1492		30.35	31.44	1
AM990945	Alexandaria Eastern harbour 70	NOR5-1A	1447		29.90	31.20	1
AM997345	An062	NOR5	1498		0.90	-9.93	7
AM997517	Ka015	NOR5-10	1494		7.35	-28.11	7
AM997625	Gu124	NOR5-10	1494		-2.42	0.00	7
AM997641	Ka140	NOR5	1496		7.35	-28.11	7
AM997729	Ka228	NOR5	1396		7.35	-28.11	7
AM997950	Gu222	NOR5	1498		-2.42	0.00	7
AM997955	Gu227	NOR5	1494		-2.42	0.00	7
AY012524	LMBA15	NOR5-13	640	+	-87.87	43.16	6
AY012525	LMBA15	NOR5-13	599	+	-87.87	43.16	6
AY038423	CD5A3	NOR5	355	+	-69.05	12.31	1
AY038560	CD4D7	NOR5	583	+	-69.05	12.31	1
AY038575	CD2C8	NOR5-1C	608	+	-69.05	12.31	1
AY095724	NL-111	NOR5-1A	450		166.41	-22.33	1
AY095735	NL-129	NOR5-1C	552		166.41	-22.33	1
AY095736	NL-130 [b]	NOR5-1C	559		166.41	-22.33	1

AY095889	pT-431	NOR5-1C	853		-150.00	-17.00	1
AY102017	HTCC160 [c]	NOR5-1B	672	+	-124.18	44.82	1
AY102018	HTCC227 [c]	NOR5-8	670	+	-124.18	44.82	1
AY102019	HTCC240 [c]	NOR5-8	646	+	-124.18	44.82	1
AY133405	BB2_187	NOR5-6	996	+	110.55	-66.26	2
AY133406	BB2_184	NOR5	995	+	110.55	-66.26	2
AY133407	OB3_172	NOR5	606	+	110.54	-66.30	2
AY133408	BB2_157	NOR5-8	994	+	110.55	-66.26	2
AY133409	BB2_126	NOR5	996	+	110.55	-66.26	2
AY133410	OB3_198	NOR5-5	995	+	110.54	-66.30	2
AY133411	OB3_138 [b]	NOR5	996	+	110.54	-66.30	2
AY133412	OB3_140	NOR5	511	+	110.54	-66.30	2
AY135664	ANT18/2_35	NOR5-5B	1434	+	21.00	-48.00	1
AY135666	ANT18/2_88	NOR5-5B	1446	+	21.00	-48.00	1
AY135673	ANT18/2_33	NOR5-5B	1449	+	21.00	-48.00	1
AY171332	s81	NOR5-5	1068				2
AY192999	BolB2	NOR5-9	519	+			2
AY193002	BolB7	NOR5-9	569	+			2
AY193031	BMSA2	NOR5-3	547	+			2
AY193236	Bol5	NOR5-6	512	+			2
AY193238	Bol19	NOR5	661	+			2
AY212565	114ds10	NOR5-13	1524	+	-84.52	39.15	5
AY214643	SL2-1-D1	NOR5-5	584	+	-89.52	45.08	4
AY214670	SL2-2-H2R	NOR5-13	353	+	-89.52	45.08	4
AY214720	SL2-2-A11F	NOR5-5	604	+	-89.52	45.08	4
AY216447	KM23	NOR5-6	1395				2
AY344391	K60-9	NOR5-9	626		-156.97	21.19	1
AY348860	4GB-70	NOR5-1C	724		149.50	-19.70	1
AY348861	4GB-61	NOR5	852		149.50	-19.70	1
AY348862	4GB-31	NOR5	672		149.50	-19.70	1
AY354853	PLY-P3-45	NOR5	942	+	-4.22	50.25	1
AY386331	HTCC2149 [c]	NOR5-1A	902	+	-124.41	44.65	1
AY386334	HTCC2148 [c]	NOR5-8	1350	+	-124.41	44.65	1
AY386336	HTCC2223 [c]	NOR5-1B	1352	+	-124.18	44.82	1
AY386337	HTCC2246 [c]	NOR5	1350	+	-124.18	44.82	1
AY386339	HTCC2080 [c]	NOR5-1B	1443	+	-124.18	44.82	1
AY458641	EBAC080-L32B05	NOR5-8	1525		-122.04	36.76	1
AY499456	RAN-36	NOR5-1A	602	+	-8.73	40.65	1
AY499457	RAN-17	NOR5-8	581	+	-8.73	40.65	1
AY499458	RAN-21	NOR5-1B	573	+	-8.73	40.65	1
AY499679	Dover89 [b]	NOR5	843	+	147.02	-43.32	2
AY499680	Dover96	NOR5-6	996	+	147.02	-43.32	2
AY499681	Dover270	NOR5-9	996	+	147.02	-43.32	2
AY499682	Dover29	NOR5	996	+	147.02	-43.32	2
AY499683	Dover366	NOR5	996	+	147.02	-43.32	2
AY499684	Dover372	NOR5-9	996	+	147.02	-43.32	2
AY499685	Dover401	NOR5-6	996	+	147.02	-43.32	2
AY499686	Dover91	NOR5-5	996	+	147.02	-43.32	2
AY499918	Nubeena69	NOR5	996	+	147.75	-43.10	2

AY499919	Nubeena293	NOR5-9	996	+	147.75	-43.10	2
AY499920	Nubeena46	NOR5	996	+	147.75	-43.10	2
AY499921	Nubeena227	NOR5-9	996	+	147.75	-43.10	2
AY499922	Nubeena378	NOR5	997	+	147.75	-43.10	2
AY499923	Nubeena54	NOR5	996	+	147.75	-43.10	2
AY499924	Nubeena27	NOR5	996	+	147.75	-43.10	2
AY499925	Nubeena256	NOR5-6	996	+	147.75	-43.10	2
AY499926	Nubeena301	NOR5	996	+	147.75	-43.10	2
AY499927	Nubeena60	NOR5-5	996	+	147.75	-43.10	2
AY499928	Nubeena63	NOR5	995	+	147.75	-43.10	2
AY499929	Nubeena392B	NOR5	613	+	147.75	-43.10	2
AY499930	Nubeena93	NOR5-8	996	+	147.75	-43.10	2
AY499957	Nubeena239 [b]	NOR5	997	+	147.75	-43.10	2
AY499970	Nubeena366	NOR5-12	657	+	147.75	-43.10	2
AY515463	GWS-K10	NOR5-3	724	+	7.72	53.71	2
AY515464	GWS-K15	NOR5-6	718	+	7.72	53.71	2
AY515468	GWS-Kdna22	NOR5	898	+	7.72	53.71	2
AY528767	CHF20	NOR5-13	584	+			5
AY533969	Therm01-55	NOR5	492	+	22.75	40.50	2
AY533983	Therm30-C09	NOR5	723	+	23.00	40.00	2
AY568767	JH10_C07	NOR5-1B	851				
AY568769	JH10_C11	NOR5-1B	851				
AY568791	JH10_C34	NOR5-7	765				
AY568820	JH10_C68	NOR5	982				
AY568854	JH12_C12	NOR5	989				
AY568920	JH12_C84	NOR5-9	608				
AY576769	3X/A02/235 [c]	NOR5-2	1485	+	3.14	42.49	1
AY580744	PI_4j5b	NOR5-1A	856	+	-70.79	42.71	1
AY580745	PI_4z10a	NOR5-1B	847	+	-70.79	42.71	1
AY580746	PI_4r10b	NOR5-1A	848	+	-70.79	42.71	1
AY580747	PI_RT273	NOR5-1A	868	+	-70.79	42.71	1
AY580748	PI_4t6h	NOR5-1B	761	+	-70.79	42.71	1
AY580750	PI_4d11d	NOR5-1B	820	+	-70.79	42.71	1
AY580752	PI_RT139	NOR5-1B	449	+	-70.79	42.71	1
AY580753	PI_4m3d	NOR5-1C	681	+	-70.79	42.71	1
AY604941	Calaveras29.17	NOR5-1C	684	+	-102.19	27.08	3
AY627376	EB000-37F04	NOR5-1A	1460	+	-122.04	36.76	1
AY627377	EB000-39D04	NOR5-1B	1469	+	-122.04	36.76	1
AY628669	LS-D9	NOR5-1B	654	+	-123.00	45.00	1
AY661626	3121724	NOR5	464	+			1
AY661629	6121724	NOR5	465	+			1
AY663906	JL-ECS-C34	NOR5-1A	1228				1
AY664075	JL-WNPG-U14 [b]	NOR5-1A	1083				1
AY665403	Sorfjord-S1-22	NOR5	626	+	6.53	60.08	2
AY678489	R8C-B11	NOR5	621	+			2
AY701454	GCHU11_C	NOR5	1452				1
AY710779	SIMO-1339	NOR5	349		-81.28	31.39	1
AY710781	SIMO-1341	NOR5	329		-81.28	31.39	1
AY710819	SIMO-1379	NOR5	583		-81.28	31.39	1

AY710823	SIMO-1383	NOR5	540		-81.28	31.39	1
AY711116	SIMO-1750	NOR5	711		-81.28	31.39	1
AY711143	SIMO-1777	NOR5	703		-81.28	31.39	1
AY711157	SIMO-1791	NOR5	731		-81.28	31.39	1
AY711190	SIMO-1824	NOR5	773		-81.28	31.39	1
AY711305	SIMO-1939	NOR5	652		-81.28	31.39	1
AY711323	SIMO-1957	NOR5	603		-81.28	31.39	1
AY711344	SIMO-1978	NOR5	775		-81.28	31.39	1
AY711364	SIMO-1998	NOR5	714		-81.28	31.39	1
AY711387	SIMO-2021	NOR5	824		-81.28	31.39	1
AY711414	SIMO-2048	NOR5	822		-81.28	31.39	1
AY711434	SIMO-2068	NOR5	767		-81.28	31.39	1
AY711437	SIMO-2071	NOR5	720		-81.28	31.39	1
AY711438	SIMO-2072	NOR5	615		-81.28	31.39	1
AY711445	SIMO-2079	NOR5	612		-81.28	31.39	1
AY711480	SIMO-2114	NOR5	732		-81.28	31.39	1
AY711488	SIMO-2122	NOR5	563		-81.28	31.39	1
AY711491	SIMO-2125	NOR5-3	629		-81.28	31.39	1
AY711507	SIMO-2141	NOR5	821		-81.28	31.39	1
AY711559	SIMO-2193	NOR5	771		-81.28	31.39	1
AY711697	SIMO-2331	NOR5-6	664		-81.28	31.39	1
AY711727	SIMO-2361	NOR5-9	576		-81.28	31.39	1
AY711829	SIMO-2463	NOR5	797		-81.28	31.39	1
AY711846	SIMO-2480	NOR5-6	716		-81.28	31.39	1
AY711847	SIMO-2481	NOR5-9	645		-81.28	31.39	1
AY711850	SIMO-2484	NOR5	665		-81.28	31.39	1
AY711871	SIMO-2505	NOR5	785		-81.28	31.39	1
AY711946	SIMO-409	NOR5	669		-81.28	31.39	1
AY712023	SIMO-486	NOR5-1A	596		-81.28	31.39	1
AY712044	SIMO-507	NOR5	609		-81.28	31.39	1
AY712088	SIMO-551	NOR5-1C	540		-81.28	31.39	1
AY712149	SIMO-612	NOR5-1A	540		-81.28	31.39	1
AY712159	SIMO-622	NOR5	540		-81.28	31.39	1
AY712207	SIMO-670	NOR5	369		-81.28	31.39	1
AY712277	SIMO-740	NOR5-1B	659		-81.28	31.39	1
AY712309	SIMO-772	NOR5-1B	483		-81.28	31.39	1
AY712315	SIMO-778	NOR5-1A	605		-81.28	31.39	1
AY712333	SIMO-796	NOR5-1B	573		-81.28	31.39	1
AY712348	SIMO-811	NOR5-9	664		-81.28	31.39	1
AY712376	SIMO-839	NOR5-4	664		-81.28	31.39	1
AY712448	SIMO-1003	NOR5-9	622		-81.28	31.39	1
AY712471	SIMO-1026	NOR5	573		-81.28	31.39	1
AY712474	SIMO-1029	NOR5-1A	416		-81.28	31.39	1
AY770727	PlS140	NOR5	804				
AY793415	GREV_244_13	NOR5-6	562				2
AY793416	GREV_244_11	NOR5-6	562				2
AY793417	GREV_255_6	NOR5-2	562				2
AY793421	GREV_255_13	NOR5-3	562				2
AY794196	F4C23S	NOR5-5	704				

AY822374	C16T9 G05	NOR5	738	+			2
AY822375	C16T9 G08	NOR5	738	+			2
AY822376	C16T9 G11	NOR5	748	+			2
AY822385	C16T12 B02	NOR5-6	725	+			2
AY822390	C5T12 B07	NOR5	734	+			2
AY822393	C5T9 G06	NOR5-6	762	+			2
AY822404	C16T12 F06	NOR5-6	562	+			2
AY830028	CONP53	NOR5-1A	655		-5.39	55.96	1
AY830029	CONW73	NOR5-1A	656		-5.39	55.96	1
AY830030	FFP89	NOR5-1A	666		-5.39	55.96	1
AY830031	FFW06	NOR5-1A	666		-5.39	55.96	1
AY830032	FFP41	NOR5-1B	838		-5.39	55.96	1
AY830033	FFP93	NOR5-1B	613		-5.39	55.96	1
AY830046	CONP07	NOR5-5B	772		-5.39	55.96	1
AY830047	CONP82	NOR5-1B	528		-5.39	55.96	1
AY830049	CONW02	NOR5-1A	667		-5.39	55.96	1
AY830059	FFW62	NOR5-1A	673		-5.39	55.96	1
AY830060	FFW70	NOR5-1C	623		-5.39	55.96	1
AY867907	G50-0019	NOR5	459	+	4.72	72.90	1
AY867934	G50-0218	NOR5	457	+	4.72	72.90	1
AY867937	G50-0300	NOR5	457	+	4.72	72.90	1
AY868086	I50-0410	NOR5	457	+	15.83	36.50	1
AY868097	I50-0440	NOR5-1A	443	+	15.83	36.50	1
AY886614	NF37-A2	NOR5-5	349				
AY897345	nubeena110	NOR5	564	+	147.75	-43.10	2
AY897346	DOVER299B	NOR5	545	+	147.02	-43.32	2
AY904496	BG27-17	NOR5-1A	471	+	-88.08	30.25	1
AY904514	BG29-16	NOR5	486	+	-88.08	30.25	1
AY904518	BG29-20	NOR5-1C	504	+	-88.08	30.25	1
DQ009132	SPOTSAPR01_5m22	NOR5-1C	1507	+	-118.40	33.55	1
DQ009133	SPOTSAPR01_5m204	NOR5-1C	1507	+	-118.40	33.55	1
DQ009134	SPOTSAPR01_5m162	NOR5-1C	1508	+	-118.40	33.55	1
DQ009135	SPOTSAPR01_5m185	NOR5-1C	1507	+	-118.40	33.55	1
DQ009137	SPOTSOCT00_5m87	NOR5-1A	1506	+	-118.40	33.55	1
DQ015807	ELB16-085	NOR5-5A	1495	+	162.33	-77.72	1
DQ015821	ELB19-149	NOR5	1498	+	162.33	-77.72	1
DQ015829	ELB19-198	NOR5-5A	1496	+	162.33	-77.72	1
DQ015838	ELB25-204	NOR5-5A	1495	+	162.33	-77.72	3
DQ015840	WLB13-127	NOR5-5A	1496	+	162.33	-77.72	1
DQ015860	WLB16-180	NOR5-5A	1489	+	162.33	-77.72	1
DQ071054	Chl1.41	NOR5-1A	1455	+			1
DQ071082		NOR5	1456	+	-122.98	48.56	1
DQ167034	g142	NOR5	643				3
DQ167094	h34	NOR5-5	695				3
DQ187798	115	NOR5-1A	948				1
DQ189854	SIMO-2879	NOR5-1C	480	+	-81.27	31.39	1
DQ189890	SIMO-2915	NOR5-4	400	+	-81.27	31.39	1
DQ189947	SIMO-2972	NOR5-4	741	+	-81.27	31.39	1
DQ200552	CD205F07	NOR5	956		-69.05	12.31	1

DQ200563	CD207A06	NOR5	920		-69.05	12.31	1
DQ200569	CD207A12	NOR5	960		-69.05	12.31	1
DQ200593	CD207C12	NOR5	993		-69.05	12.31	1
DQ200612	CD207E08	NOR5	938		-69.05	12.31	1
DQ200616	CD207E12	NOR5	942		-69.05	12.31	1
DQ200618	CD207F02	NOR5	997		-69.05	12.31	1
DQ234113	DS029	NOR5-4	1535	+	121.45	25.16	1
DQ234158	DS074 [b]	NOR5-4	1532	+	121.45	25.16	1
DQ252467	JRCKIII11 [b]	NOR5	921				
DQ256675	Fitz2_15	NOR5	996				2
DQ256676	Flyn25a	NOR5-3	998				2
DQ256708	Fitz2_21	NOR5-9	990				2
DQ256709	Flyn2_29	NOR5	989				2
DQ263710	GA456	NOR5	1358				
DQ295347	PA-A18	NOR5	486		31.20	121.95	2
DQ295348	PA-A19	NOR5	491		31.20	121.95	2
DQ295349	PA-A20	NOR5	486		31.20	121.95	2
DQ295388	PA-B49	NOR5-9	488		31.20	121.95	2
DQ295922	SIMO-3977	NOR5-4	735	+	-81.27	31.39	1
DQ295935	SIMO-3990 [b]	NOR5-1B	749	+	-81.27	31.39	1
DQ300630	HF10_D4_P1	NOR5-1A	768	+	-158.00	22.75	1
DQ300656	HF10_G2_P1	NOR5-1A	906	+	-158.00	22.75	1
DQ300662	HF10_G10_P1	NOR5-1A	911	+	-158.00	22.75	1
DQ300664	HF10_G12_P1	NOR5-1A	868	+	-158.00	22.75	1
DQ300842	HF70_H5_P2	NOR5-1A	881	+	-158.00	22.75	1
DQ300853	HF70_A5_P2	NOR5-1A	880	+	-158.00	22.75	1
DQ300855	HF70_G7_P2	NOR5-1A	831	+	-158.00	22.75	1
DQ300872	HF70_B10_P2	NOR5-1A	861	+	-158.00	22.75	1
DQ312245	OTU 12 [b]	NOR5	1278				1
DQ330798	02D2Z51	NOR5-3	1370	+	-114.03	27.89	3
DQ330814	01D2Y33	NOR5-3	868	+	-114.03	27.89	3
DQ330815	01D2Y37	NOR5-1B	869	+	-114.03	27.89	3
DQ330816	01D2Y89	NOR5-1B	867	+	-114.03	27.89	3
DQ330829	02D2Y91	NOR5	881	+	-114.03	27.89	3
DQ330830	02D2Y55	NOR5-3	865	+	-114.03	27.89	3
DQ330831	02D2Y54	NOR5-3	848	+	-114.03	27.89	3
DQ330841	01D210B	NOR5	1104	+	-114.03	27.89	3
DQ330860	02D26B	NOR5	1093	+	-114.03	27.89	3
DQ330879	01D1Y95	NOR5-3	938	+	-114.03	27.89	3
DQ330916	02N1A250	NOR5	661	+	-114.03	27.89	3
DQ330925	01D1Za55	NOR5-3	857	+	-114.03	27.89	3
DQ330927	01D1Za83	NOR5-3	904	+	-114.03	27.89	3
DQ330945	05D2Z66	NOR5-3	1368	+	-114.03	27.89	3
DQ334654	IF-14-13	NOR5-6	1288				2
DQ334663	IF-50-4	NOR5-11	1293				2
DQ334665	IF-60-4	NOR5-11	1303				2
DQ351746	Belgica2005/10-120-14	NOR5	1492		2.70	51.19	2
DQ351747	Belgica2005/10-120-16	NOR5	1493		2.70	51.19	2
DQ351753	Belgica2005/10-130-4	NOR5-8	1492		2.90	51.27	2

DQ351759	Belgica2005/10-130-14	NOR5-9	1492		2.90	51.27	2
DQ351785	Belgica2005/10-140-17	NOR5-6	1492		3.05	51.33	2
DQ351790	Belgica2005/10-140-11	NOR5	1492		3.05	51.33	2
DQ351807	Belgica2005/10-ZG-13	NOR5-6	1490		2.33	51.26	2
DQ351816	Belgica2005/10-ZG-25	NOR5-6	1492		2.33	51.26	2
DQ372852	NH10_44 [b]	NOR5-1A	1487	+	-124.18	44.82	1
DQ374218	Souza16s835	NOR5-1A	944				
DQ394897	VHS-B1-16	NOR5-5A	1495		114.17	22.29	2
DQ394949	VHS-B3-55	NOR5	1511		114.17	22.29	2
DQ396099	ctg_NISAA05	NOR5-5B	1526				1
DQ396306	ctg_NISA244 [b]	NOR5	1550				1
DQ416247	s2ua29	NOR5-9	560	+	34.88	32.48	1
DQ416381	w1ub48	NOR5-8	786	+	34.88	32.48	1
DQ416493	w3ub20	NOR5	560	+	34.88	32.48	1
DQ417915	RS.Water.259	NOR5-4	520				1
DQ421605	SIMO-4240	NOR5-4	883	+	-81.27	31.39	1
DQ421665	SIMO-4300	NOR5-4	851	+	-81.27	31.39	1
DQ421720	SIMO-4355 [b]	NOR5-4	857	+	-81.27	31.39	1
DQ421756	SIMO-4391	NOR5-1B	900	+	-81.27	31.39	1
DQ421776	SIMO-4411	NOR5-1B	828	+	-81.27	31.39	1
DQ424125	NE34G09cA	NOR5	706				
DQ424264	NE36F09cA	NOR5	748				
DQ424381	NE40C07cA	NOR5	800				
DQ424384	NE40C12cA	NOR5	822				
DQ424400	NE40F04cA [b]	NOR5	804				
DQ424570	NE42H06cA	NOR5	751				
DQ436638	T31_3	NOR5-1B	638				1
DQ436639	T31_10	NOR5-1A	793				1
DQ436644	T31_21	NOR5-1B	780				1
DQ436649	T31_60	NOR5-1C	789				1
DQ436650	T31_73	NOR5-1A	788				1
DQ436653	T31_124	NOR5-1B	687				1
DQ436659	T31_176	NOR5-8	706				1
DQ436660	T31_179	NOR5-1A	638				1
DQ436666	T32_45	NOR5-1A	774				1
DQ436667	T32_48	NOR5-1C	790				1
DQ436669	T32_61	NOR5-1C	776				1
DQ436678	T32_198	NOR5-1C	784				1
DQ438435	ECS-P7-C41	NOR5-1A	985				1
DQ446111	BBD_216_18	NOR5-1C	687	+	-76.09	23.77	1
DQ473628	302T-9	NOR5	480				1
DQ473630	303T-2	NOR5	480				1
DQ513020	FS142-64B-02	NOR5-10	1425	+	-127.78	47.70	7
DQ513021	FS142-8B-02	NOR5-1A	1496	+	-127.78	47.70	7
DQ517273	SE1205	NOR5	606				1
DQ661786	DGGE gel band B20-1	NOR5	592	+	-80.39	25.03	1
DQ778149	BL03-WIN12	NOR5-1A	895	+	2.80	41.67	1
DQ778267	BL03-AUT48	NOR5-1A	839	+	2.80	41.67	1
DQ778269	BL03-AUT94	NOR5-1C	844	+	2.80	41.67	1

DQ810304	ESP10-K27I-2 [b]	NOR5-1C	813				1
DQ810313	ESP10-K27I-3	NOR5-1A	707				1
DQ810315	ESP10-K27I-4	NOR5-1A	775				1
DQ810378	ESP10-K9II-54	NOR5-1A	1236				1
DQ810384	ESP10-K9II-64	NOR5-1C	957				1
DQ828907	DOK_CONFYM_clone693	NOR5-13	515	+			4
DQ829007	DOK_CONFYM_clone807	NOR5-13	515	+			4
DQ830363	CON5_D02	NOR5-13	735		-54.33	-33.33	5
DQ839256	NS11	NOR5-1A	565		3.45	56.65	1
DQ839257	NS12	NOR5-1A	560		3.45	56.65	1
DQ880947	CB_059	NOR5-1C	477		-81.27	31.39	1
DQ881017	CT_041	NOR5-1A	476		-81.27	31.39	1
DQ881089	DB_037	NOR5-1A	477		-81.27	31.39	1
DQ881143	DB_099	NOR5-1A	477		-81.27	31.39	1
DQ881366	VT_061	NOR5-1A	477		-81.27	31.39	1
DQ881434	CB_047	NOR5-1A	477		-81.27	31.39	1
DQ889883	EC179	NOR5	1493				1
DQ906720	AntCL1D1	NOR5-5B	1508				1
EF019499	Elev_16S_1001 [b]	NOR5	1395				4
EF061963	XME57	NOR5	1496				2
EF092655	b1pl1H05	NOR5-4	722		-43.15	-22.85	1
EF111188	RBE2CI-108	NOR5-13	1279				5
EF125404	MSB-1E11	NOR5	1495				2
EF125458	MSB-2G4	NOR5-9	1495				
EF137399	Bv39	NOR5	615	+	-70.55	41.56	1
EF142025	KF047	NOR5-13	685				4
EF192886	Rap2_4A	NOR5-3	803	+	-71.41	-34.14	6
EF192904	Rap2_23C	NOR5-6	1092	+	-71.41	-34.14	6
EF192914	RAp1_9C	NOR5-3	372	+	-71.47	-34.13	6
EF206850	c1uc14	NOR5-1C	701				1
EF207031	b3ub20	NOR5	643				1
EF207107	b3cb20 [c]	NOR5	686				1
EF215764	PV1-35	NOR5	829				1
EF215776	PV2-29	NOR5	851				1
EF221653	DGGE gel band B15	NOR5-1B	549	+			1
EF221655	DGGE gel band B17	NOR5-1A	549	+			1
EF379446	TLC-PA3-26	NOR5-8	598	+	114.27	22.25	1
EF379702	VH-FL8-8	NOR5-4	1508	+	114.19	22.29	1
EF379800	PC-PA11-51	NOR5-1C	1517	+	114.05	22.31	1
EF379865	PC-FL10-45	NOR5-1B	1492	+	114.05	22.31	1
EF433163	BBS16S-8	NOR5	597	+	34.94	29.51	1
EF433169	BBS16S-14	NOR5	600	+	34.94	29.51	1
EF459868	140b1	NOR5-5	910				2
EF459991	313c2	NOR5-6	910				2
EF460022	282c2	NOR5-6	910				2
EF491301	S1-40	NOR5-11	833				1
EF491408	S3-25	NOR5-8	909				1
EF491411	S3-29	NOR5-1B	922				1
EF491419	S3-37	NOR5-6	928				1

EF491420	S3-38	NOR5	851				1
EF572388	S23_487	NOR5-1A	1498		-86.57	5.64	1
EF573742	S25_86	NOR5-1A	1498		-87.09	5.55	1
EF573771	S25_115	NOR5-1A	669		-87.09	5.55	1
EF573772	S25_116	NOR5-1A	786		-87.09	5.55	1
EF573840	S25_184	NOR5-1A	1497		-87.09	5.55	1
EF573844	S25_188	NOR5-1A	1497		-87.09	5.55	1
EF574004	S25_348	NOR5-1A	1497		-87.09	5.55	1
EF574418	S25_762	NOR5-1A	1497		-87.09	5.55	1
EF574432	S25_776	NOR5-1A	1498		-87.09	5.55	1
EF574499	S25_843	NOR5-1A	1497		-87.09	5.55	1
EF574750	S25_1094 [b]	NOR5-1A	1499		-87.09	5.55	1
EF574820	S25_1164	NOR5-1A	1497		-87.09	5.55	1
EF575029	S25_1373	NOR5-1A	1497		-87.09	5.55	1
EF575100	S25_1444	NOR5-1A	1475		-87.09	5.55	1
EF575261	S25_1605	NOR5-1A	1497		-87.09	5.55	1
EF575314	S25_1658	NOR5-1A	1497		-87.09	5.55	1
EF575338	S25_1682	NOR5-1A	710		-87.09	5.55	1
EF582172	JL-BS-61N0m007	NOR5-5B	650				1
EF632657	Asc-s-45	NOR5-3	1522		-68.32	-21.52	3
EF632658	Asc-s-2 [b]	NOR5-3	1528		-68.32	-21.52	3
EU005288	G1-45	NOR5-10	824				1
EU005325	G3-63	NOR5	964				1
EU005337	G7-27	NOR5	981				1
EU005342	G7-32	NOR5	972				1
EU005344	G7-34	NOR5-8	923				1
EU005345	G7-35	NOR5	911				1
EU010139	B24	NOR5-1B	1480				1
EU010152	B81	NOR5-1C	1476				1
EU010155	B90	NOR5-1A	1474				1
EU010181	C22	NOR5-1A	1475				1
EU010207	C61	NOR5-1A	1430				1
EU010228	B36	NOR5-1A	1359				1
EU672847	Ivo 10 red [c]	NOR5-3	1492		8.42	55.04	2
EU672848	Ivo 11 [c]	NOR5-3	1492		8.42	55.04	2
EU672849	Ivo 14 [c]	NOR5-1B	1492		8.42	55.04	2
EU672850	Ivo 19 [c]	NOR5-3	1492		8.42	55.04	2
EU672851	Mel 7 [c]	NOR5-3	1492		8.42	55.04	2
EU672852	Mo 10 red [c]	NOR5-3	1492		8.42	55.04	2
EU672853	Pao 12 [c]	NOR5-3	1492		8.42	55.04	2
EU672854	RAp 1 red [c]	NOR5-3	1492		8.42	55.04	2
EU672855	RAp 6 [c]	NOR5-3	1492		8.42	55.04	2
EU672856	RAp 7 [c]	NOR5-3	1492		8.42	55.04	2
EU672857	RAp 9 [c]	NOR5-3	1492		8.42	55.04	2
EU672858	RAp 14 1B [c]	NOR5-1B	1492		8.42	55.04	2
EU672859	Mel 5 [c]	NOR5-3	1492		8.42	55.04	2
EU672860	RAp 11 [c]	NOR5-3	1492		8.42	55.04	2
EU672861	Mo 12 red [c]	NOR5-3	1492		8.42	55.04	2
EU672862	Mo 4 [c]	NOR5-1B	1492		8.42	55.04	2

EU672863	Mel 6 [c]	NOR5-3	1492		8.42	55.04	2
EU672864	RAp 5 [c]	NOR5-3	1492		8.42	55.04	2
EU672865	RAp 13 red [c]	NOR5-1B	1492		8.42	55.04	2
EU672866	RAp 8 [c]	NOR5-3	1492		8.42	55.04	2
EU672867	Mo 5 [c]	NOR5-1B	1492		8.42	55.04	2
EU672868	RAp 14 3 [c]	NOR5-3	1492		8.42	55.04	2
EU672869	RAp 2 [c]	NOR5-3	1492		8.42	55.04	2
EU707293	JSS S04 clone 488	NOR5	1492		7.68	53.72	2
EU707294	JSS S05 clone 128	NOR5-3	1498		7.68	53.72	2
EU707295	JSS S04 clone 292	NOR5-8	1492		7.68	53.72	2
EU707296	JSS S04 clone 334	NOR5-11	1491		7.68	53.72	2
EU707297	JSS S04 clone 336	NOR5	1492		7.68	53.72	2
EU707298	JSS S04 clone 359	NOR5	1490		7.68	53.72	2
EU707299	JSS S04 clone 388	NOR5-8	1492		7.68	53.72	2
EU707300	JSS S04 clone 452	NOR5	1492		7.68	53.72	2
EU707301	JSS S04 clone 457	NOR5-6	1492		7.68	53.72	2
EU707302	JSS S04 clone 466	NOR5-11	1492		7.68	53.72	2
EU707303	JSS S04 clone 482	NOR5-6	1492		7.68	53.72	2
EU707304	JSS S04 clone 516	NOR5	1492		7.68	53.72	2
EU707305	JSS S04 clone 527	NOR5	1492		7.68	53.72	2
EU707306	JSS S04 clone 536	NOR5-6	1492		7.68	53.72	2
EU707307	JSS S04 clone 549	NOR5-9	1493		7.68	53.72	2
EU707308	JSS S04 clone 554	NOR5-3	1494		7.68	53.72	2
EU707309	JSS S04 clone 557	NOR5-11	1492		7.68	53.72	2
EU707310	JS9_43	NOR5	1491		7.68	53.72	2
EU707311	JS9_29	NOR5	1492		7.68	53.72	2
EU707312	Dangast_21	NOR5-11	1496		8.12	53.45	2
EU707313	Dangast_27	NOR5-11	1493		8.12	53.45	2
EU707318	JSS S04 clone 320	NOR5-6	699		7.68	53.72	2
EU707319	JSS S04 clone 315	NOR5	699		7.68	53.72	2
EU707320	JSS S05 clone 106	NOR5	699		7.68	53.72	2
EU707321	JSS S04 clone 547	NOR5-9	699		7.68	53.72	2
EU707322	JSS S04 clone 422	NOR5	699		7.68	53.72	2
EU707323	JSS S04 clone 495	NOR5	699		7.68	53.72	2
EU707324	JSS S04 clone 538	NOR5	699		7.68	53.72	2
EU707325	JSS S04 clone 535	NOR5-5	699		7.68	53.72	2
EU707326	JSS S04 clone 561	NOR5	699		7.68	53.72	2
EU707327	JSS P05 clone 886	NOR5	701		7.68	53.72	2
EU707328	JSS_PIB6	NOR5-6	699		7.68	53.72	2
L35466	SAR125	NOR5-1A	284	+	-64.38	32.07	1
U43635	EH-10	NOR5	326	+	-122.50	47.60	2
U43637	EH-15	NOR5-11	326	+	-122.50	47.60	2
U70696	OM60	NOR5-1B	1525	+	-75.13	35.98	1
U70702	OM241	NOR5-4	1491	+	-75.13	35.98	1
U84619	pB7-104R	NOR5-1C	291		-64.75	32.40	1
Z77586	1-700 C32	NOR5	284				
Z77618	2-100 C26	NOR5-11	314				
Z77624	2-400 C2.14	NOR5-5	347				
Z77630	2-400 C2.26	NOR5-5	362				

Z77634	2-400 C2.34	NOR5-11	307				
Z77635	2-400 C2.35	NOR5-5	306				
Z77647	2-400 C2/2.5	NOR5	303				
Z88579	HRV1	NOR5-1A	784		13.66	45.13	1
	JCVI_SCAF_1093012159849	NOR5-4	890	+	-75.39	36.00	1
	JCVI_SCAF_1093012223429	NOR5-4	632	+	-75.39	36.00	1
	JCVI_SCAF_1097156386993	NOR5-1B	1472	+	-90.43	-1.23	3
	JCVI_SCAF_1097156710436	NOR5-1A	801	+	-66.22	42.85	1
	JCVI_SCAF_1097159028986	NOR5-1C	990	+	-63.64	44.14	1
	JCVI_SCAF_1097159028986	NOR5-1C	581	+	-63.64	44.14	1
	JCVI_SCAF_1097169027043	NOR5-1C	910	+	-71.35	41.49	1
	JCVI_SCAF_1097173022133	NOR5-1B	1535	+	-74.69	38.94	1
	JCVI_SCAF_1097205716456	NOR5-1A	548	+	-83.78	18.04	1
	JCVI_SCAF_1097205742466	NOR5-1A	1533	+	-79.26	32.51	1
	JCVI_SCAF_1097208169768	NOR5-1A	1366	+	-82.90	6.49	1
	JCVI_SCAF_1097263268721	NOR5-1C	1365	+	-90.28	-0.38	1
	JCVI_SCAF_1097263565737	NOR5-1A	548	+	-147.44	-15.14	1
	JCVI_SCAF_1097263577073	NOR5-1A	468	+	-147.44	-15.14	1
	JCVI_SCAF_1097263759176	NOR5-1C	445	+	-90.32	-1.22	1
	JCVI_SCAF_1099266679416	NOR5-1A	1533	+	-64.32	31.18	1
	JCVI_SCAF_1099266943237	NOR5-1C	472	+	-65.50	32.17	1
	JCVI_SCAF_1101370270864	NOR5-1A	1089	+	-91.07	-0.59	1
	JCVI_SCAF_1101669020732	NOR5-1C	548	+	-67.24	42.50	1
	JCVI_SCAF_1101669038421	NOR5-1A	844	+	-63.64	44.14	1
	JCVI_SCAF_1101669039626	NOR5-1A	302	+	-63.64	44.14	1
	JCVI_SCAF_1101669085068	NOR5-1C	497	+	-71.35	41.49	1
	JCVI_SCAF_1101669447951	NOR5-1C	1534	+	-79.69	8.13	1
	JCVI_SCAF_1101669483308	NOR5-1C	359	+	-82.90	6.49	1
	JCVI_SCAF_1101669568955	NOR5-1C	874	+	-90.42	-1.22	1
	JCVI_SCAF_1101670021427	NOR5-1A	1083	+	-90.84	-0.20	1
	JCVI_SCAF_1101670048457	NOR5-1C	321	+	-91.63	0.27	2
	JCVI_SCAF_1101670167881	NOR5-1C	666	+	-91.65	-0.30	1
	JCVI_SCAF_1101670174504	NOR5-1C	1282	+	-91.65	-0.30	1
	JCVI_SCAF_1101670359330	NOR5-1A	775	+	-90.28	-0.38	1

[a] environmental categories: 1-marine water and other marine habitats; 2-marine sediment; 3-hypersaline; 4-soil; 5-fresh water; 6-fresh sediment; 7-deep sea

[b] bad aligned sequence or suspected chimera

[c] isolated strain

SI Table 4 Published strains of NOR5/OM60 group

Strain	Accession number	Length	Subclade	Environment	Position of isolation	Refe-rence
KT71	AAOA 01000004	1535	NOR5-3	marine coastal water	Germany, Helgoland	[19]
HTCC160	AY102017	672	NOR5-1B	marine coastal water	US, Oregon	[14]
HTCC227	AY102018	670	NOR5-8	marine coastal water	US, Oregon	[14]
HTCC240	AY102019	646	NOR5-8	marine coastal water	US, Oregon	[14]
HTCC2080	AAVV 01000015	1533	NOR5-1B	marine pelagic water	US, Oregon	[14]
HTCC2148	AY386334	1350	NOR5-8	marine pelagic water	US, Oregon	[11]
HTCC2149†	AY386331	902	NOR5-1A	marine pelagic water	US, Oregon	[11]
HTCC2223	AY386336	1352	NOR5-1B	marine pelagic water	US, Oregon	[11]
HTCC2246	AY386337	1350	-	marine coastal water	US, Oregon	[11]
3X/A02/235	AY576769	1485	NOR5-2	marine coastal water	France, Banyuls-sur-Mer	[1, 2]
NEP1	AB212800	1456	NOR5-7	marine coastal sediment	Japan, Shimonoseki	[45]
NEP3	AB212802	1455	NOR5-7	marine coastal sediment	Japan, Shimonoseki	[45]
NEP17	AB212804	1456	NOR5-2	marine coastal sediment	Japan, Shimonoseki	[45]
NEP65	AB212805	1456	NOR5-7	marine coastal sediment	Japan, Kochi	[45]
NEP68	AB212806	1456	NOR5-6	marine coastal sediment	Japan, Tokyo	[45]

† Could not be recovered

Unit 2

Potential novel photoautotrophy in the NOR5/OM60 clade of *Gammaproteobacteria* discovered by genome comparison

Shi Yan[a], Bernhard M. Fuchs[a]*, Jens Harder[b], Rudolf Amann[a]

Affiliations:

[a]Department of Molecular Ecology and [b]Department of Microbiology, Max Planck Institute for Marine Microbiology, Bremen, D-28359, Germany

* to whom correspondence should be addressed

Summary

Based on the analysis of the isolate Candidatus *Congregibacter litorasis* KT71, it was recently demonstrated that the "missing" gammaproteobacterial aerobic anoxygenic phototrophs (AAnP) are from the NOR5/OM60 clade. The aim of this study was to gain additional insights into the functional potential of this clade by comparative genome analysis of 5 strains of the NOR5/OM60 clade (KT71, RAp1red, Ivo14, HTCC2080 and HTCC2148) and one strain of its sister group BD1-7 clade (HTCC2143). We identified a complete photosynthesis superoperon, several genes for 3-hydroxypropionate pathway of CO_2 fixation as well as *sox* operon for sulfur compound oxidation in four strains: KT71, RAp1red, Ivo14 and HTCC2080. The discovery of two key genes for 3-hydroxy-propionate pathway (malonyl-CoA reductase (NADPH) and propionyl-CoA synthase) is the first time in *Gammaproteobacteria*. On the other hand, the proteorhodopsin genes were found in HTCC2148 and HTCC2143. These findings provide important hints about the possible novel living strategy and function of the NOR5/OM60 clade of marine gammaproteobacteria.

Key words: phototrophy, carbon fixation, sulfur oxidation

Introduction

The NOR5/OM60 clade encompasses a group of *Gammaproteobacteria*, which is widespread in marine habitats (Yan, Fuchs et al. 2009). Members of this clade are particularly abundant in coastal settings like the North Sea (Eilers, Pernthaler et al. 2001), but are also found in pelagic surface water, deep sea sediment or freshwater sediment (Yan, Fuchs et al. 2009). Several strains were isolated from many locations, including the North Sea coastal regions and northeast Pacific Ocean (Cho and Giovannoni 2004). Candidatus *Congregibacter litoralis* KT71 is the first isolated strain of the NOR5/OM60 clade. It was obtained in the year 1998 from surface water sample at station Kabeltonne, close to the island Helgoland, German Bight, North Sea (Eilers, Pernthaler et al. 2001). Strain HTCC2080, HTCC2143 and HTCC2148 were isolated in the year 2001, using a high-throughput culturing (HTC) method (Cho and Giovannoni 2004). The first two strains were from a Pacific Ocean surface water sample taken at the jetty of Newport, while the latter was obtained from pelagic water sampled at 10 m depth, 27.6 km off the coast of Oregon. Furthermore, 20 other strains were isolated from the oxic layer of sediment of island Sylt at the German North Sea coast in the year 2005 (Yan, Fuchs et al. 2009), among them are the strains RAp1red and Ivo14. The 16S rRNA phylogeny of the NOR5/OM60 clade was investigated in detail and 13 subclades were recognized (Yan, Fuchs et al. 2009). HTCC2080 and Ivo14 belong to the subclade NOR5-1B, while KT71 and RAp1red belong to NOR5-3, and HTCC2148 to subclade NOR5-8. Based on comparative 16S rRNA analysis, strain HTCC2143 is more distant to the other sequences, and was classified as a member of BD1-7 group, a sister group of NOR5/OM60 (Cho and Giovannoni 2004).

The genome analysis of KT71 revealed the presence of a full photosynthesis (PS) operon, and subsequently bacteriochlorophyll *a* was proved to be expressed (Fuchs, Spring et al. 2007). Therefore it was recognized as the first strain of gammaproteo-bacterial aerobic anoxygenic phototroph (AAnP). In contrast to many other photo-synthetic organisms, AAnPs were not able to grow photoautotrophically. Only marginal proportion of carbon anabolism was obtained through fixation of inorganic carbon compounds (Yurkov and Beatty 1998; Kolber, Plumley et al. 2001).

Till now, there are two known mechanisms for prokaryotes to utilize light energy: with chlorophyll or bacteriochlorophyll (Kolber, van Dover et al. 2000), and with rhodopsin (Béjà, Aravind et al. 2000). The former type exists in *Cyanobacteria*, *Chloroflexi* (green non-sulfur bacteria), *Chlorobi* (green sulfur bacteria), *Heliobacteraceae*, and *Alpha-*, *Beta-* and *Gammaproteobacteria* (purple sulfur bacteria, purple non-sulfur bacteria and AAnPs), while rhodopsin is widely distributed in *Archaea*, *Proteobacteria*, *Flavobacteria* etc. (McCarren and DeLong 2007).

In this study, we compared the genomes of strains KT71, RAp1red, Ivo14, HTCC2080, HTCC2148 (all NOR5/OM60) and HTCC2143 (BD1-7). All strains have been fully sequenced and preliminarily annotated by the J. Craig Venter Institute (JCVI). The focus of this study is less on general aspects, but more on particular genes such as those are needed to utilize light.

Materials and Methods

Sequence retrieval

All genomes investigated in this study were sequenced using the shot gun approach. The sequence of *Congregibacter litoralis* KT71 is available from Genbank (CH672401, CH672402), while all the other five genomes RAp1red (under the name "NOR5-3"), Ivo14 ("NOR51-B"), HTCC2080, HTCC2148, HTCC2143 can be retrieved from J. Craig Venter Institute (https://moore.jcvi.org/moore/).

Genome-wide comparison

Genome pairwise alignment was done using PROmer program (version 3.06) of MUMmer package (http://mummer.sourceforge.net/) (Delcher, Phillippy et al. 2002; Kurtz, Phillippy et al. 2004). PROmer was run with the default settings but with parameter "--mum" for "using anchor matches that are unique in both sequences". The program finds the orthologous regions of the two genomes with all six reading frames and translates into amino acid sequences. However, other orthologous regions like rRNA or tRNA as well as non-coding homologous regions can also be found. The "mummerplot" program then generates the dot plot for pairwise alignment. A java script was used to calculate from the output file (generated by "show-coords" command of

MUMmer package), how many bases on each genome were covered by the orthologous regions in pairwise comparison. For each compared genome pair, the average percentage of the length of orthologous regions on both genomes was calculated and filled into a similarity matrix. Then the matrix was used for calculating a neighbor joining tree (NJ tree) for the strains using "neighbor" version 3.65 from the Phylip package (http://evolution.genetics.washington.edu/phylip.html).

Amplification of partial *puf*LM sequences from NOR5/OM60 strains

The sequence from the end of *puf*L to the end of *puf*M was amplified from most of the NOR5/OM60 strains isolated from the island of Sylt (Table 1). Hundreds of *puf*LM sequences were collected from GenBank and aligned according to their amino acid sequences. Old primers were checked and new primers were designed on the most conservative regions in order to cover most of the sequences. The primers used in this study for amplification were pufL_WW_F (5′-Y TAV TGG TGG VVN TGG TGG-3′, forward, which locates at a tryptophan-rich region at the end of *puf*L, designed in this study) and pufM_uni_R (5′-YC CAT NGT CCA NCK CCA RAA-3′ reverse, which locates at the end of *puf*M (Yutin, Suzuki et al. 2005)). The polymerase chain reaction (PCR) was performed with the annealing temperature of 60 °C, 35 cycles. The PCR products (about 1000 bp) were ligated in TOPO vector and cloned into TOP10 *E. coli* competent cells and finally sequenced using primers M13F and M13R.

Table 1 List of the strains isolated from the North Sea

	Source	Subclade	Color	*puf*LM sequence obtained
KT71	Helgoland, surface water	NOR5-3	white	+
RAp1red	Sylt, aerobic sediment	NOR5-3	dark red	+
RAp2	Sylt, aerobic sediment	NOR5-3	dark red	+
RAp5	Sylt, aerobic sediment	NOR5-3	dark red	+
RAp6	Sylt, aerobic sediment	NOR5-3	dark red	+
RAp7	Sylt, aerobic sediment	NOR5-3	dark red	- [b]
RAp8	Sylt, aerobic sediment	NOR5-3	dark red	- [b]
RAp9	Sylt, aerobic sediment	NOR5-3	dark red	+
RAp11	Sylt, aerobic sediment	NOR5-3	dark red	- [b]
RAp13red	Sylt, aerobic sediment	NOR5-1B	pink	+
RAp14	Sylt, aerobic sediment	NOR5-3 / NOR5-1B [a]	dark red	+
Ivo10red	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red	+
Ivo11	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red	+
Ivo14	Sylt, top oxic layer of muddy sediment	NOR5-1B	pink	+
Ivo19	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red	-
Pao12	Sylt, top oxic layer of muddy sediment	NOR5-3	dark red	+
Mo4	Sylt, oxic layer of sandy sediment	NOR5-1B	pink	+
Mo5	Sylt, oxic layer of sandy sediment	NOR5-1B	pink	+
Mo10red	Sylt, oxic layer of sandy sediment	NOR5-3	dark red	+
Mo12red	Sylt, oxic layer of sandy sediment	NOR5-3	dark red	+
Mel5	Sylt, 15 cm depth of muddy sediment	NOR5-3	dark red	+
Mel6	Sylt, 15 cm depth of muddy sediment	NOR5-3	dark red	+
Mel7	Sylt, 15 cm depth of muddy sediment	NOR5-3	dark red	+

[a] From strain RAp14 we have retrieved two different 16S rRNA sequences, therefore it might be a mixture from two strains.
[b] *puf*M successfully amplified, but sequence not fully obtained

Gene searching and comparison

The genome sequences were all automatically annotated by JCVI. These sequences were imported into GenDB gene annotation system (http://www.cebitec.uni-bielefeld.de/groups/brf/software/gendb_info/index.html) (Meyer, Goesmann et al. 2003) and the genes were further analyzed. The genomes were edited with the JCoast tool (http://www.megx.net/jcoast/) (Richter, Lombardot et al. 2008).

Although the sequences have been automatically annotated, all the gene findings in this study were based on manual annotation rather than automatic, since in several cases the automated annotation was imperfect. In order to determine whether a gene has homologs existing in the six genomes, we used the "local BLAST" function in the software BioEdit (version 7.0.5.2) (Altschul, Madden et al. 1997; Hall 1999). The query sequence was either a gene sequence from GenBank, or a sequence from one of the six genomes. The full nucleotide sequences of the six genomes were made as the subject

database, and tblastn (search translated nucleotide database using a protein query) algorithm was used for searching.

The phylogeny of several genes was investigated in this study. The sequences were retrieved either by amplification (only for *puf*M) or from genomic nucleotide sequences from the public database. The sequences were aligned either using ClustalW multiple alignment (Thompson, Higgins et al. 1994) based on amino acid sequences in BioEdit, or using E-INS-i algorithm of MAFFT (Version 6) on web server (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/) (Katoh and Toh 2008). The aligned sequences were imported into ARB (http://www.arb-home.de/) (Ludwig, Strunk et al. 2004), and different trees were made based on amino acid sequences. The base columns with scores higher than a certain value (30 – 50, depending on the alignment of the protein) were filtered for treeing. Neighbor-joining (NJ) and maximum parsimony (MP) trees were made inside ARB, while maximum likelihood (ML) tree were done using either RAxML on web server (http://phylobench.vital-it.ch/raxml-bb/) (Stamatakis 2006; Stamatakis, Hoover et al. 2008) or MrBayes (Version 3.1, http://mrbayes.csit.fsu.edu/) for posterior probability trees (Huelsenbeck and Ronquist 2001).

Results

Genome characteristics

All the genomes are middle-sized, ranging from 3.2 to 4.4 Mbp (Table 2). In all cases except for HTCC2148, the sequences could be assembled into only a few scaffolds. Each genome likely contains 1 or 2 rRNA operon(s), although some rRNA operons are located at the end of a scaffold and possibly caused difficulty in assembly. Therefore, one rRNA operon may appear in more than one scaffolds, and the maximal number could be 4. All sequences contain less than 0.5% of N bases (not determined as A/G/C/T).

Table 2 Basic information of the six genomes in this study

	KT71	RAp1red	Ivo14	HTCC2080	HTCC2148	HTCC2143
Clade	NOR5-3	NOR5-3	NOR5-1B	NOR5-1B	NOR5-8	BD1-7
Scaffolds	2	6	1	2	31	4
Total length (bp)	4,344,414	4,208,084	3,261,541	3,582,105	4,326,936	3,940,784
Percentage of N*	0.40%	0.10%	0.43%	0.17%	0.39%	0.38%
G+C content	57.68%	56.34%	56.74%	51.82%	52.96%	47.16%
rRNA operons	2	2 – 4	1	1	1 – 3	1

*N indicates undetermined nucleotides (besides A, G, C and T)

The G+C content varies between 51 – 58% in the five NOR5/OM60 strains, and is with 47% lower in HTCC2143. The G+C content of most of the scaffolds from one strain is very similar, with usually less than 1% deviation from the average value. A few short scaffolds show greater deviation to the average, e.g. scaffold HTCC2148_12 (27,779 bp) has a G+C content of 46.19% compared to the strain's average of 52.96%.

General comparison of the genomes

To estimate the whole genome relationship, we calculated the whole genome orthology via genome-wide pairwise alignment. Dot plots (Fig. 1) show the aligned homologous regions in the both directions. The six scaffolds of RAp1red genome were rearranged in order to show a better collinearity to KT71 (Fig. 1a). The aligned regions are almost on the diagonal, which means that no big gene insertion/deletion (indels) or rearrangement events have occurred since the speciation of RAp1red and KT71. For the other pairwise comparisons, the collinearity is not so clear. HTCC2080 is less collinear to KT71 than RAp1red, with obvious stretches of indels and rearrangement (Fig. 1b), whereas Ivo14 has a large stretch of genome reversed (Fig. 1c). The genome of HTCC2148 was separated in 31 small scaffolds, therefore cannot be well aligned to KT71. Between HTCC2143 and KT71 only short stretches are in synteny (Fig. 1d).

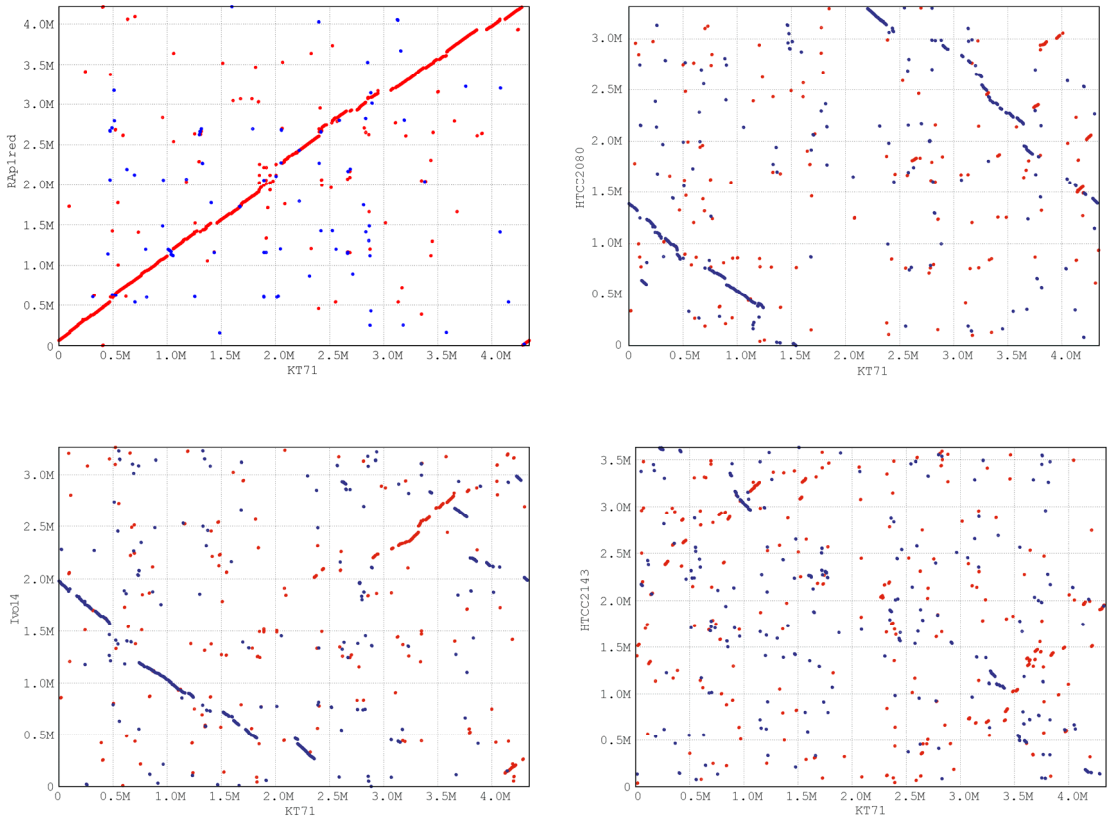

Fig. 1 Pairwise alignment of the genomes. The aligned region in the same direction was plotted as red, and reverse complement as blue. (a) KT71 (x-axis) – RAp1red (y-axis), with scaffolds of RAp1red rearranged in the order 1-3-5-2-4-6 and reverse-complemented; (b) KT71 – HTCC2080; (c) KT71 – Ivo14; (d) KT71 – HTCC2143

For each pairwise alignment, the orthologous regions were summed up, and the percentage of total orthologous regions were calculated for each pairs of genomes. The similarity matrices of genome and 16S rRNA were made. The neighbor-joining trees calculated from the both matrices showed identical topology (Fig. 2).

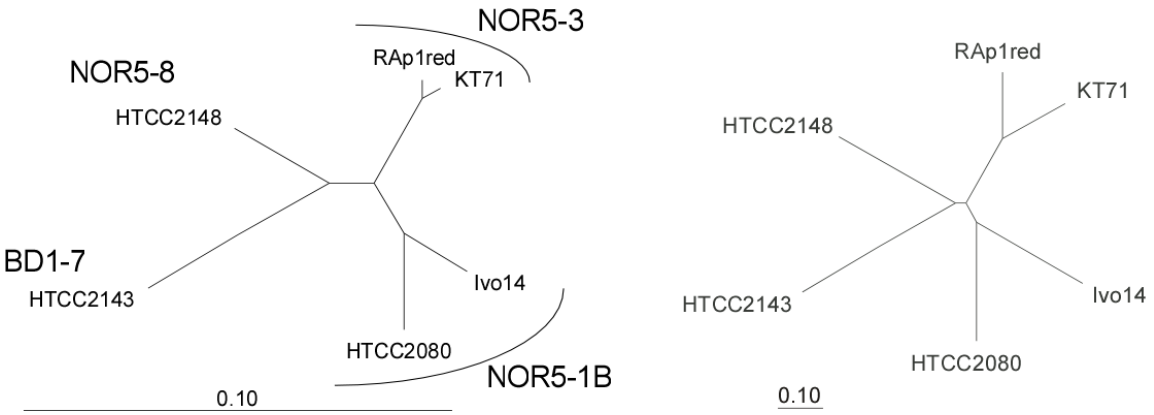

Fig. 2 Neighbor joining trees based on 16S rRNA similarity (left) and genome homology (right).

An overview on the distribution of clusters of orthologous groups (COG) is shown in Fig. 3. The composition of the COG categories in the six genomes are by large similar, although with some variations: e.g., Category L – DNA replication, recombination and repair genes in KT71 are markedly more than in other strains.

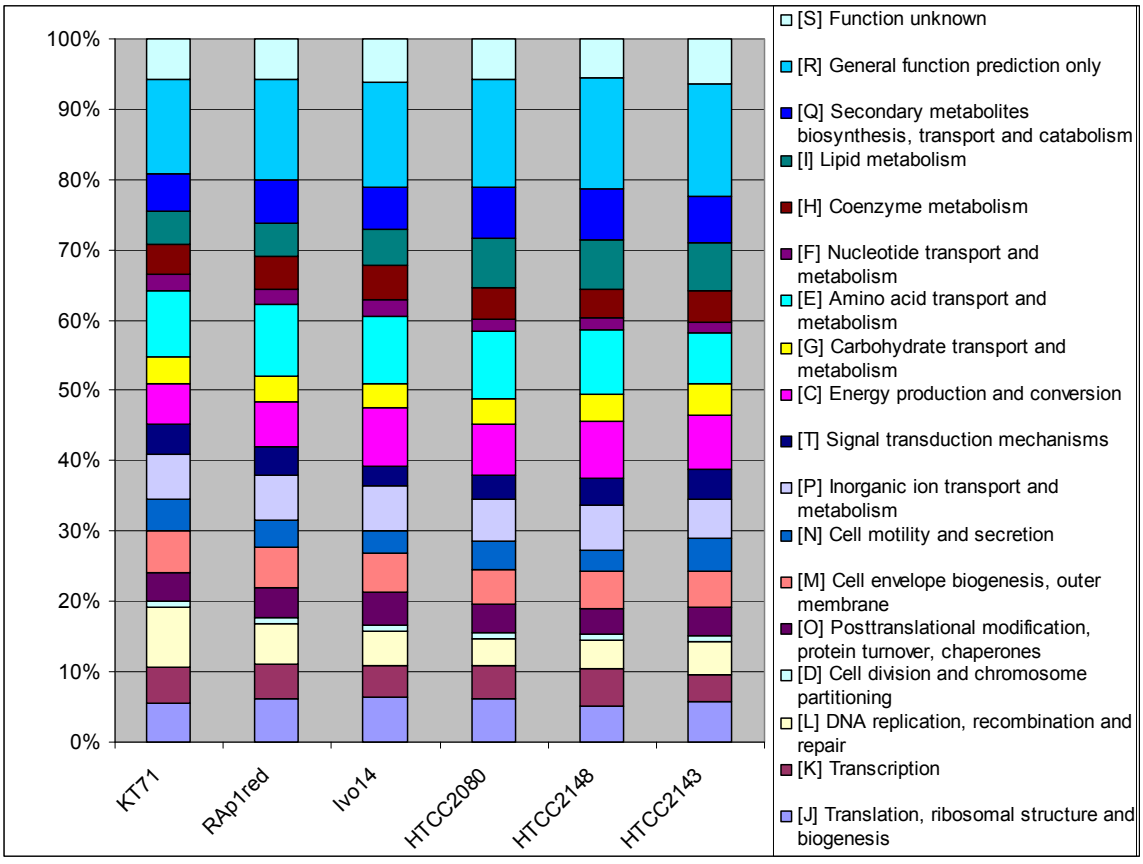

Fig. 3 COG (Clusters of Orthologous Groups) codes distribution in the six genomes.

Specific genes

PS superoperon

The photosynthesis (PS) superoperon was found in four genomes: KT71, RAp1red, Ivo14 and HTCC2080 (Fig. 4), but was completely absent in the genomes of HTCC2148 and HTCC2143. The superoperon contains *bch* (bacteriochlorophyll synthesis), *puf* (light-harvesting complex I (LHC I) and reaction center) and *crt* (carotenoid synthesis) genes. The organization of the PS superoperon is highly similar, at a length of 40 – 45 kbp, but not exactly identical. The *puf*LMCBA arrangement is the same and unique for *Gammaproteobacteria* (Yutin and Béjà 2005). Ivo14 and HTCC2080 lack the *crt*EJ genes at the end of the operon and the "hypothetical alpha/beta hydrolase or acyltransferase" between *crt*I and *crt*E is reversed. For Ivo14 and HTCC2080, downstream of *bch*ODI and this hypothetical protein are three genes: *crt*D, a conserved membrane protein, and cytochrome P450 (not shown in the figure), which are located at another site on KT71 and RAp1red genomes. The *phe*A (phenol 2-mono-oxygenase) gene is inserted at the beginning of the HTCC2080 PS operon, after the BLUF (sensor of blue-light using FAD) domain protein and *acs*F (magnesium-proto-porphyrin IX monomethyl ester aerobic oxidative cyclase) genes. The *bch*H (magnesium chelatase H subunit) and *bch*M (magnesium protoporphyrin O-methyltransferase) genes of RAp1red are located about 24 kbp upstream of the PS superoperon. The similarity of *bch*H between RAp1red and KT71 is lower than that between HTCC2080 and KT71, and in RAp1red, an end part of *bch*H remains between *bch*B and *bch*L. Therefore, the *bch*HM genes of RAp1red might be obtained through lateral gene transfer and have substituted their paralogs in the PS superoperon.

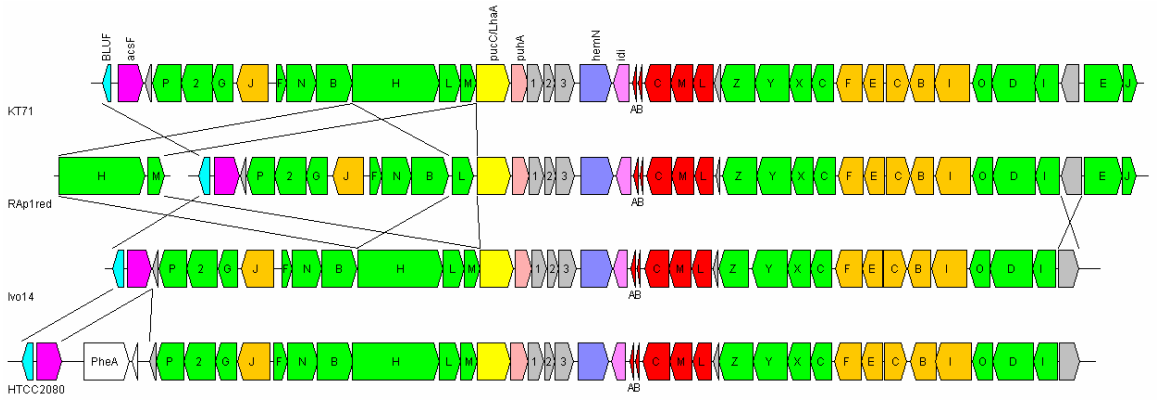

Fig. 4 Comparison of PS operons of KT71, RAp1red, Ivo14 and HTCC2080. Green, *bch* genes; red, *puf* genes; orange, *crt* genes; light grey, unknown conserved genes. The *bch*HM genes of RAp1red locate on ~24 kbp upstream of PS superoperon on the same scaffold (scaffold 4).

The *puf*L and *puf*M genes that code for PS reaction center proteins were described by Yutin et al. (Yutin, Suzuki et al. 2007). The sequences of *puf*M from these four genomes as well as the amplicons from the other NOR5/OM60 isolates are closely related. The *puf*M phylogeny is congruent to the 16S rRNA tree (Yan, Fuchs et al. 2009) (Fig. 5).The *puf*M genes within subclade NOR5-3 are nearly identical (at most 1 amino acid differs), except for those of KT71 and Mo10red. The same case applies to the *puf*M genes of the North Sea strains from the NOR5-1B subclade. According to the former study (Yutin, Suzuki et al. 2007), the *puf*M genes from the NOR5-1B group are located inside the Group K, while the gene from KT71 is the closest relative of Group K. As studied before (Cho, Stapels et al. 2007), the *puf*M sequences of HTCC2148 and HTCC2246 (also a NOR5/OM60 member), which were obtained using PCR, do not group with other NOR5/OM60 sequences.

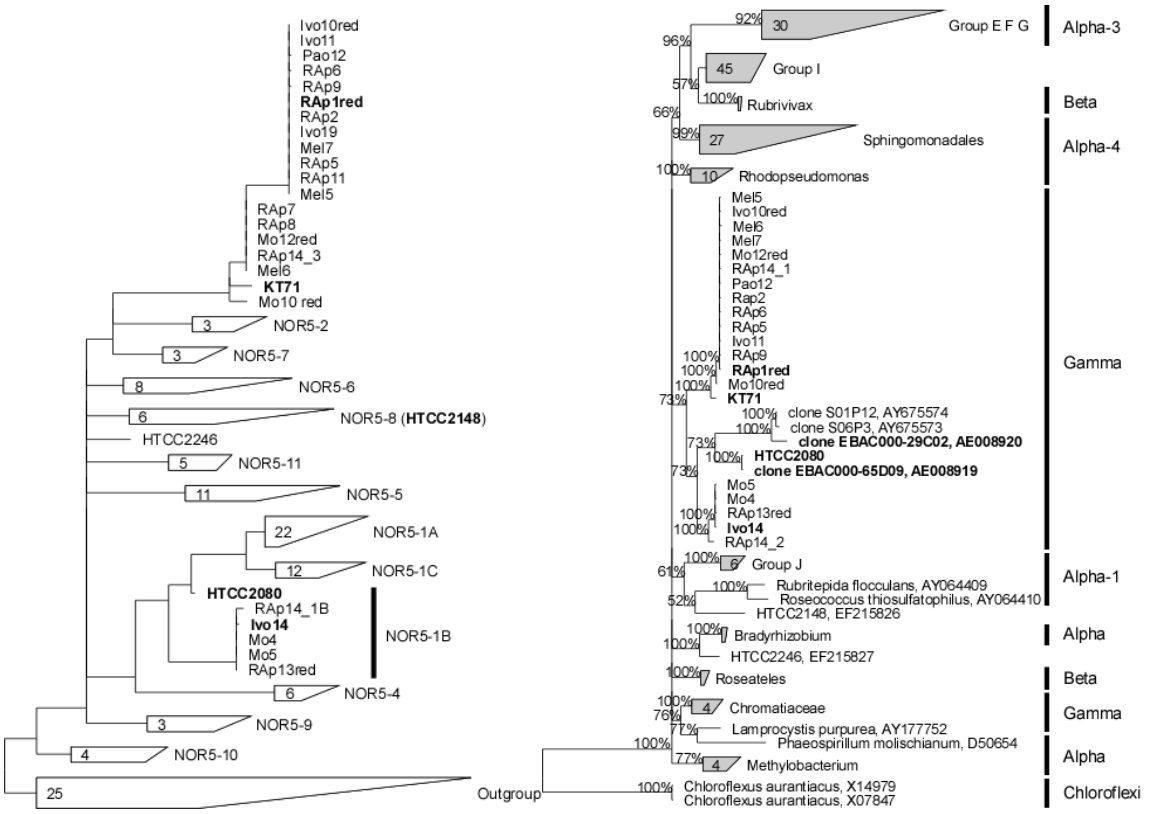

Fig. 5 A comparison of 16S rRNA (left) and *puf*M (right) consensus trees of NOR5/OM60 strains. The 16S tree was reduced from the consensus tree published before (Yan, Fuchs et al. 2009). The pufM tree was calculated using MrBayes (with 239 valid residues), made consensus and reduced. The percentage numbers at the nodes show the posterior probability. The names in bold indicate sequences acquired from genomes of isolated strains or metagenomic library, while the other sequences were retrieved through PCR amplification.

The operon coding for LHC II, *puc*BAC, has only been found in the strains KT71 and RAp1red. It is in both cases separated from the PS superoperon.

Proteorhodopsin

The proteorhodopsin genes (*pop*) were found only in the genomes of HTCC2148 and HTCC2143, but not in the four genomes in which PS superoperon was present. The *pop* gene of HTCC2148 is located at the beginning of a very short scaffold (scaffold 18, 4490 bp) and the sequence is not complete (540 bp), while the *pop* gene in HTCC2143 is complete (690 bp). The proteorhodopsin of HTCC2148 and HTCC2143 both contain a leucine residue (L) at the position 105 (SAR86 EBAC31A08 numbering) which is

indicative for absorbing green light (Béjà, Spudich et al. 2001; Man, Wang et al. 2003; Giovannoni, Bibbs et al. 2005; Fuhrman, Schwalbach et al. 2008).

The *pop* gene of HTCC2143 is the closest relative of the SAR92 group (Fig. 6), while that of HTCC2148 also cluster with other *Alpha-* and *Gammaproteobacteria*, although the exact position cannot be determined due to incompleteness of its sequence. The SAR92 group is one of the most closely related groups to the NOR5/OM60 clade and HTCC2143 (16S rRNA sequence identity between the groups is 88 – 92%). Proteorhodopsin genes were found in strain HTCC2207 and three other strains, which all belong to the subgroup B of the SAR92 clade (Stingl, Desiderio et al. 2007), as the first discovered examples of proteorhodopsin from cultivated gammaproteobacterial strains.

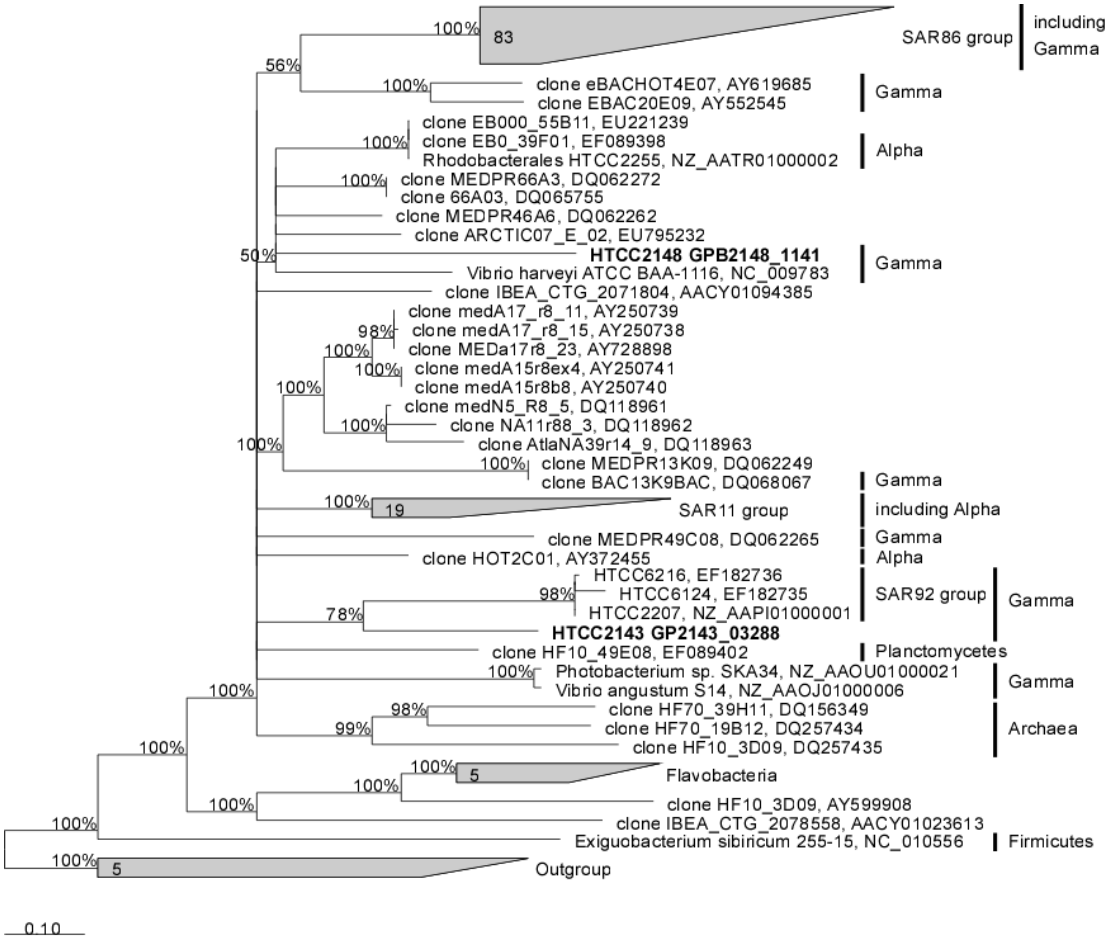

Fig. 6 Consensus tree of proteorhodopsin (based on amino acid) based on maximum likelihood and posterior probability trees, showing the positions of HTCC2143, which are closest to the sequences HTCC2207 of SAR92 group, and the partial sequence of HTCC2148, of which position cannot be well solved in the tree.

Downstream of the HTCC2143 proteorhodopsin gene are the genes for retinal synthesis, in the order *pop-crt*EIBY-*blh-fni* (*crt*E = *ids*A), all translated in the same direction. This gene arrangement is identical as in HTCC2207 (Stingl, Desiderio et al. 2007). The *crt*EIB genes also exist in the PS superoperon of the four strains, but are arranged in different order: *crt*IBCEF (Fig. 5). The *crt*E and *crt*B of HTCC2143 show low similarity with the ones in the PS superoperon. However, the genes for retinal synthesis are not found in the genome of HTCC2148, and the downstream of *pop* are functionally unrelated genes.

Carbon fixation

The key genes of Calvin Cycle (ribulose-1,5-bisphosphate-carboxylase/oxygenase (RuBisCO)), reverse TCA cycle (pyruvate synthase (EC: 1.2.7.1)) and reductive acetyl-CoA pathway (CO-dehydrogenase/acetyl-CoA-synthase) were not found in any of the six genomes.

We have found several genes of the 3-hydroxypropionate cycle in the four genomes that contain PS superoperon (Fig. 7). This includes malonyl-CoA reductase (*mcr*) and propionyl-CoA synthase (*pcs*), which have not been found involved in any pathway other than the carbon-fixing 3-hydroxypropionate cycle (Hügler, Menendez et al. 2002). The two genes were found in the tandem arrangement as *pcs-mcr* in the genomes of RAp1red, Ivo14 and HTCC2080. We have found only *pcs* in KT71, while *mcr* is missing as reported before (Friedmann, Alber et al. 2007). Until now, these large genes (for HTCC2080, *mcr* 3651 bp and *pcs* 5421 bp) can be found only in a few strains: *Chloroflexus spp.* and *Roseiflexus spp.* (both *Chloroflexi*) and *Erythrobacter sp.* NAP1 (*Alphaproteobacteria*); a single *pcs* gene was found in *Chloroherpeton thalassium* ATCC35110 (*Chlorobi*).

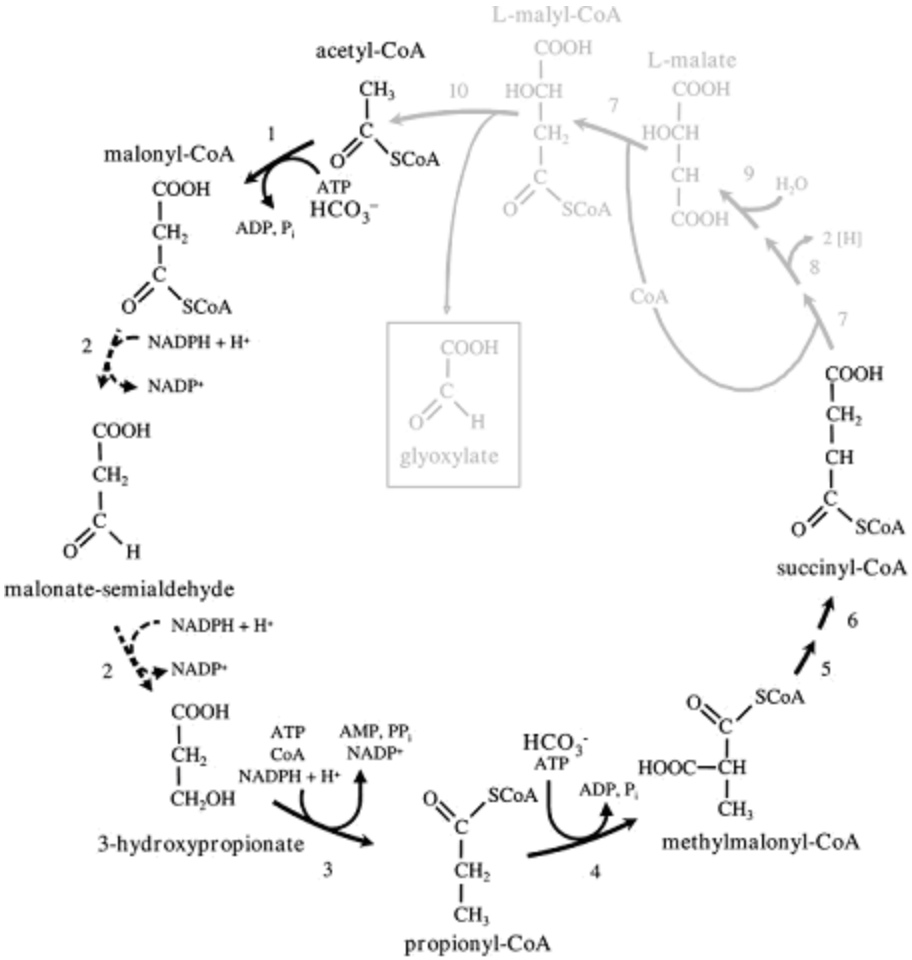

Fig. 7 The 3-hydroxypropionate pathway (adapted from the figure for *Chloroflexus aurantiacus* by Alber et al. (Alber, Olinger et al. 2006)). 1, acetyl-CoA carboxylase (*acc* genes); 2, malonyl-CoA reductase (NADPH) (*mcr*); 3, propionyl-CoA synthase (*pcs*); 4, propionyl-CoA carboxylase (*pcc*); 5, methylmalonyl-CoA epimerase (*mce*); 6, methylmalonyl-CoA mutase (*mcm*); 7, succinyl-CoA:L-malate CoA transferase (*smt*); 8, succinate dehydrogenase; 9, fumarate hydratase; 10, L-malyl-CoA lyase. The *mcr* enzyme for step 2 (dashed arrows) is absent in KT71, but present in RAp1, Ivo14 and HTCC2080, while the genes for steps 1 and 3 – 6 (solid black arrows) exist in all the four strains. The enzymes for step 7 – 10 (in gray) cannot be proved from the genome sequences yet.

A comparative sequence analysis for all the available genomic *pcs* genes to date (Fig. 8) shows clustering of the NOR5/OM60 sequences. The *pcs* sequence of the strain Ivo14 is closer to that of the other North Sea strains than to HTCC2080, which means that the *pcs* phylogeny is not parallel to 16S rRNA phylogeny. The similarity of *pcs* from all the sources is high (e.g. 46 – 54% amino acid identity between the NOR5/OM60 and

Chloroflexi sequences). Therefore it is highly possible that the *pcs* genes in NOR5/OM60 have the same function as in *Chloroflexi*.

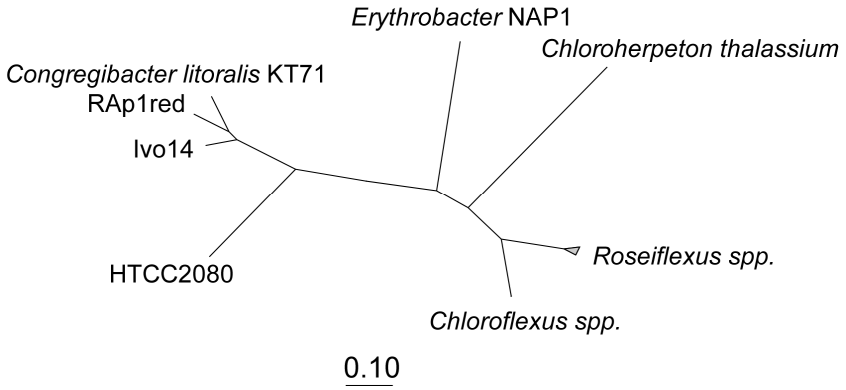

Fig. 8 Maximum likelihood tree of genes for propionyl-CoA synthase (*pcs*). Both *Chloroflexus* and *Roseiflexus* belong to the phylum *Chloroflexi*, while *Chloroherpeton* belongs to *Chlorobi* and *Erythrobacter* belongs to *Alphaproteobacteria*.

The enzymes for the first step of 3-hydroxypropionate pathway, *acc*A, *acc*BC and *acc*D for acetyl-CoA carboxylase were found in all the five strains of the NOR5/OM60 clade, all separated at three isolated locations on a genome. Propionyl-CoA carboxylase (*pcc*BA), methylmalonyl-CoA epimerase (*mce*), methylmalonyl-CoA mutase (*mcm*) and a putative arginine/ornithine transport system ATPase occur tandemly in all the six genomes.

Sulfur compound oxidation

The *sox* operon encoding enzymes for the oxidation of sulfur compounds is present in all the PS-superoperon-containing genomes, i.e. KT71, RAp1red, Ivo14 and HTCC2080 (Fig. 9), but not in HTCC2148 and HTCC2143. Among all the *sox* genes, *sox*CDXYZAB are the core genes for reducing thiosulfate (Friedrich, Bardischewsky et al. 2005). SoxYZ bind to the sulfur compounds; SoxAX are composed of two *c*-type cytochromes, which bind and reduce thiosulfate onto a cystein residue of SoxY (Quentmeier and Friedrich 2001); SoxB hydrolyzes off one sulfate molecule; and $Sox(CD)_2$ further oxidize the remaining sulfur atoms on SoxYZ, which are then further hydrolyzed by SoxB. The presence of *sox*CD is in accordance with the fact that no elemental sulfur deposits exist in the strains, since the $Sox(CD)_2$ function as sulfur dehydrogenase (Hensen, Sperling et al. 2006).

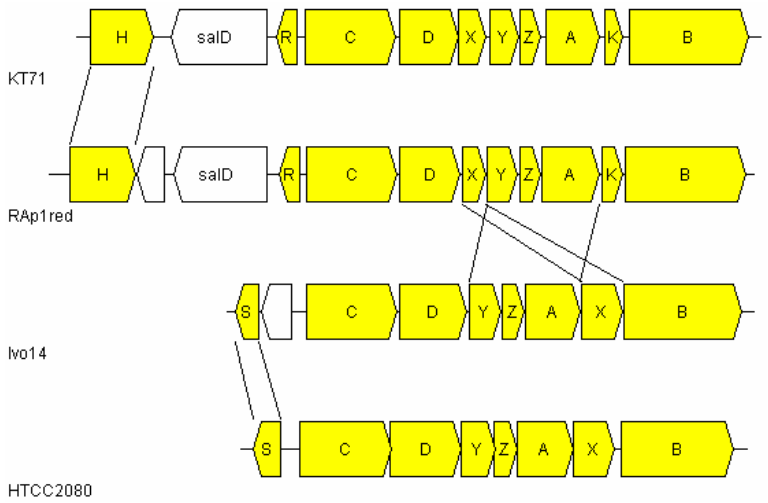

Fig. 9 Arrangement of *sox* operon in KT71, RAp1red, Ivo14 and HTCC2080 genomes. The *sox*X of KT71 and RAp1red show low similarity and different length with those of Ivo14 and HTCC2080.

The operon arrangement *sox*CDXYZAKB in KT71 and RAp1red is the same as in *Neptuniibacter caesariensis* (*Oceanospirillum sp.*) MED92 (*Gammaproteobacteria*), *Methylobacterium sp.* 4-46 and *Acidiphilium cryptum* JF-5 (both *Alphaproteobacteria*). Strain 4-46 contains also *sox*R in the same position.

The *sox* operon of Ivo14 and HTCC2080 differ to that in KT71 and RAp1red in several aspects: The *sox*K gene (SoxAX binding protein) (Ogawa, Furusawa et al. 2008) is absent in Ivo14 and HTCC2080. The *sox*X gene is partially homologous to the one in KT71 and RAp1red, but essentially longer (~200 bp) at the beginning. The *sox*D gene also shows relative low similarity between the two group. The arrangement of Ivo14 and HTCC2080, *sox*CDYZAXB, is the same as in *Dechloromonas aromatica* RCB, *Polynucleobacter sp.* QLW-P1DMWA-1 and *Polaromonas sp.* JS666 (all *Betaproteobacteria*). The gene between *sox*S and *sox*C in Ivo14 is annotated as DUF1791, which occurs often in *Betaproteobacteria*. Most genes in the *sox* operons of Ivo14 and HTCC2080 also have the closest relatives in *Beta*- and *Deltaproteobacteria* rather than *Gammaproteobacteria* (see e.g. *sox*B tree, Fig. 10). All these features suggest a lateral gene transfer of the whole *sox* operon from *Beta*- or *Deltaproteobacteria*.

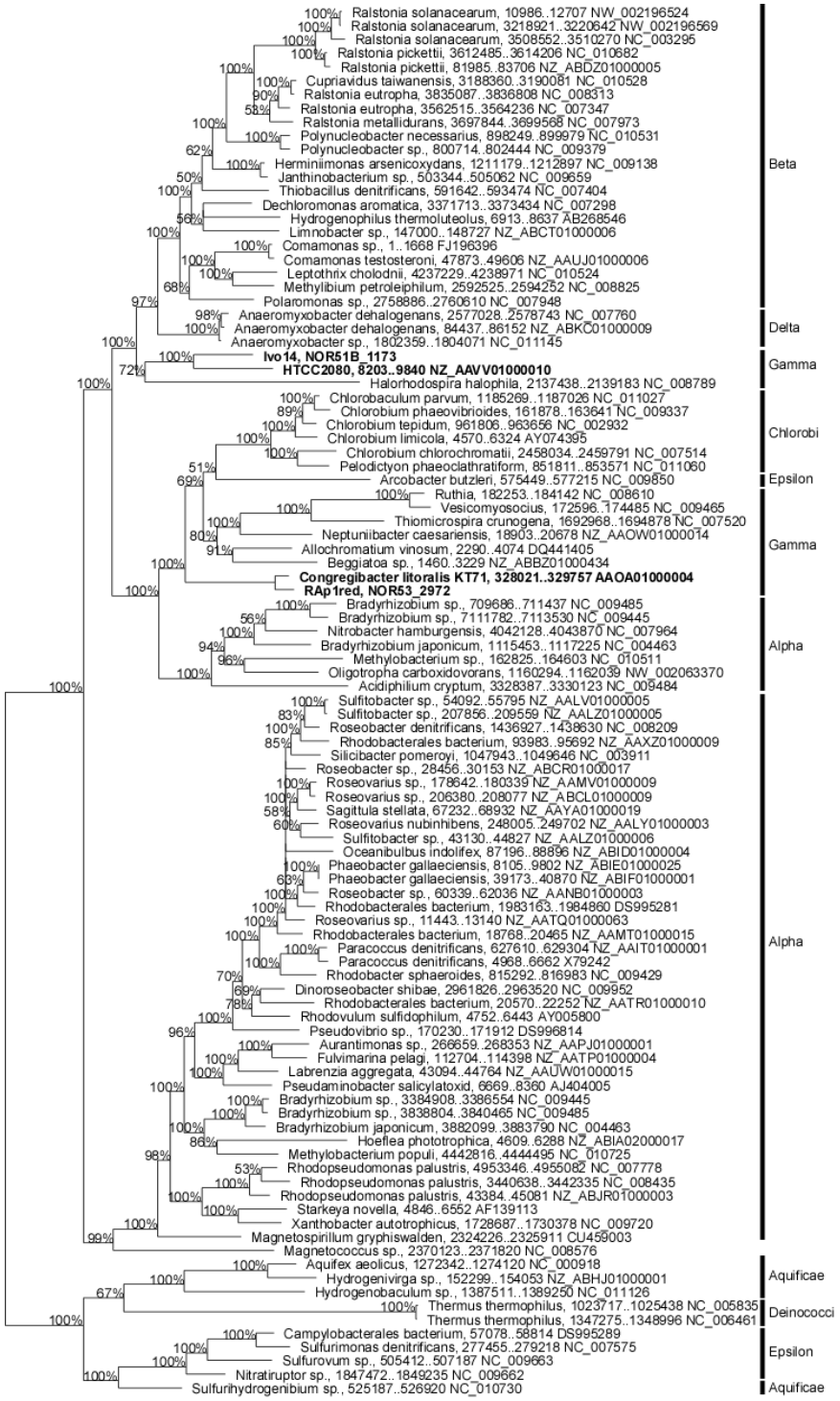

Fig. 10 Posterior probability tree of *sox*B gene from genomic sequences, using MrBayes, calculated from 528 valid amino acids. This tree shows nearly identical topology compared to the RAxML (maximum likelihood) tree (not shown).

Although other known *sox* genes, like *sox*E, F, G and V appear in some of the genomes, they are separated from the *sox* operon and usually have relatively low

similarity to the known sequences. Therefore they may have different functions. The sulfite reductase genes *dsr*AB are absent in all the six strains, so that they possibly lack the function for sulfate reduction.

Motility, pili formation

Four genomes, except for Ivo14 and HTCC2148, contain the whole set of genes needed for motility. The gene order of the three NOR5/OM60 strains is similar: *flg*NMA-?-*flg*BCDEFGHIJKL-?-*fli*C-*fla*G-*fli*DS-??-*fle*QSR-*fli*EFGHIJKLMNOPQR-*flh*BAF-*fle*N-*fli*A-*mot*AB-?-*mot*BA (here the "?" marks indicate homologous ORFs with unknown function), with all these genes grouped together in a superoperon. In HTCC2143, the genes up to *fle*R and the genes from *fli*E are separated into two large operons on two scaffolds, and the order for a few genes also varies.

The genes for pili formation, *pil*SR and *pil*MNOPQ exist in all the six genomes.

Housekeeping genes

All the six strains contain ATP synthase *atp*(I)BEFHAGDC in the same order in one operon. Four strains, except for Ivo14 and HTCC2080, contain *atp*I gene, which might function under low Mg^{2+} conditions (Hicks, Wang et al. 2003).

Most of the genes that constitute glycolysis/gluconeogenesis pathway exist in all the genomes. However, none of the genomes contains all the genes of the enzymes for both glycolysis and gluconeogenesis. All the five NOR5/OM60 genomes contain the gene of phosphofructokinase (*pfk*A) for glycolysis, while HTCC2143 contains the gene of the enzyme for its reverse reaction: fructose-1,6-bisphosphatase for gluconeogenesis, but no *pfk*A. The glucokinase (*glk*) for the first step of glucose consumption cannot be found in KT71, and accordingly KT71 is unable to use glucose (Fuchs, Spring et al. 2007), while the gene can be found in all the other five genomes. The genes for glycolysis/gluconeogenesis are dispersed in several sites on the genomes, without forming large operons.

All six strains contain the full sets of genes for pyruvate dehydrogenase complex, citric acid cycle (TCA cycle) and glyoxylate shunt. These genes are also dispersed across the genomes.

Redox activity and elemental metabolism

There are two groups of cytochrome *c* oxidases found in these genomes: *cta*CDGE in all the six genomes, and *fix*NOQP (or *cco*NOQP) in five genomes except HTCC2148. Catalase/peroxidase gene *kat*G exists in all the five NOR5/OM60 genomes, but not in HTCC2143. Superoxide dismutase (*sod*B) was found in KT71, RAp1red and Ivo14. A cluster of nickel-containing hydrogenase for hydrogen uptake was found in Ivo14, therefore it may also use hydrogen as electron donor. None of the molybdenum-, vanadium- or iron-containing nitrogenase genes for nitrogen fixation (Larimer, Chain et al. 2004) was found in any of the six genomes.

The selected gene groups discussed in this study are summarized in Table 3.

Table 3 Distribution of the genes and operons that are discussed in this study.

	KT71	**RAp1red**	**Ivo14**	**HTCC2080**	**HTCC2148**	**HTCC2143**
Phylogeny	NOR5-3	NOR5-3	NOR5-1B	NOR5-1B	NOR5-8	BD1-7
Photosynthesis superoperon	+	+	+	+	-	-
LHC II (*puc*)	+	+	-	-	-	-
Proteorhodopsin (*pop*)	-	-	-	-	+	+
3-hydroxypropionate pathway (*pcs*)	+	+	+	+	-	-
3-hydroxypropionate pathway (*mcr*)	-	+	+	+	-	-
Sulfur oxidation operon (*sox*)	+	+	+	+	-	-
Motility	+	+	-	+	-	+
Type IV pilus	+	+	+	+	+	+
ATP synthase	+	+	+	+	+	+
Glycolysis	-	+	+	+	+	-
Gluconeogenesis	-	-	-	-	-	+
Pyruvate dehydrogenase complex	+	+	+	+	+	+
Citric acid cycle	+	+	+	+	+	+
Glyoxylate shunt	+	+	+	+	+	+
aa3-type cyt *c* oxidase (*cta*CDGE)	+	+	+	+	+	+
cbb_3-type cyt *c* oxidase (*fix*NOQP)	+	+	+	+	-	+
Catalase (*kat*G)	+	+	+	+	+	-
Superoxide dismutase (Fe) (*sod*B)	+	+	+	-	-	-
Nickel-dependent hydrogenase	-	-	+	-	-	-
Nitrogenase (*nif*D)	-	-	-	-	-	-

Discussion

In the strain KT71, the expression of bacteriochlorophyll *a* (BChl *a*) and carotenoid, which are encoded by PS superoperon, has been proved with HPLC analysis (Fuchs, Spring et al. 2007). Significant amount of BChl *a* was synthesized when growing under light in oligotrophic medium. Enhanced cell yield of KT71 under light could be observed. Since the full PS superoperons with high sequence similarity and nearly

identical arrangement were found in four of the genomes, it can be speculated that utilization of light might be a common trait for the NOR5/OM60 group. This needs to be tested for more strains under various culturing conditions.

Surprisingly, although the *puf*LM genes of HTCC2148 were reported to be amplified and shown to be clustered with "α-1 group" (Cho, Stapels et al. 2007), they were not found in the genome of HTCC2148. Since we found no indication of the whole PS superoperon, we assume that the HTCC2148 culture used by Cho et al. might have been contaminated by other photosynthetic bacteria. This discrepancy might also be due to missing of the whole PS superoperon during the genome shotgun sequencing, or the PS superoperon was deleted from the strain by a mutation event after the amplification of *puf* genes and before genome sequencing. However, the possibility of missing of the whole PS superoperon from HTCC2148 during genome sequencing is not high. Several groups of housekeeping genes, which are located separately on the genome, can all be found, thus it seems that the coverage of the HTCC2148 genome was very high.

Since the PS superoperon exists in four strains and in two subclades of the NOR5/OM60 group and both the gene arrangement and gene sequences indicate no sign for a main lateral gene transfer event, photosynthesis might be an intrinsic common trait for the NOR5/OM60 group inherited from their common ancestor, and thus differentiate them from the other *Gammaproteobacteria*. More strains should be tested for the existence of PS superoperon in order to prove this hypothesis. Comparing the *puf*M and 16S rRNA trees (Fig. 5), supposing that the topology inside *puf*M Group K (Yutin, Suzuki et al. 2007) and 16S rRNA of the NOR5/OM60 bacteria are fully parallel, we assume that the many *puf*M environmental sequences such as EBAC000-29C02 probably originate from the NOR5-1A/C subclade.

In this study, the existence of proteorhodopsin in HTCC2143 is convincing. It was supported by the high sequence similarity of *pop* gene to that of the physiologically tested SAR92 clade, as well as the identical gene arrangement of retinal synthetic genes. All the genomes of NOR5-3 and NOR5-1B clades investigated in this study do not contain the gene for proteorhodopsin. The existence of proteorhodopsin in HTCC2148 is not fully proved. The absence of the retinal synthetic genes in HTCC2148 makes the occurrence of the *pop* gene in the HTCC2148 genome questionable. It is possible that the proteo-

rhodopsin is not functional, or has other functions. It is also possible that it is contamination. However, since the genome size of HTCC2148 is comparable to the other strains, it is not likely that another bacterial genome is contaminating. Either a re-sequencing of the genome or a PCR amplification of *pop* gene from the strain is needed to answer the question whether the existence of *pop* gene in HTCC2148 genome is true. Although it was observed that for the rhodopsin-containing flavobacterial strain *Dokdonia* MED134, light has enhanced growth markedly, however, in the proteo-rhodopsin-containing alphaproteobacterial strain *Pelagibacter ubique* HTCC1062 and gammaproteobacterial strain HTCC2207, no light-enhanced growth was observed (Giovannoni, Bibbs et al. 2005; Stingl, Desiderio et al. 2007). Since the *pop* sequence of HTCC2143 and HTCC2148 are phylogenetically closer to the other proteobacterial sequences, they may also not utilize proteorhodopsin for cell growth, but rather for an unknown function.

If the existence of proteorhodopsin in HTCC2148 can be proved, the distribution of PS superoperon and proteorhodopsin in the six genomes are complementary. The PS superoperon is large, and therefore relatively expensive for the bacteria to maintain, some bacteria may introduce rhodopsin as an alternative for retaining some sort of phototrophy in order to compensate the loss of PS superoperon. This hypothesis can be supported by the fact that the *puf* and rhodopsin genes rarely co-exist in the same genome. Only three prokaryote genomes known to-date contain both *puf* and rhodopsin genes: cyanobacterial *Nostoc* sp. PCC7120, alphaproteobacterial *Methylobacterium* sp. 4-46 and chloroflexal *Roseiflexus* sp. RS-1. All the three rhodopsin sequences are distantly related to the proteobacterial rhodopsins and possibly have different functions.

KT71 was not able to grow autotrophically in physiological tests (Fuchs, Spring et al. 2007). This is in accordance with the fact that the *mcr* gene was not found in its genome and thus the 3-hydroxypropionate pathway is incomplete. The reason why it still keeps the huge *pcs* gene is not clear yet. The only other example that contains *mcr* and *pcs* genes of 3-hydroxypropionate pathway rather than the NOR5/OM60 group and green sulfur and non-sulfur bacteria is the alphaproteobacterial AAnP *Erythrobacter sp.* NAP1, which was proved to be able to assimilate CO_2 (Kolber, Plumley et al. 2001). The daily

cellular CO_2 fixation rate was 3% of the cellular carbon content and contributed to about 1% of total carbon anabolism.

The last steps of 3-hydroxypropionate cycle in *Chloroflexus* are more complicated than previously thought and are still under investigation (Friedmann, Alber et al. 2007). For the supposed succinyl-CoA:L-malate CoA transferase and L-malyl-CoA lyase, homologs with relatively low similarity to those in *Chloroflexus* can be found in the NOR5/OM60 genomes, and it is hard to judge whether the NOR5/OM60 strains use these enzymes to close the cycle. On the other hand it is possible that the NOR5/OM60 strains may use a different pathway to recycle succinyl-CoA and to regenerate acetyl-CoA. For example, some enzymes in citric acid cycle or glyoxylate shunt might be involved. Physiologic experiments need to be done on the NOR5/OM60 cultures in order to know whether they are able to fix CO_2, which intermediates are produced, and how the pathway is constructed.

Since it is the first time that the *pcs* gene is found in *Gammaproteobacteria*, we searched for its homologous sequence using BLAST against metagenomic databases. Hundreds of homologous sequences were found from the Global Ocean Survey (GOS) project (http://camera.calit2.net/index.php) (Rusch, Halpern et al. 2007), of which many are obviously more similar to the sequences of the NOR5/OM60 strains than to the other groups (e-values differentiate more than 10^{30} times). The sampling locations are also widely distributed. Therefore the 3-hydroxypropionate pathway might be a common route for carbon fixation in the surface layer of the ocean, and more studies in detail are expected to determine if they belong to the NOR5/OM60 group.

The same distribution of PS superoperon, 3-hydroxypropionate pathway as well as *sox* operon provides the possibility that some members of the NOR5/OM60 group (including NOR5-3 and NOR5-1B subclades) might be able to oxidize sulfur compounds and reduce and fix CO_2 using light energy. However, none of the three points were yet proved in the first physiological tests of KT71 (Fuchs, Spring et al. 2007). If these points can be experimentally proved in the other strains of the NOR5/OM60 group, it might lead to the discovery of an important style of photoautotrophy in the ocean. These three sets of genes do not seem to have been acquired at the same time during the evolution. Through the observation of topology of the phylogenetic trees of *puf*M (Fig. 5), *pcs* (Fig. 8) and

*sox*B (Fig. 10) trees as well as the gene arrangement in the operons, we can postulate that the PS superoperon is intrinsic to the NOR5/OM60 group, while *sox* operon have been gained by the NOR5-3 and NOR5-1B strains during two separate events, and for *pcs-mcr* genes the gene transfer is more complicated: these genes of Ivo14 might have been lost and then re-gained from its geographic neighbor from the NOR5-3 subclade.

The genes for flagella were found in four genomes (including KT71, HTCC2080 and HTCC2143 from coastal water and RAp1red from coastal surface sediment). KT71 is microaerophilic and actively swim to a layer with ≈10% oxygen saturation in an agar shake experiment (Fuchs, Spring et al. 2007). Thus the flagella enable the bacterium to swim to a proper niche for light, oxygen and nutrients in the coastal region where frequent mixing of water and sediment occurs. The cbb_3-type cytochrome *c* genes *fix*NOQP exist in all the five coastal strains (except HTCC2148). This oxidase with high oxygen affinity that is expressed only under low oxygen conditions in *Bradyrhizobium japonicum* (Preisig, Anthamatten et al. 1993) fits the need for the coastal strains that they might be temporally buried in the microaerobic sediment.

Considering its potential of utilizing light and fixing CO_2, as well as the high abundance (level of several percent) in many coastal waters and sediments, the NOR5/OM60 clade should be further studied. Additional comparative genomics as well as physiological tests on the strains are needed for solving these questions and to address the role of NOR5/OM60 in the marine carbon cycle.

Acknowledgement

The authors would like to thank to Gordon and Betty Moore Foundation for financing the sequencing and the J. Craig Venter Institute for carrying out sequencing and annotation, and Microbial Genomics Group of Max Planck Institute for Marine Microbiology for technical supports. This work is part of Shi Yan's PhD thesis in the framework of the MarMic program, financed by the Max Planck Society.

References

Alber, B., M. Olinger, et al. (2006). "Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal Metallosphaera and Sulfolobus spp." Journal of Bacteriology **188**(24): 8551-8559.

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**(17): 3389-3402.

Béjà, O., L. Aravind, et al. (2000). "Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea." Science **289**(5486): 1902-1906.

Béjà, O., E. N. Spudich, et al. (2001). "Proteorhodopsin phototrophy in the ocean." Nature **411**(6839): 786-789.

Cho, J. C. and S. J. Giovannoni (2004). "Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria." Applied and Environmental Microbiology **70**(1): 432-440.

Cho, J. C., M. D. Stapels, et al. (2007). "Polyphyletic photosynthetic reaction centre genes in oligotrophic marine Gammaproteobacteria." Environmental Microbiology **9**(6): 1456-1463.

Delcher, A. L., A. Phillippy, et al. (2002). "Fast algorithms for large-scale genome alignment and comparison." Nucleic Acids Research **30**(11): 2478-2483.

Eilers, H., J. Pernthaler, et al. (2001). "Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton." Applied and Environmental Microbiology **67**(11): 5134-5142.

Friedmann, S., B. E. Alber, et al. (2007). "Properties of R-citramalyl-coenzyme A lyase and its role in the autotrophic 3-hydroxypropionate cycle of Chloroflexus aurantiacus." Journal of Bacteriology **189**(7): 2906-2914.

Friedrich, C. G., F. Bardischewsky, et al. (2005). "Prokaryotic sulfur oxidation." Current Opinion in Microbiology **8**(3): 253-259.

Fuchs, B. M., S. Spring, et al. (2007). "Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis." Proceedings of the National Academy of Sciences of the United States of America **104**(8): 2891-2896.

Fuhrman, J. A., M. S. Schwalbach, et al. (2008). "Opinion - Proteorhodopsins: an array of physiological roles?" Nature Reviews Microbiology **6**(6): 488-494.

Giovannoni, S. J., L. Bibbs, et al. (2005). "Proteorhodopsin in the ubiquitous marine bacterium SAR11." Nature **438**(7064): 82-85.

Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.

Hensen, D., D. Sperling, et al. (2006). "Thiosulphate oxidation in the phototrophic sulphur bacterium Allochromatium vinosum." Molecular Microbiology **62**(3): 794-810.

Hicks, D. B., Z. X. Wang, et al. (2003). "A tenth atp gene and the conserved atpl gene of a Bacillus atp operon have a role in Mg2+ uptake." Proceedings of the National Academy of Sciences of the United States of America **100**(18): 10213-10218.

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-755.

Hügler, M., C. Menendez, et al. (2002). "Malonyl-coenzyme A reductase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation." Journal of Bacteriology **184**(9): 2404-2410.

Katoh, K. and H. Toh (2008). "Recent developments in the MAFFT multiple sequence alignment program." Briefings in Bioinformatics **9**(4): 286-298.

Kolber, Z. S., F. G. Plumley, et al. (2001). "Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean." Science **292**(5526): 2492-2495.

Kolber, Z. S., C. L. van Dover, et al. (2000). "Bacterial photosynthesis in surface waters of the open ocean." Nature **407**(6801): 177-179.

Kurtz, S., A. Phillippy, et al. (2004). "Versatile and open software for comparing large genomes." Genome Biology **5**(2).

Larimer, F. W., P. Chain, et al. (2004). "Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris." Nature Biotechnology **22**(1): 55-61.

Ludwig, W., O. Strunk, et al. (2004). "ARB: a software environment for sequence data." Nucleic Acids Research **32**(4): 1363-1371.

Man, D. L., W. W. Wang, et al. (2003). "Diversification and spectral tuning in marine proteorhodopsins." Embo Journal **22**(8): 1725-1731.

McCarren, J. and E. F. DeLong (2007). "Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla." Environmental Microbiology **9**: 846-858.

Meyer, F., A. Goesmann, et al. (2003). "GenDB - an open source genome annotation system for prokaryote genomes." Nucleic Acids Research **31**(8): 2187-2195.

Ogawa, T., T. Furusawa, et al. (2008). "SoxAX binding protein, a novel component of the thiosulfate-oxidizing multienzyme system in the green sulfur bacterium Chlorobium tepidum." Journal of Bacteriology **190**(18): 6097-6110.

Preisig, O., D. Anthamatten, et al. (1993). "Genes for a Microaerobically Induced Oxidase Complex in Bradyrhizobium-Japonicum Are Essential for a Nitrogen-Fixing Endosymbiosis." Proceedings of the National Academy of Sciences of the United States of America **90**(8): 3309-3313.

Quentmeier, A. and C. G. Friedrich (2001). "The cysteine residue of the SoxY protein as the active site of protein-bound sulfur oxidation of Paracoccus pantotrophus GB17." Febs Letters **503**(2-3): 168-172.

Richter, M., T. Lombardot, et al. (2008). "JCoast - A biologist-centric software tool for data mining and comparison of prokaryotic (meta) genomes." BMC Bioinformatics **9**: Article No. 177.

Rusch, D. B., A. L. Halpern, et al. (2007). "The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific." Plos Biology **5**(3): 398-431.

Stamatakis, A. (2006). "RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models." Bioinformatics **22**(21): 2688-2690.

Stamatakis, A., P. Hoover, et al. (2008). "A Rapid Bootstrap Algorithm for the RAxML Web Servers." Systematic Biology **57**(5): 758-771.

Stingl, U., R. A. Desiderio, et al. (2007). "The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin." Applied and Environmental Microbiology **73**(7): 2290-2296.

Thompson, J. D., D. G. Higgins, et al. (1994). "Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Research **22**(22): 4673-4680.

Yan, S., B. M. Fuchs, et al. (2009). "Biogeography and phylogeny of the NOR5/OM60 clade of Gammaproteobacteria." Systematic and Applied Microbiology **32**(2): 124-139.

Yurkov, V. V. and J. T. Beatty (1998). "Aerobic anoxygenic phototrophic bacteria." Microbiology and Molecular Biology Reviews **62**(3): 695-+.

Yutin, N. and O. Béjà (2005). "Putative novel photosynthetic reaction centre organizations in marine aerobic anoxygenic photosynthetic bacteria: insights from metagenomics and environmental genomics." Environmental Microbiology **7**(12): 2027-2033.

Yutin, N., M. T. Suzuki, et al. (2005). "Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs." Applied and Environmental Microbiology **71**(12): 8958-8962.

Yutin, N., M. T. Suzuki, et al. (2007). "Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes." Environmental Microbiology **9**(6): 1464-1475.

Unit 3

Characterization of the NOR5/OM60 strains from the North Sea

(In preparation)

Introduction

Twenty-two bacterial strains were isolated from the North Sea coastal sediment in the year 2005, in an attempt for isolating *Planctomycetes* strains. These strains are characterized with red or pinkish color and slow growth rate. Through 16S rRNA sequence analysis, they were found to be members of the NOR5/OM60 clade of *Gamma-proteobacteria* and closely related to the previously well-studied strain KT71 (Yan et al., 2009), which was isolated from the island Helgoland, and proved to be the first discovered gammaproteobacterial aerobic anoxygenic phototroph (AAnP) (Fuchs et al., 2007). The NOR5/OM60 members occur in high number in the North Sea water column and sediment. In the surface water of the North Sea, the yearly NOR5/OM60 percentage by DAPI counts varies between 0.2 and 2.8% (Keller, 2003), while 8% (Eilers et al., 2001) and even 11% (Pernthaler and Pernthaler, 2005) of DAPI counts were also reported. The strains from Sylt together with strain KT71 were preserved for investigations.

In this work, the North Sea strains were preliminarily studied through culturing, physiology and genomics. In order to know whether the strains with identical 16S rRNA sequences are identical in the genome level, PFGE fingerprint was done to compare the patterns. The pigments from the cultures were extracted in order to study the composition, especially, whether photosynthetic pigments, such as bacteriochlorophyll *a* (BChl *a*), are expressed.

Materials and methods

Strain sources

In attempt to isolate *Planctomycetes* strains, J. Harder and students of the MarMic class 2009 sampled the coastal marine sediment at the North Sea island Sylt in October, 2006. The samples were sandy or muddy, aerobic or anaerobic (Unit 2, Table 1). They were diluted with sterilized sea water, and streaked on the "PLA1-rich" plate (Table 1, Harder, personal communication), which contained two antibiotics: cycloheximide for inhibiting growth of fungi, and ampicillin for inhibiting most of the bacteria which contain peptidoglycan and do not contain genes for penicillin-resistance. Since the *Planctomycetes* do not have a peptidoglycan-containing cell wall, the "PLA1-rich" medium was supposed to be able to select *Planctomycetes* strains.

Table 1 The PLA1-rich medium

in 1L of water:	
NaCl	26.37 g
$MgCl_2$ $6H_2O$	5.67 g
$MgSO_4$ $7H_2O$	6.8 g
$NaHCO_3$	0.19 g
$CaCl_2$ $2H_2O$	1.47 g
KCl	0.72 g
KBr	0.1 g
H_3BO_3	0.02 g
$SrCl_2$	0.02 g
NaF	0.003 g
Add 50 mM Tris to pH 7.5	
Carbon Sources:	
Yeast extract	0.25 g
Peptone	0.25 g
Glucose	0.25 g
50 g L^{-1} NH_4Cl	(10 / 0) mL
50 g L^{-1} KH_2PO_4	10 mL
Trace element solution	2 mL
Se-W-solution	1 mL
7-vitamine solution	(1 / 11) mL
Vitamin-B12 solution	(1 / 0) mL
Riboflavine solution	1 mL
Thiamine solution	(1 / 12) mL
Ampicillin	2 g
Cycloheximide	2 g
pH 7.5	
Agar	15 g

Strain cultivation

On the PLA1-rich plates, several red or pink colonies grew up after a few days. According to the results of 16S rRNA sequencing, many of them were categorized as NOR5/OM60 members. These isolates were transferred to liquid or plates of SYPG rich medium (Table 2). The strains were kept on the agar plates by transferring once every three weeks or one month. On the plates, the colonies can be visible in 7 – 15 days after transfer. The liquid cultures were usually kept in 50 or 200 mL plastic bottles for cell culturing (Greiner), filled with 1/4 – 1/3 volume of SYPG liquid media and normal air.

Table 2 The SYPG medium

in 1L of water:	
Sigma sea salts	30 g
Yeast extract	0.5 g
Peptone, triptic digest	0.25 g
Sodium glutamate	0.1 g
pH 7.5 – 8.0	
for agar plates:	
Agar	15 g

Pulse field gel electrophoresis (PFGE)

The PFGE were run for the 12 strains of which 16S rRNA sequences are identical, including Ivo10red, Ivo11, Ivo19, Mel5, Mel7, Pao12, RAp1red, RAp2, RAP5, RAp6, RAp9 and RAp11. KT71 was used as control. From two-month-old liquid cultures, 20 – 25 mL was collected for making plugs. Restriction digest was done with endonuclease *Swa*I (BioLabs), which recognizes the signature 5′-ATTT|AAAT-3′. In KT71 genome, this signature occurs 14 times, and separates the genome into fragments of 25 kbp – 1.17 Mbp.

The PFGE was done with the above mentioned 13 strains and two markers: lambda DNA-*Hin*dIII fragments and Yeast chromosome PFG marker (both BioLabs). The setting was: initial switch time: 1.2 s, final switch time: 68.7 s, run time: 23.9 h, voltage: 6 V, angle: 120°. This setting was used for separating genomic fragments of 15 – 750 kbp.

Pigment extraction

The cell pellets were obtained by either centrifugation of liquid culture, or directly picking using pipette tips from colonies on the agar plates. The pellets were re-suspended in 300 µL of extraction mixture of aceton:methanol = 7:2, and stored in dark under 4°C, under the protection of nitrogen gas.

The extractions went through high performance liquid chromatography (HPLC) by the washing solvent of acetonitril:methanol:tetrahydrofurane = 15:3:2. For each peak, absorption of wavelength range of 300 – 800 nm was measured.

Results and discussion

Pulse field gel electrophoresis (PFGE)

The result of PFGE is shown in Figure 1. The result showed that all the strains, for which fingerprint could be obtained had different band patterns. Therefore it indicates that all the North Sea strains, in spite of their 16S rRNA identity, are genetically different. This can be also supported by the different cell morphology, growth and colony forms among those 16S-rRNA-identical strains.

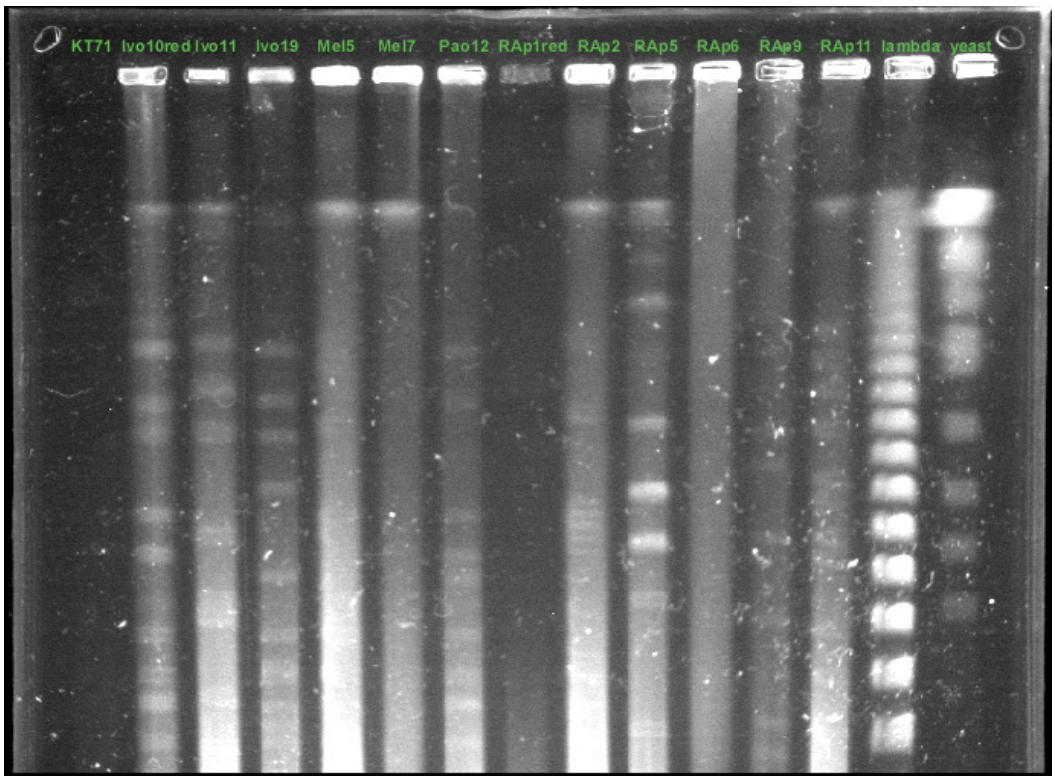

Figure 1 PFGE picture of KT71 and the 12 strains, of which 16S rRNA sequences are identical. The genome DNA samples were cut by the endonuclease *Swa*I.

Pigment extraction

The results of HPLC (Table 3) showed that a pigment composition (with 3 highest peaks at the 363.7, 753.1 and 589.3 nm), which is possibly the bacteriochlorophyll *a* (Bchl *a*), was found in the extract of strains Ivo10red and Mo12red (both NOR5-3), but not in the strains KT71 and Mo10red (both NOR5-3), RAp14red (NOR5-3/1B), Mo4 and Ivo14 (both NOR5-1B). However, since the BChl *a* has been proved to be expressed

from the culture of KT71 (Fuchs et al., 2007), the expression must be influenced by cultivation conditions. All the other four strains, in which BChl *a* was not found to be expressed, also contain the *puf*LM genes. A carotenoid-like compound (retention time 12.8 min), of which absorption maxima are very close to spirilloxanthin in *Roseateles depolymerans* (Suyama et al., 1999), could be found in strains Ivo10red, Mo12red and RAp14red in great amount, which has also been shown existing in KT71 (Fuchs et al., 2007).

Table 3 HPLC results of 7 strains of NOR5/OM60 group

	Absorption (nm)			NOR5-3				NOR5-1B		
Retention time (min)	Peak 1	Peak 2	Peak 3	Ivo10 red	Mo12 red	Mo10 red	KT71	RAp14 red	Mo4	Ivo14
12.6	439.9	486.9	552.5					+		
12.8 (spirillo-xanthin)	494.9	531.2		+++	+++			+++	+	
12.9	363.7			++	++			++	?	
13.4	439.9			+				+		
13.6 (BChl *a*)	363.7	753.1	589.3	+++	+++					
13.6	360.6							++	++	
14.1	363.7			+						
16.8	354.6			++	++			++	++	++
18.9	404.6			++		+	+	+	++	

Acknowledgement

Here I address special thanks to Birgit H. Söller for helping me with the HPLC analysis for the pigment of the cells. I also thank Stefan Spring for suggestions about cultivation and pigment extraction, and Jörg Wulf for instructions of PFGE.

References

Eilers, H., Pernthaler, J., Peplies, J., Glöckner, F.O., Gerdts, G., and Amann, R. (2001) Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton. *Appl Environ Microbiol* **67**: 5134-5142.

Fuchs, B.M., Spring, S., Teeling, H., Quast, C., Wulf, J., Schattenhofer, M. et al. (2007) Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *P Natl Acad Sci USA* **104**: 2891-2896.

Keller, L. (2003) Herbstsukzessionen und Aktivität der pelagischen Bakteriengemeinschaft in der Deutschen Bucht. In *Max-Planck-Institut für Marine Mikrobiologie*. Bremen, p. 89.

Pernthaler, A., and Pernthaler, J. (2005) Diurnal variation of cell proliferation in three bacterial taxa from coastal North Sea waters. *Appl Environ Microbiol* **71**: 4638-4644.

Suyama, T., Shigematsu, T., Takaichi, S., Nodasaka, Y., Fujikawa, S., Hosoya, H. et al. (1999) Roseateles depolymerans gen, nov., sp. nov., a new bacteriochlorophyll a-containing obligate aerobe belonging to the beta-subclass of the Proteobacteria. *International Journal of Systematic Bacteriology* **49**: 449-457.

Yan, S., Fuchs, B.M., Lenk, S., Harder, J., Wulf, J., Jiao, N.Z., and Amann, R. (2009) Biogeography and phylogeny of the NOR5/OM60 clade of Gammaproteobacteria. *Syst Appl Microbiol* **32**: 124-139.

Acknowledgement

First I want to express my greatest thank to Rudi AMANN. As my supervisor, he is the person I can fully trust. I am grateful that he gave me the chance to work in this nice department – Molecular Ecology. His revision for the publications and thesis is always clear, careful and without delay. He is my exemplar for becoming a good scholar, a good leader and a good person.

I thank very much to my direct supervisor, Bernhard FUCHS. He gave me all kinds of daily instructions for both scientific and technical questions, and he scrutinized and corrected my publications and thesis. It is a great pleasure to work with him.

Many thanks to Prof. JIAO Nianzhi, our cooperation partner in China and member of my PhD thesis committee. Benefited from his generous invitation to his Xiamen, I did some samplings in China and acquainted with many friends from his students and co-workers. I thank many colleagues in his group as well, especially ZHANG Yao and ZHANG Fan for the organizing the cruise at Yangtze estuary, and LIU Yongqin from the Institute of Tibetan Plateau Research of Chinese Academy of Science.

Thank to Stefan SPRING, member of my committee thesis. He gave me many suggestions for the physiological tests as well clues for gene finding.

Jörg WULF, our chef for the lab and my office-mate, he knows everything in the lab, and is always ready for help, whenever I have troubles in the lab. Marc MUßMANN, who also shared the office with me, helped me with many instructions and by revising the paper. Sabine LENK, my classmate and office-mate, has helped me with the data of my first publication and many discussions as well. They are my best friends for both in and outside the work.

Thanks very much to Prof. Ulrich FISCHER for becoming reviewer of my thesis in a shortened time. And special thank to Birgit H. SÖLLER, who helped me with pigment analysis of the strains.

I also want to thank all the colleagues who work or have worked in the department of Molecular Ecology, including the Genomic group, especially Ulrike BUCK, Jill PETERSEN, Cristina MORARU, Silke WETZEL, Lisa KEMP, Ilaria PIZETTI, Paola GOMEZ,

Birgit RATTHUNDE, Sabine KÜHN, HUANG Sixing, KUBOTA Kengo, Ivo KONSTADINOV, Renzo KOTTMANN and Shalin SEEBAH. I am sorry not able to name all the people here that have helped me. They supported me in works, in daily life, and spiritually as well.

I thank all the MarMic staff and fellow students. At the first place is Jens HARDER, who offered the North Sea strains as well as many useful data and experimental instructions. And Christiane GLÖCKNER and Karl-Heinz BLOTEVOGEL, for organizing the MarMic program and helped me through all the procedures during my PhD.

I acknowledge the Max Planck Society (MPG) for supporting the MarMic program and my researches, and German Academic Exchange Service (DAAD) and China Scholarship Council (CSC) for funding the project-based personal exchange program (PPP) with Xiamen University.

Last but not least, I thank my dearest parents, and all my friends in Bremen. They are always supporting me.

Erklärung
gem. § 6 (5) Nr. 1 – 3 PromO

Ich erklärte, dass ich

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

Bremen, den 16. Februar, 2009

YAN Shi