University of Bremen
International Max Planck Research School for Marine Microbiology

GENOMIC ANALYSIS OF THE ENDO SYMBIOTIC COMMUNITY OF A GUTLESS MARINE WORM (OLAVIUS ALGARVENSIS)

by
Amelia Elena Rotaru

Supervised
by
Nicole Dubilier

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Marine Microbiology

Max Planck Institute for Marine Microbiology,
28359 Bremen, Germany
April, 2005
Erklärung

Statement

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

I herewith confirm that I have written this thesis unaided and that I used no other resources than those mentioned.

______________________________  ______________________________
(Ort und Datum / Place and Date)  (Unterschrift / Signature)
Table of content

Summary
Chapter 1: Introduction
  1.1. General introduction to marine oligochaetes symbioses
  1.2. Molecular identification and phylogeny of O. algarvensis endosymbionts
  1.3. Function of O. algarvensis symbionts
  1.4. Metagenomics: microbial community genome analyses
Chapter 2: Materials and methods
  2.1. Symbionts DNA isolation and BAC library construction
  2.2. Plasmid preparation
  2.3. BAC insert size determination via pulsed field gel electrophoresis
  2.4. BAC sequencing and primer walking
  2.5. Primer design
  2.6. Sequence analysis
Chapter 3: Results
  3.1. Symbionts DNA isolation and BAC library construction
  3.2. Sequence analysis and primer walking
Chapter 4: Discussions
  4.1. Carbon metabolism
  4.2. Sulfur metabolism
  4.3. Nitrogen metabolism
  4.4. Transport proteins
  4.5. Pathogenic / symbiotic associated traits
Annexes
Acknowledgements
Summary

The gutless marine worm, *Olavius algarvensis*, harbors a microbial community of five co-occurring symbionts belonging to sulfide-oxidizers, sulfate-reducers, and spirochetes (Rühland et al., In Prep.). In order to get access to the genetic capabilities of these symbionts a metagenomic approach was used. Previously performed clone insert end-sequencing of the entire BAC library had revealed sequences that are considered to be relevant for the symbiotic community. In this study clones carrying interesting genes were further analyzed using primer walking, in order to proceed further into the insert. For the obtained contiguous sequences similarities were searched against the non-redundant GenBank protein database applying the BLASTx algorithm.

Several sequences attracted special attention because their deduced amino acid sequences exhibited a strong similarity to proteins involved in both oxidative and reductive sulfur metabolisms. This indicates that the analyzed genomic fragments are derived from sulfur-oxidizing and sulfate-reducing symbionts. Some of the genes offer insights into the symbionts metabolic pathways that are presently not known. Particularly the degradation of urea, the respiration of nitrate, one carbon metabolism, and the synthesis of cyanophycin demand attention in the near future. Genes putatively encoding virulence associated proteins can provide information on how the association between bacteria and oligochaetes has developed and how bacteria adapt to host dependent lifestyle.

Although it has to be taken into consideration that the assigned functions by similarity search is potential and that the actual role can be only validated experimentally, the sequence information of the bacterial symbionts from *O. algarvensis*, obtained in this work, offers the opportunity to examine the metabolic pathways and genetic interactions within the symbiotic bacterial consortium in more detail, expanding the understanding of this symbiosis.
1. Chapter 1: Introduction

1.1. General introduction to marine oligochaete symbioses

Symbiosis, an interdependent relationship between two species is an important driver of evolutionary novelty and ecological diversity (Margulis and Chapman, 1998; Wernegreen, 2004). The ties between symbiotic partners range from rather loose and occasional associations over regular ectosymbiosis to obligatory incorporation into the host body. While in ectosymbiosis the bacteria occur externally on the host surface, in endosymbiosis the microbial partner lives within the host either extra- or intracellularly. Symbioses are diverse and ubiquitous in aquatic and terrestrial habitats, and are particularly prominent in marine ecosystems, such as coral reefs, hydrothermal vents or shallow waters. The first symbiotic association between chemoautotrophic sulfur-oxidizing bacteria and marine invertebrates was proposed for the tube worm *Riftia pachyptila*, inhabitant of the hydrothermal vent fauna (Cavanaugh et al., 1981; Felbeck, 1981).

Within the annelid class Oligochaeta two genera of marine Tubificidae, *Inanidrilus* and *Olavius*, are characterized by obligate endosymbiosis with extracellular bacteria. Gutless marine oligochaetes are abundant in marine sediments through the tropics and subtropics of the world and approximately 80 species have been described (Erseus, 1979; Felbeck, 1981; Giere et al., 1998; Erseus, 2003). These marine worms are lacking completely both their digestive and excretory systems having no gut, mouth or nephridia. Given the complete reduction of digestive and excretory system an obligate association with autotrophic bacteria is clearly obligate for the oligochaete partner (Giere et al., 1995; Dubilier et al., 2001).

*Olavius algarvensis* was first found in the subtidal sands of Algarve, Portugal and described as a small tubificid worm (0.1 mm x 15-25 mm) found in subsurface layers at depths of 10-15 cm (Giere et al., 1998). Later on, *O. algarvensis* was also found off the coast of Elba, Italy in coarse-grained sands surrounding sea grass (Dubilier et al., 2001). *O. algarvensis* endosymbiotic community was under focus, particularly the bacterial mutualistic relationships in addition to their symbiotic relation to the host ((Dubilier et al., 2001).

Ultrastructural studies were performed on *O. algarvensis* and was observed a similar
arrangement of the symbionts as in other gutless species, occurring in a multicellular layer in
the subcuticular space between the cuticle and epidermis, called the symbiotic region (Giere
et al., 1998; Dubilier et al., 2001). Transmission electron micrography (TEM) revealed two
morphotypes: i) a large oval morphotype of 3-5 µm diameter with numerous intracellular
inclusions, ii) a smaller rod-shaped morphotype of 0.5-1 µm diameter without any inclusions
(Giere et al., 1995; Dubilier et al., 2001).

1.2 Molecular identification and phylogeny of *O. algarvensis* endosymbionts

In order to study the phylogeny of the bacterial symbionts the 16S ribosomal RNA gene was
used as a phylogenetic anchor. The full cycle rRNA approach (Amann et al., 1995) has been
used to correlate the morphotype of the symbionts with their phylotype. This approach
combines comparative rRNA sequence analysis with fluorescence in situ hybridization
(FISH).

The full cycle rRNA approach revealed five endosymbiotic phylotypes in *O. algarvensis*. The large bacterial morphotype, with sulfur globules and polyhydroxybutyrate (PHB)
vesicles has been identified as a Gamma 1 phylotype and is closely related to the
endosymbionts of other marine oligochaetes (e.g. *O. crassitunicatus*, *O. ilvae*, *I.
leukodermatus*). These symbionts are distributed throughout the entire symbiotic region
positioned below the thin cuticle of the worm. The second gamma proteobacterial symbiont
is closely related to clone sequences from cold-seep communities form the Japan Trench (Li
et al., 1999c) and FISH analyses showed that corresponded to the smaller symbiotic
morphotype (Rühland et al., In Prep.; Dubilier et al., In press).

Two delta proteobacterial symbionts were found in *O. algarvensis*, both belonging to
*Desulfococcus/Desulfosarcina/Desulfonema* subgroup of sulfate-reducing bacteria. In situ
hybridization studies showed that these symbionts are small and cocci shaped and they are
distributed through the entire symbiotic region. The delta proteobacterial symbionts are in
close contact with the gamma proteobacterial symbionts suggesting a dependency on each
others metabolites (Rühland et al., In Prep.; Dubilier et al., In press).

The spirochete symbiont falls on a neighboring branch with free living marine spirochetes,
like *Spirochaeta isovaleric* (Harwood & Canale-Parola, 1983) and *S. litoralis* (Hespel &
Canale-Parola, 1973). TEM analyses of *O. algarvensis* showed no spirochete morphotype; instead FISH with spirochete specific probes provided evidence for their presence (Rühland et al., In Prep.).

**1.3 Function of *O. algarvensis* symbionts**

**1.3.1. Physiological diversity**

For long time was considered that symbiosis between bacteria and invertebrates is fueled by symbionts chemoautotrophy. Chemoautotrophic bacteria derive metabolically useful energy from the oxidation of inorganic compounds such as hydrogen, carbon monoxide, inorganic reduced sulfur and nitrogen compounds and from divalent ions (Lengeler et al., 1999). Thioautotrophs obtain energy by oxidation of reduced sulfur while their carbon is incorporated from carbon dioxide via Calvin cycle. The gamma proteobacterial symbionts from *O. algarvensis* were determined as being evolutionary related to sulfur-oxidizing bacteria like *A. vinosum*. TEM analysis showed the presence of sulfur globules in Gamma 1 symbiont and elemental sulfur is considered a good indicator for chemoautotrophic symbioses (Fisher, 1990). Furthermore combined molecular and immunocytochemical studies brought the final proof for the thioautotrophic nature of Gamma 1 symbiont. Immunocytochemical analyses with an antiserum directed against the form I of ribulose bisphosphate carboxilase consistently labeled this symbiont (Dubilier et al., 2001). 16S rRNA FISH with probes designed for Gamma 1 phylotype together with TEM and the immunocytochemical studies linked the phylogeny and morphology with the autotrophic function of Gamma 1 in the endosymbiotic community.

The physiologic nature of Gamma 2 is presently unknown while their evolutionary relationship with cold seeps bacteria suggest that they may also participate in chemosynthetic pathways. Currently immunofluorescence studies with antiserum against the form I and II of RubisCO are being used for a greater understanding of their metabolism (Rühland et al., In Prep.; Dubilier et al., In press). The lack of any intracellular sulfur inclusions is not a proof against thiotrophy (Fisher, 1990). It is well known that the oxidation of reduced sulfur compounds is not performed using a single unifying enzymatic pathway and that many
sulfur-oxidizing bacteria do not store elemental sulfur (Robertson and Kuenen, 1992; Kelly et al., 1997; Kappler and Dahl, 2001).

The delta proteobacterial symbionts, Delta 1 and Delta 2, are sulfate reducing bacteria based on their close phylogenetic relationship to free living sulfate reducers, the presence of the enzyme that catalyze the reduction of sulfite to sulfide, dissimilatory sulfite reductase, and the detection of sulfate reduction rates at comparable levels with those of free-living sulfate reducers (Dubilier et al., 2001).

Sulfate reducing bacteria (SRB) respire sulfate to sulfide and oxidize their substrates either partially to acetate or completely to CO$_2$ (Widdel and Hansen, 1992a). Some sulfate reducers (e.g. Desulfovibrio) are non acetate oxidizers which utilize lactate, ethanol, short fatty acids (succinate, fumarate, malate) and pyruvate, as electron donors. The resulting metabolic end product is acetate. Other sulfate reducers are acetate oxidizers (e.g. Desulfobacter, Desulfosarcina) and they are able to use acetate and other substrates like lactate, higher fatty acids and phenyl-substituted organic acids. For these acetate oxidizers the end product is CO$_2$. Some acetate and non-acetate oxidizers are able to use molecular hydrogen as an electron donor. (Widdel and Hansen, 1992a)

Both $O. algarvensis$ delta symbionts were classified by molecular analyses as part of the Desulfovibrio/Desulfooccus/Desulfonema group. SRB are metabolically diverse in particular within the Desulfovibrio group, where both chemoorganotrophy and chemosynthesis occurs and dissolved organic carbon and hydrogen are possible sources of reducing power. Hypothesizing that endosymbiotic sulfate reducers take up fermentation products (succinate, acetate, propionate) from a host that would otherwise excrete them (Dubilier, 2004) and knowing that the free living relatives of the symbionts produce CO$_2$ the following questions arise. Is CO$_2$ used by the chemoautotrophic symbiont so that there is no need anymore for an external CO$_2$ source? Can the delta proteobacterial symbionts use H$_2$, given that many SRB can utilize molecular hydrogen as an electron donor (facultative chemolithotrophs)? During this mode of energy conservation cell material may be synthesized from acetate and CO$_2$ (chemolithoheterotrophs) or only from CO$_2$ (autotrophs). Perhaps one phylotype is using the fermentation products of the host and the other phylotype uses molecular hydrogen. As the Desulfovibrio group is so diverse metabolically it is
possible for the two different delta endosymbionts to perform different functions in the community (Dubilier et al., In press).

The spirochete symbiont is closely related to free living spirochetes like *Spirochaeta litoralis* and *S. isovalerica* isolated from marine sediments and salt marshes. Spirochetes are extremely diverse physiologically ranging from aerobes to facultative and obligate anaerobes. Within these groupings there is a large diversity of nutritional requirements and energy yielding mechanisms. Some spirochetes derive their energy exclusively from the fermentation of plant polymers; others ferment a wide variety of sugars and amino acids, while some dissipilate only certain long-chain fatty acids. *Spirochaeta litoralis* and *S. isovalerica* are free living, strictly anaerobic spirochetes which are able to ferment glucose mainly to acetate, ethanol, CO2 and H2 (Hespell and Canale-Parola, 1973; Harwood and Canale-Parola, 1983, 1984). A spirochete symbiont utilizing glucose would be disadvantageous for the worm. Recent research has showed that the metabolic possibilities of spirochete symbionts are even broader than fermentation. Termite symbiotic spirochetes were discovered as being chemoautotrophs, with the ability of H2-CO2 acetogenesis (Leadbetter et al., 1999) and dinitrogen fixation (Lilburn et al., 2001). This type of metabolism would be beneficial for the oligochaete host, providing an important carbon and energy source beside the source of nitrogen (Dubilier et al., In press).

### 1.3.2. Mutualistic relationships: syntrophic sulfur cycle

The discovery of multiple symbionts was controversial and contrasting theories arouse about whether competition or cooperation occurs between the symbionts. Recent advances in the molecular characterization of uncultivable organisms, which many symbionts are, revealed a large diversity of mutualistic relationships. Gutless oligochaetes are a good example for the successful evolution of mutualistic associations with multiple symbionts. The sulfide-oxidizing and sulfate-reducing symbionts do not appear to compete for host derived resources, but cooperate instead in the use of resources from each other and the environment (Dubilier, 2004).

The coexistence of sulfate reducing and sulfide oxidizing bacteria as endosymbionts in *O. algarvensis* indicates that these are engaged in a syntrophic sulfur cycle in which oxidized
and reduced sulfur compounds are recycled between the symbionts (Fig. 1) (Dubilier et al., 2001). The SRB are producing reduced sulfur compounds using as reducing power either organic carbon or hydrogen. The reduced sulfur compounds are used by the sulfide oxidizer as electron donors for autotrophic CO2 fixation via Calvin cycle. This syntrophic sulfur cycle resembles a “perpetuum mobile”, and it is necessary for both host and symbionts to take up form the environment energy sources like dissolved organic compounds (Dubilier et al., 2001; Dubilier, 2004; Dubilier et al., In press). The role of the spirochete in this syntrophy is unclear.

1.3.3. Host-symbiont relations: possible interactions

The existence of a sulfide oxidizer, the Gamma 1 symbiont, that donates chemoautotrophically synthesized organic carbon to its host, is clearly benefic for the host, nourishing the worm. While anaerobic metabolites like succinate, acetate, and propionate produced by the worm when oxygen is limited might be taken up by the endosymbiotic sulfate reducers. The absence of excretory and digestive systems for these oligochaetes made them dependent on their symbionts The cycling of metabolic products between symbionts could increase their energy yields, which in turn would benefit the host. SRB supply also the sulfide oxidizing bacteria with internal produced sulfide, creating independence of environmental sulfide sources. This increases the ability of the worm to colonize new habitats with low sulfide concentration. Another advantage of having genetically diverse symbionts would be the host versatility in adapting different environments (Dubilier, 2004).
1.4. Metagenomics: microbial community genome analyses

The pure culture approach to the study of the microbial world seriously constrained the view of microbial diversity because most microbes defy cultivation by standard methods (Pace, 1997). The study of oligochaete symbiotic associations remains limited by the fact that bacterial symbionts are uncultivable despite numerous attempts. Nowadays, new tools are available for investigating these associations without separating the partners; one of them, the metagenomic approach, is defined as culture independent genomic analyses of microbial communities (Schloss and Handelsman, 2003; Riesenfeld et al., 2004).

Two approaches functional driven analyses and sequence driven analyses have emerged to extract information from metagenomic libraries. Functional analyses are based on identification of clones that express a desired trait, followed by characterization of the active clones by sequence and biochemical analysis. The limitations of this approach are that it requires expression of the function of interest in the host cells and clustering of all of the genes required for the function. It also depends on the availability of an assay for the function of interest that can be performed efficiently on vast clone libraries, knowing that the frequency of active clones is quite low (Schloss and Handelsman, 2003; Riesenfeld et al., 2004). Sequence based approach relies on the use of conserved DNA sequences to design hybridization probes or primers in order to screen metagenomic libraries for clones containing genes of interest. This has proved effective for identification of clones carrying phylogenetic anchors and genes encoding enzymes with highly conserved domains. (Piel, 2002) Significant discoveries resulted have also resulted from random sequencing of metagenomic clones (Beja, 2004). There are contradictory opinions on the utility of random sequencing of metagenomic clones, the method being considered too indirect to yield biological understanding, but it is stressed that there is so little known about some divisions of bacteria that any genomic sequence is helpful in guiding the design of new experiments to reveal their biology (Schloss and Handelsman, 2003; Riesenfeld et al., 2004). Taking in consideration advantages and disadvantages of both approaches can be emphasized that function driven approach has the potential to identify genes that would not be recognizable based on their sequences, but sequence based screening can identify sequences that are not
possible to express in the host species carrying the library. A combination of sequence-based methods and functional screening is critical in order to define the full diversity of gene function in the libraries. (Riesenfeld et al., 2004)

Metagenomic technology has been successful at all scales—it has been used to study single genes e.g. for cellulases (Healy et al., 1995), pathways e.g. antibiotic synthesis (Rondon et al., 2000), organisms e.g. Archaea (Stein et al., 1996) and communities e.g. acid mine drainage biofilm (Tyson et al., 2004). The ultimate goal of any genomic analyses is to use the information for a better understanding of the biology of the organism, in the present study—the biology of *O algarvensis* endosymbionts.

The objective of the study of *O. algarvensis* symbionts using the metagenomic approach is to obtain information on the interactions between the endosymbionts with their host and their environment. Therefore important issues are under question. How has the bacterial genome been modified in the context of the demands of symbioses? What pathways are used for the syntrophic interactions between the bacterial symbionts? Do the bacteria compete with each other for substrates from the host or the environment? Which metabolic pathways are used by the different symbionts to sequester inorganic and organic carbon, sulfur, and nitrogen from the environment? (Dubilier et al., Proposal for the Community Sequencing Programmm of the Joint Genome Institute, USA).

The present survey is based on sequence-driven metagenomic analysis of a BAC (Bacterial Artificial Chromosome) library, using vector primed insert sequencing and continuing with primer walking. The obtained contiguous sequences were subjected to similarity searches using BLASTx which aligns their inferred translation products against protein sequences from protein databases. This project is focused on providing row data in order to become able to discern interaction patterns in between symbiotic partners and also with their oligochaete host. The goal of the project is to assemble as many fragments as possible to gain a better understanding of the endosymbiots genetic potential and metabolic diversity.
2. Chapter 2: Materials and Methods

2.1. Symbionts DNA isolation and BAC library construction

DNA isolation and BAC library construction was performed at Texas A&M University, Texas, USA, by Hong-Bin Zhang, Chengcang Wu, and Zhanyou Xu in 2004. High molecular weight DNA was extracted from a pooled sample of approximately 500 Olavius algarvensis individuals. The extracted DNA was used for the BAC library construction. Enzymatic restriction was carried out and restriction fragments were ligated into the cloning vector pECEBAC1. Competent E. coli cells, strain ElectroMAX DH10B were transformed by electroporation.

2.2. Plasmid preparation

For BAC DNA extraction the precultured, transformed Escherichia coli cells were grown 18 to 20 hours at 37°C in 5 ml Luria Bertani medium, containing 17 mg/l chloramphenicol (see below). Addition of chloramphenicol assures that only transformed E. coli cells are growing. The following BAC extraction protocol represents the standard alkaline lysis method (Sambrook et al., 1989) and contains the next important steps: removal of Luria Broth, lysis of the cells, DNA precipitation and purification.

Transformed E. coli cells from the overnight culture were precipitated by centrifuging on an Eppendorf 5810 R centrifuge (Eppendorf) at 3000 rpm for 10 min. The supernatant was removed to eliminate the Luria Broth. The remaining pellet was transferred into a 2 ml microcentrifuge tube and resuspended by adding 200 µl alkaline lysis solution I (glucose 50 mM / EDTA 10 mM / TrisHCl 25 mM). The suspension was mixed by inverting the tube and incubated on ice for 5 minutes. The bacterial cells were lysed in 400 µl solution II (SDS 1 % / NaOH 0.2 N) and the mixture was gently shaken prior to ice incubation for another 5 minutes. Cell debris and chromosomal DNA were precipitated by adding 300 µl solution III (potassium acetate 3 M / acetic acid 1.15 M). The mixture was gently shaken, incubated on ice for 15 min and centrifuged at 5600 rpm, 15 min, on an Eppendorf 5415 R microcentrifuge (Eppendorf). The supernatant was transferred to a 1.5 ml microcentrifuge tube. The next
steps are important for precipitating the BAC DNA and removing RNA. Using 450 µl isopropanol, plasmidial DNA was precipitated by centrifuging for 5 minutes at 11400 rpm on an Eppendorf 5415 R microcentrifuge (Eppendorf). The supernatant was discarded from the tube and the pellet was cleaned of salts and other small molecules by adding 450 µl of ethanol 70 % and centrifuging at 11400 rpm for 2 minutes on an Eppendorf 5415 R microcentrifuge (Eppendorf). The ethanol was removed, the nucleic acid pellet was air dried and resuspended in 40 µl TE buffer (TrisCl 100 mM / EDTA 10 mM) prior to incubation at 65°C for to dissolve the BAC DNA and inhibit DNAses. The presence of the BAC, after extraction, was verified by agarose gel electrophoresis using 1 % [w/v] SeaKem LE Agarose (BioZym) gels and run in TAE buffer 1x (Tris acetate 45mM / EDTA 1mM) at 75 mV, 20 minutes. Lambda digested with Hind III was used as ladder. Gel pockets were loaded with 5 µl of BAC extract and 3µl loading buffer (0.4 % [w/v] bromphenol blue solution) Agarose gels were stained with ethidium bromide (5 µg/ml), and photographed.

2.2.1. Liquid Luria Bertani medium

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptone from casein tryptic digest</td>
<td>10 g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>5 g</td>
</tr>
<tr>
<td>Sterile water</td>
<td>to1000 ml</td>
</tr>
</tbody>
</table>

The pH was adjusted to 7.00 with NaOH 5 N and was autoclaved 25 min; chloramphenicol was added to 17 mg/l final concentration;

2.3. BAC insert size determination via pulsed field gel electrophoresis

For BAC insert size determination pulse field gel electrophoresis in contour-clamped homogenous electric field (CHEF) has been performed. This device comprises a hexagonal array of electrodes in a voltage divider circuit and produces homogenous fields (oriented at 120°) approximating those of infinitely long parallel electrodes (Sambrook et al., 1989).
To obtain fragments of the desired size (2 kb to 100 kb) the BAC DNA was digested with a rare cutter restriction enzyme, Not I, which has GC↓GGCCGC as recognition site. The size of the restriction fragments depends on the GC content of the BAC insert since a higher GC content raises the probability that Not I finds more restriction sites.

Analysis of inserted DNA was performed by digestion of each BAC with Not I at 37°C for 3 hours. The restriction reaction was performed with 40 mM spermidine, 4 µl Not I buffer 10x, 0.075 U/µl Not I (final concentration), 10 µl BAC DNA and adjusted to 40 µl with sterile water. Loading dye 10x was added to the restriction reaction in a proportion of 1 to 10 [v/v] and the restriction reaction was stopped by heating at 65° C for 10 min. Digests were analyzed with PFGE using 1 % SeaKem LE agarose gel in TBE buffer 0.5x (AppliChem). Gel pockets were loaded with 5 µl of restricted BAC and 1 µl loading buffer. PFGE was performed on a BioRad CHEF-DR III system using the following parameters: linear pulse time ramp from 5 to 15 sec, 10 h run time, 120° field angle and 6 V/cm field strength. The CHEF gels were stained in an ethidium bromide bath (5 µg/ml) for 30 min. The size of the restricted fragments was evaluated by comparison to high molecular weight markers in the form of agarose plugs (Lambda Ladder PFG Marker, Low Range PFG Marker (New England, BioLabs). For smaller fragments Lambda digested with Hind III (Invitrogen) was used.

2.4. BAC sequencing and primer walking

Primer walking is a sequencing method used to obtain contiguous sequence information (contigs). The initial sequences from each end of the BAC insert were obtained using standard vector primers, T7 and SP6 (Biomers) (Table 1). To sequence further into the insert, new primers were designed and synthesized (Biomers) (Table 2).

Sequencing on vectors that were earlier identified as containing cbb L gene, have been done during this study with specific cbb L primers (Table 1).

Sequencing reactions were performed in 15 µl final volume using Big Dye reaction mix (Applied Biosystems) or ET reaction mix (Amersham) with 0.67 µM primer (final concentration) and 5 µl BAC DNA template. The sequencing reaction was carried out in an Eppendorf Master Cycler (Eppendorf) starting with the denaturizing step at 96°C for 30 sec,
primer annealing at primer dependent temperatures (see Table 2) for 30 sec, and elongation for 4 min at 60°C, repeated for 98 cycles. The products of the sequencing reactions were purified on Sephadex G-50 Superfine (Amersham) columns according to the manufacturer’s manual and run on an Applied Biosystems 3130 xl Genetic Analyzer (Applied Biosystems).

<table>
<thead>
<tr>
<th>Vector primers</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td><strong>Sequence</strong></td>
<td><strong>Tm</strong></td>
</tr>
<tr>
<td>T7</td>
<td>5'-taa tac gac tea cta tag gg-3'</td>
<td>56°C</td>
</tr>
<tr>
<td>SP6</td>
<td>5'-tta ggt gac act ata gaa tac-3'</td>
<td>50°C</td>
</tr>
</tbody>
</table>

Table 1  List of primers used for end-sequencing (T7 and SP6) and Cbb L primers for sequencing the gene which codifies for RubisCO

2.5. Primer design

Primers were designed manually starting with aligning and comparing the end-sequences using the BioEdit program (www.mbio.ncsu.edu/BioEdit/bioedit.html). The priming site was chosen with 100 nucleotides upstream the first region of poor quality (areas beyond single peak resolution of a chromatogram). During primer design the following issues were considered: i) melting temperature (Tm) in the range of 52°C to 65°C; ii) absence of dimerization capability; iii) absence of significant hairpin formation; iv) low specific binding at the 3' end (lower GC content to avoid mispriming); v) high specific binding at the 5' end; vi) G+C to A+T ratio should be close to 1.

All primers were designed accordingly, having from 20 to 24 bases in length, a GC content of 40 to 60 % and a Tm range from 50° to 66°C (see Table 2).

2.6. Sequence analysis

Sequences were analyzed using BLASTx software which compares a nucleotide query sequence translated in all reading frames against a protein sequence database (http://www.ncbi.nlm.nih.gov/blast). The BLASTx software evaluates the quality of a pairwise sequence alignment using the substitution matrix BLOSUM 62 with the word size 3. The sequences were assembled into contiguous units with SEQUENCER® Contig Assembly.
Program (Gene Codes Corporation). These contiguous sequences were again compared against protein sequences from the non-redundant protein data base (http://www.ncbi.nlm.nih.gov/blast) using BLASTx. Known genes and putative functions were assigned for each individual sequence by inspection of the search output. Information of protein functions and their metabolic roles was searched using enzyme and protein data bases easily accessible from DBGET (http://www.genome.jp/dbget/) which is an integrated database retrieval system for a diverse range of molecular biology databases. In order to assess the metabolic functions most of the retrieved sequence information was searched against KEGG Pathway, release 33 (http://www.genome.jp/kegg/pathway.html). Diverse roles in cellular functions, pathogenesis or symbiotic interactions were assessed by seeking against protein databases like PIR release 79 (http://www.genome.jp/dbget-bin/www_bfind?pir).
Table 2 List of primers designed and used during this study; the determined dimensions of the BACs, in kilo bases (kb), are also listed below. Tm stands for the melting temperatures of the primers.

<table>
<thead>
<tr>
<th>No</th>
<th>BAC</th>
<th>Size (kb)</th>
<th>Vector</th>
<th>Primer</th>
<th>Name</th>
<th>Sequence</th>
<th>Tm</th>
<th>Name</th>
<th>Sequence</th>
<th>Tm</th>
<th>Name</th>
<th>Sequence</th>
<th>Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1H4</td>
<td>71kb</td>
<td>T7</td>
<td>Primer I</td>
<td>dsrF_1H4_199</td>
<td>5'-agt ggt tgt gaa acc gca cc-3'</td>
<td>62°C</td>
<td>dsrE-1H4_250</td>
<td>5'-cca aca tca ggc gta ctg cc-3'</td>
<td>62°C</td>
<td>dsrE_1H4_611(p3)</td>
<td>5'-tgt gcc gca tgt gaa ctg cg-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>2</td>
<td>1B4</td>
<td>90kb</td>
<td>T7</td>
<td>Primer II</td>
<td>Dh_B4_119</td>
<td>5'-agg tta agg cta tta cc-3'</td>
<td>60°C</td>
<td>FeSpr_1B4_475</td>
<td>5'-agg tta tgt tca tgt gct gc-3'</td>
<td>60°C</td>
<td>H2ase_1B4355(p3)</td>
<td>5'-tgt gcc aca tgt gcc gc-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>3</td>
<td>2D12</td>
<td>67kb</td>
<td>SP6</td>
<td>Primer III</td>
<td>napA_1267</td>
<td>5'-tgt gcc tgt cc gta cca gc-3'</td>
<td>62°C</td>
<td>napG_2D12_341</td>
<td>5'-ctt cca gac tgt cca cg-3'</td>
<td>62°C</td>
<td>napF_2D12_341(p3)</td>
<td>5'-tgt gcc aca tgt gcc gc-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>4</td>
<td>2F5</td>
<td>48kb</td>
<td>SP6</td>
<td>Primer IV</td>
<td>apr_F5_260</td>
<td>5'-aat ggg tgt gta aac aat gcc cc-3'</td>
<td>62°C</td>
<td>apr_2F5_254</td>
<td>5'-gtg gta aac ggc gta ctg cc-3'</td>
<td>62°C</td>
<td>Apr³_2F5_472(p3)</td>
<td>5'-gtg gta aac ggc gta ctg cc-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>5</td>
<td>2D4</td>
<td>45kb</td>
<td>SP6</td>
<td>Primer V</td>
<td>ITP_B_272</td>
<td>5'-ctt cca cat aat ggt caa cc-3'</td>
<td>60°C</td>
<td>ITP_B_272</td>
<td>5'-ctt cca cat aat ggt caa cc-3'</td>
<td>60°C</td>
<td>FeoB_2D4_363(p3)</td>
<td>5'-cca cat aat ggt caa cc-3'</td>
<td>60°C</td>
</tr>
<tr>
<td>6</td>
<td>3B6</td>
<td>39kb</td>
<td>SP6</td>
<td>Primer VI</td>
<td>metr_B6_259</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>60°C</td>
<td>metr_B6_259</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>60°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>3E11</td>
<td>67kb</td>
<td>SP6</td>
<td>-</td>
<td>soxA_E11,6_361</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>soxA_E11,6_361</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>4H3</td>
<td>63kb</td>
<td>SP6</td>
<td>Primer VIII</td>
<td>SR_H3_329</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>60°C</td>
<td>SR_H3_329</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>60°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>5C6</td>
<td>13kb</td>
<td>SP6</td>
<td>-</td>
<td>COdh_C6_72</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>COdh_C6_72</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>8G3</td>
<td>29kb</td>
<td>T7</td>
<td>-</td>
<td>OctdG_G3_65</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>OctdG_G3_65</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>8E6</td>
<td>41kb</td>
<td>T7</td>
<td>-</td>
<td>soxA_E11,6_361</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>soxA_E11,6_361</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>8D1</td>
<td>250kb</td>
<td>T7</td>
<td>-</td>
<td>ureF_D1_250</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>ureF_D1_250</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>9H6</td>
<td>64kb</td>
<td>SP6</td>
<td>-</td>
<td>Mep_H6_151</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>Mep_H6_151</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>9A11</td>
<td>35kb</td>
<td>T7</td>
<td>-</td>
<td>metr_A11_320</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>metr_A11_320</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>9D3</td>
<td>46kb</td>
<td>T7</td>
<td>-</td>
<td>ureC_9D3_128(p1)</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>ureC_9D3_128(p1)</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>9C10</td>
<td>38kb</td>
<td>T7</td>
<td>-</td>
<td>SR_C10_267</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>SR_C10_267</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>9B5</td>
<td>570kb</td>
<td>SP6</td>
<td>-</td>
<td>COdh_B5_278</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>COdh_B5_278</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>10H7</td>
<td>-</td>
<td>SP6</td>
<td>-</td>
<td>vapB_H7_357</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>vapB_H7_357</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>10H10</td>
<td>44kb</td>
<td>T7</td>
<td>-</td>
<td>cya_F5_H10_396</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>cya_F5_H10_396</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>10F5</td>
<td>46kb</td>
<td>T7</td>
<td>-</td>
<td>cya_F5_H10_396</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>cya_F5_H10_396</td>
<td>5'-tta gaa gcc gca tgt cc-3'</td>
<td>62°C</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
</tbody>
</table>
3. Chapter 3: Results

Vector primed end sequencing on the 500 clones from *O. algarvensis* BAC library revealed functional genes of potential importance for the metabolism of the symbionts. Twenty from these clones were previously chosen for primer walking based on the reason that they might bear genes that encode for proteins involved in carbon, sulfur, nitrogen metabolisms, and symbiotic/pathogenic related traits.

3.1. BAC insert size determination

Dimension of the chosen BACs was determined using pulsed field gel electrophoresis. The NotI restricted fragments of the BAC inserts were compared to markers which ranged from 125 bp (Lambda digest of Hind III) to 1000 kb (Lambda Ladder PFG Marker). For each clone the size of its insert was assessed by summing up the size of the restriction fragments. The average insert size was determined as being 50 kb with the smallest value 13 kb and the largest 90 kb (Table 2).

3.2. Sequence analysis and primer walking

The sequences of the functional genes discovered during the first screening of the library by Anna Blazejack were confirmed during this study and subjected to primer walking. Numerous contiguous sequences were obtained and their translated consensus sequences showed similarity to proteins involved in carbon, sulfur and nitrogen metabolism, transport processes and also other relevant pathogenic or symbiotic functions. Assigning a function based on sequence similarity is a matter of prediction; proofs for function can only be brought using functional tests.

Similarity search against protein databases of the obtained contiguous sequences was done using BLASTx and similarities with E-values less than 1E-10 were considered relevant (Annex 1). Similarities higher than 40 % for identical amino acid residues and 60 % for positive amino acids together with similarity scores less than 1E-10 were considered significant for maintaining the protein function and were taken into account.
The encoded protein information was classified into five categories: carbon metabolism, sulfur metabolism, nitrogen metabolism, transport proteins and pathogenic/symbiotic interactions.

### 3.2.1 Carbon metabolism

Autotrophy requires the ability to synthesize cell material from inorganic carbon. The key enzyme for autotrophy is RubisCO which catalyzes the fixation of CO$_2$. $cbbL$ is the gene that encodes for the large subunit of RubisCO. In this study were sequenced five $cbbL$ genes, from five BAC clones, which were previously identified by Anna Blazejack as presenting this gene.

In this project were sequenced, from the vector, other genes encoding for enzymes which might be involved in one carbon metabolism like CO dehydrogenase, formate dehydrogenase, trimethylamine methyltransferase, sarcosine oxidase and opine oxidase.

#### 3.2.1.1 RubisCO key enzyme of Calvin cycle

The gene encoding for the large subunit of RubisCO form I ($cbbL$) was identified by PCR screening with specific $cbbL$ primers on five clones in the BAC library of *O. algarvensis* (by Anna Blazejak).

In this study sequencing of the $cbbL$ gene on these five BAC clones was performed using as primer $cbbL$ 1b. The five sequences obtained were aligned with BioEdit and showed to be identical (see Annex 2). The consensus sequence was analyzed with BLASTx and the inferred translation was found similar to the large chain of RubisCO from the gammaproteobacterial sulfur oxidizer, *Allochromatium vinosum* (92 % identical and 97 % positive amino acids).

#### 3.2.1.2 CO dehydrogenases

Two BAC clones were identified as having genes putatively encoding for CO dehydrogenase. Using BioEdit no similarity between the two sequences has been found.
Primer walking on both clones ended with two contiguous sequences. One contig was only 545 nucleotides long (Fig. 2) and the BLASTx analysis showed that the inferred translation was similar to anaerobic CO dehydrogenase form I from the alphaproteobacterial marine phototroph *Rugeria* sp.. BLASTx showed a shift in the translation and both fragments were assigned to neighboring regions in CO dehydrogenase from *Rugeria* sp. (Annex 1) The percentages of identical and positive amino acids was calculated by making the average of identical and positive residues from the two fragments which had the same E-value (85 % identical and 89 % positive amino acids).

![Contig clone 5C6](image)

![Contig clone 9B5](image)

**Fig. 2** Graphic representation of two contigs which contain regions encoding for CO dehydrogenase; the arrows shows the correspondence between nucleotide sequence and amino acid sequences found similar to CO dehydrogenases from other organisms; The red stripe represent the cut in the translation.

The second contig was of 937 nucleotides (Fig. 2) and the first part of its translated consensus sequence showed similarity to the end of CO dehydrogenase from the methanogen *Methanosarcina mazei* (36 % identical and 55 % positive amino acids).
The second part of its deduced translation product was similar to protein 3 from the ComE operon of the deltaproteobacterial sulfate reducer *Desulfotaea psychrophila* (32 % identical and 48 % positive amino acids). The end of this translated contig was related to a sensory transduction histidine kinase from an uncultured archaean (52 % identical and 85 % positive amino acids).

### 3.2.1.3 Formate dehydrogenase

Primer walking on a clone, which was previously identified as containing a gene encoding for formate dehydrogenase, provided a contiguous sequence of 1485 nucleotides which was analyzed with BLASTx (Fig. 3). The first part of the translated consensus sequence was similar to the end of formate dehydrogenase α subunit from the alphaproteobacterial marine isolate *Silicibacter pomeroyi* (59 % identical and 70 % positive amino acids). The second part of the translated consensus sequence was similar to Fe-S containing formate dehydrogenase from *S. pomeroyi*. This region presented a cut in its conceptual translation and the fragments were assigned by BLASTx to neighboring parts form the iron sulfur subunit of formate dehydrogenase. The percentage (63 % identical and 74 % positive amino acids) represents the average of the identical and positive amino acids from the two fragments. The end of the translated consensus sequence had no similarity to any protein in the protein data bases.

![Fig. 3](Image) Graphic representation of one contig which contain regions encoding for formate dehydrogenase; the arrows shows the correspondence between nucleotide sequence and amino acid sequences found similar to formate dehydrogenases from other organisms; α, β, γ and Fe/S represent the α, β, γ and Fe/S subunits of FDH. The red stripe represents the cut in the translation. “//” represents a non scaled region.
3.2.1.4 Trimethylamine methyltransferase

Two BACs in the library were identified by end sequencing as containing trimethylamine methyltransferase genes. The two end sequences showed no similarity to each other. Therefore primer walking was initiated on both clones and two contiguous sequences were obtained. The BLASTx analysis of a 1512 nucleotides contig (Fig. 4) showed that the inferred translation product was similar to trimethylamine methyltransferase (MttB) from the alphaproteobacterial plant symbiont *Mesorhizobium loti* (58 % identical, 73 % positive amino acids). The end of the conceptual translation of this contig had no similarity to any protein stored in the protein data bases.

![Figure 4](image)

*Fig. 4* Graphic representation of two contigs which contain regions encoding for trimethylamine methyltransferase (MttB); the arrows shows the correspondence between nucleotide sequence and amino acid sequences found similar to MttB from other organisms; The red stripe represent the cut in the translation.

The second contig of 1516 nucleotides long (Fig. 4) have been identified as encoding in the first part a protein similar to MttB1 form *Methanosarcina acetivorans* (31 % identical and 51 % positive amino acids). A cut in the translation resulted in two fragments.
assigned to neighboring regions from MttB1. The average similarity percentage was calculated for the entire region that putatively encodes for a trimethylamine methyltransferase. The second part of the conceptual translation had no similarity to any proteic sequence from the available data bases, while the third region was similar to spermidine/putrescine ABC transporter from the anaerobic pathogen Fusobacterium nucleatum (35 % identical and 57 % positive amino acids). This region presented two shifts in its translation and the three fragments obtained were related to three near regions in the amino acid sequence of the spermidine/putrescine transporter F. nucleatum. The average similarity percentage was calculated and the E-value was the same for all three fragments (Annex 1).

### 3.2.1.5 Sarcosine oxidase

End sequencing on two clones showed that these BACs have sarcosine oxidase genes and BioEdit alignment of these two sequences proved that they are identical. Primer walking on one clone resulted in a contiguous sequence of 1592 nucleotides (Fig. x) whose inferred translation showed similarity to sarcosine oxidase, subunit A from S. pomeroyi (identical 45 %, positive 59 % amino acids). The conceptual translation presented a cut in its sequence and the two resulted fragments were similar to adjacent regions from SoxA of S. pomeroyi. The similarity of the entire inferred translation was considered by making an average of the percentages of identical and positive amino acids, and knowing that the E-value was identical for both fragments (Annex 1).

![Graphic representation of the contigs which encodes for subunit A of sarcosine oxidase; SoxA, B, D and G represents the subunits of sarcosine oxidase; the arrows shows the correspondence between nucleotide sequence and amino acid sequences; The red stripe represent the cut in the translation.](image)

**Fig. 5** Graphic representation of the contigs which encodes for subunit A of sarcosine oxidase; SoxA, B, D and G represents the subunits of sarcosine oxidase; the arrows shows the correspondence between nucleotide sequence and amino acid sequences; The red stripe represent the cut in the translation.
3.2.1.6 Opine oxidase

The contig of 1894 nucleotides (Fig. 6) was the result of primer walking on a clone that was previously identified as possible bearing one gene for opine oxidase. This contig was subjected to BLASTx analysis and the inferred translation of the first part resembled opine oxidase subunit B from the pSymA plasmid of *Sinorhizobium meliloti* (43 % identical and 54 % positive amino acids).

The next part from the translated contig was related to an uncharacterized NAD / FAD dependent dehydrogenase from the betaproteobacteria *Burkholderia fungorum* (48 % identical and 60 % positive amino acids). The final part of this translated contig was similar to opine oxidase subunit A from the alphaproteobacterial plant symbiont *Bradyrhizobium japonicum* (38 % identical and 55 % positive amino acids).

Fig. 6 Representation of the contig which encodes for subunit A and B of opine oxidase; OoxA, OoxB, represents the subunits of opine oxidase SoxD and SoxB represents subunits of sarcosine oxidase; the arrows shows the correspondence between nucleotide sequence and amino acid sequences; The red stripe represent the cut in the translation.

3.2.2 Sulfur metabolism

Four contiguous sequences obtained by primer walking were putatively encoding for enzymes involved in assimilatory and dissimilatory sulfur metabolism.
BLASTx analysis of one contig revealed the existence of an encoded beta subunit of sulfite reductase which is involved in assimilatory sulfur metabolism. Two enzymes which appear to be involved in both oxidative and reductive sulfur metabolism APS reductase and sirohaem sulfite reductase might be encoded by sequences from contigs obtained on two different clones from the BAC library. Another contig restrained a possible polysulfide reductase encoding region.

### 3.2.2.1 Sulfite reductase

A clone possibly bearing a *cysI* gene was identified and primer walking gave a contiguous sequence of 1923 nucleotides. The inferred translation product of the first region of this contig was similar to sulfite reductase from *Chromobacterium violaceum*, from the Betaproteobacteria class. BLASTx assigned the three fragments that were the result of two cuts in the translation to adjacent regions in the amino acid sequence of sulfite reductase. The average percentage of identical and positive amino acids was calculated (75 % identity and 89 % positive amino acids).

The next region from the translated consensus of this contig shows similarity to a hypothetical protein from the alphaproteobacterial plant symbiont *Bradyrhizobium japonicum* (32 % identical and 46 % positive amino acids). Next region resembled protein chain releasing factor A from the gammaproteobacterial marine isolate *Microbulbifer degradans* (71 % identical and 83 % positive). The rest of the inferred translation showed no similarity to any protein in the available data bases.

### 3.2.2.2 Adenosine-5'-phosphosulfate (APS) reductase

The α subunit of adenosine-5'-phosphosulfate reductase was previously identified as being encoded on a clone from *O. algarvensis* BAC library and using primer walking on a contig of 1956 nucleotides, have been established (Fig. 8). The BLASTx analysis showed that the first region from the conceptual translation was similar to α subunit of APS reductase from the gammaproteobacterial sulfur oxidizer *A. vinosum* (54 % identical and 60 % positive amino acids). The middle part was found not similar to known proteins.
while the end of the translated contig resembled the beginning of FKBP type peptidyl-prolyl cis-trans isomerase 2 from *T. denitrificans* (72 % identical and 84 % positive amino acids).

**Fig. 7** Representation of the contig which encodes for AprA which is one of the three subunits of APS reductase, the other two are AprM and AprB; Sat is sulfate adenyllyl transferase; the arrows shows the correspondence between nucleotide sequence and amino acid sequences;

### 3.2.2.3 Dissimilatory sulfite reductase

Four BACs were identified as containing genes which potentially encode for different proteins involved in dissimilatory sulfite reduction from *A. vinosum*. Primer walking was performed on one clone and a 1232 nucleotides contig was assembled (Fig. 8). BLASTx analysis of the consensus sequence showed that the translation product of the first region resembled DsrF from *A. vinosum* (71 % identical and 80 % positive amino acids). The next region form the translated consensus resembled DsrE also from *A. vinosum* (90 % identical and 95 % positive amino acids). The end of the translated contig showed similarity with the end of dissimilatory sirohaem sulfite reductase beta subunit from the same *A. vinosum* (58 % identical and 69 % positive amino acids).

Using BioEdit the consensus sequence of this contig was aligned against the *dsr* gene cluster (see Annex 2) of an entirely sequenced BAC (Anna Blazejack, unpublished data). For a better perception of this difference these two sequences were subjected to the bl2seq which gave a nucleotide similarity of 75 % for *dsrE* and *dsrF* genes and 88% for the end of *dsrB* gene.
3.2.2.4 Polysulfide reductase

Primer walking was initiated on a clone that was identified as containing a psrA gene that encodes for subunit A of polysulfide reductase. The length of the obtained contig was 1225 nucleotides (Fig. 9) and its inferred translation product resembled polysulfide reductase subunit A from the photosynthetic, anaerobic green-sulfur bacterium *Chlorobium tepidum* (40 % identical and 57 % positive amino acids).
3.2.3 Nitrogen metabolism

During this survey several regions were discovered that codify for proteins which could play a role in assimilatory, dissimilatory nitrogen metabolism and nitrogen storage for the symbionts of *O. algarvensis*. These included genes which encode for urease, periplasmatic nitrate reductase involved in the dissimilatory nitrate respiration and cyanophycin synthetase involved in nitrogen storage.

3.2.3.1 Urease

Two BAC clones were previously identified as bearing genes encoding for different subunits of urease. Therefore primer walking was initiated on both BAC clones and two contiguous sequences were assembled on which BLASTx analyses were performed. One contig of 803 nucleotides (Fig. 9) in the first part of its inferred translation was similar to the end of urease alpha subunit from the alphaproteobacterial plant pathogen *Agrobacterium tumefaciens* (81 % identical and 84 % positives). No similarity was found for the middle part of the contig while the end of its translation was vaguely similar to a region from anthranilate synthase component II of *Corynebacterium efficiens* (37 % identical and 48 % positive residues).

The second contig of 2019 nucleotides presents in the first part a region whose inferred translation resembled UreF from the gammaproteobacterial pathogen *Vibrio parahaemolyticus* (40 % identical and 56 % positive amino acids).

The next region from the translated contig was similar to UreG from cyanobacterium *Nostoc* sp. (72 % identical and 89 % positive amino acids). This region presents a shift in its conceptual translation and both fragments are aligned to neighboring segments in the amino acid sequence of UreG. The average content of identical and positive amino acids was calculated. The end of the translated consensus sequence was similar to a transposase from the betaproteobacterium *Polaromonas* sp. (37 % identical and 56 % positive amino acids)
3.2.3.2 Nitrate reductase

A 1672 nucleotides contig was obtained by primer walking on a clone that was previously determined as containing a \textit{napA} gene that encodes for periplasmatic nitrate reductase. This contig was subjected to BLASTx analysis and the first part of its inferred translation was similar to the end of periplasmatic nitrate reductase subunit A from the epsilonproteobacterium \textit{Wolinella succinogenes}. For this region a shift in the translation was observed, resulted fragments were assigned to adjacent regions from NapA and the percentage of identical and similar aminoacids was calculated by making an average of the identical and positive residues in both fragments (54 % identical and 67 % positive amino acids). Inferred translation of the second part of the contig was similar to Nap G from the gammaproteobacterial human pathogen \textit{Salmonella enterica} (50 % identical and 67 % positive amino acids). The end of this translated contig was slightly similar to the
beginning of NapH from \textit{W. succinogenes}. BLASTx output revealed, for this part of the inferred translation, a translation cut that resulted in two fragments similar to near regions from NapH. For these two fragments an average percentage of identical and positive amino acids was calculated and assigned to the entire region (43 % identical and 56 % positive amino acids).

### 3.2.3.3 Cyanophycin synthetase

Two BACs were determined as having genes which putatively codify for cyanophycin synthetase. BioEdit alignment showed that these two sequences were identical so that only one clone was sequenced. The result of the primer walking was a contig of 1587 nucleotides which was analyzed using BLASTx. Three cleavages occur in the translation and the fragments were all assigned to neighboring regions in the amino acid sequence of cyanophycin synthetase from the gammaproteobacterial intracellular pathogen \textit{Francisella tularensis}. The average of identical and similar aminoacids for all these three fragments was calculated (74 % identical and 88 % positive aminoacids).

### 3.2.4 Transport proteins

Transport systems allow the uptake of essential nutrients and ions, excretion of end products, metabolites and deleterious substances but also facilitate the communication between cells and the environment. These facts make the ‘transporter protein’ category, appropriate for studying the interactions of the symbiotic community from \textit{O. algarvensis}.

During this study three BAC clones were analyzed and contiguous sequences were assembled. The translation products were related to proteins like ferrous iron transport protein, natrium driven multidrug efflux pump and spermidine/putrescine transporter. The last one was analyzed previously as part of a contig described in the carbon metabolism section.
3.2.4.1 Ferrous iron transport protein (FeoB)

Primer walking was performed on a BAC which previously was identified that might encode FeoB. All sequences obtained with the designed primers were assembled into a contig of 1345 nucleotides. Similarity searches were performed using BLASTx and the inferred translation product of this contig was similar to ferrous iron transport protein B from the pathogen Porphyromonas gingivalis member of the Bacteroides. BLASTx showed that the translated contig had three regions of similarity to neighboring segments from FeoB. The E-value was identical for all three fragments whether the ratio of identical and positive aminoacid was different. For addressing the entire region assigned to FeoB an average of the percentages was calculated (47 % identical and 65 % positive amino acids).

3.2.4.2 Na+ driven multidrug efflux pump (NorM)

A BAC clone previously identified as possible bearing a gene encoding for a multidrug efflux pump was subjected to primer walking and the assembled contig of 831 nucleotides was analyzed using BLASTx. The translated consensus was related to a natrium driven multidrug efflux pump, NorM from the alphaproteobacterium Rhodopseudomonas palustris (35 % identical and 52 % positive amino acids). BLASTx showed a cut in the conceptual translation and the fragments were similar to near segments from the proteic sequence of NorM. For a general idea of the similarity to NorM of the translated contig an average value of the identical and positive amino acids from both fragments together.

3.2.5 Pathogenic/symbiotic associated traits

Traits that are generally associated with pathogenic or symbiotic lifestyle were found encoded in the genome of O. algarvensis symbionts: urease, ferrous iron transporter, and virulence associated proteins. Urease contigs that were analyzed are described in the ‘nitrogen metabolism’ section whether the ferrous iron transporter is explained in the
‘transporter proteins’ section. In this section only the virulence associated proteins are detailed.

### 3.2.5.1 Virulence associated proteins (VapB and VapC)

A contig of 1482 nucleotides was obtained on a clone which has been determined as possibly containing a gene for a protein associated with virulence. BLASTx analyses of this contig showed that the first part of the translated consensus sequence was similar to virulence associated protein B form *Leptospira interogans*. BLASTx output showed a shift in the translation and the resulted fragments were assigned to neighboring regions in virulence associated protein B. The E-value is identical for both fragments and the average of their identical and positive amino acid content was calculated (52 % identical and 77 % positive). The next part from the translated contig was found similar to virulence associated protein C from the deltaproteobacterium *Geobacter sulfurreducens* and also in this region a modification was detected and two slightly separated fragments were similar to neighboring regions that cover the entire VapC, and the percentages of identical and positive amino acids assigned to the entire region was calculated taking in consideration the values for both fragments (47 % identical and 68 % positive amino acids). The next region in the conceptual translation showed no similarity to any known protein.
4. Chapter IV Discussions

The symbiotic community of *O. algarvensis* consists of five phylotypes: two gammaproteobacterial, two deltaproteobacterial and one spirochete (Rühland et al., In Prep.). To address the genomic encoded function of these uncultured symbionts the sequence driven metagenomic approach was used. End-sequencing was previously performed on the 500 clones of the BAC library and sequences that have been considered relevant for the symbiotic community were confirmed during this study, prior to primer walking for proceeding further into the inserts. The outcome was represented by longer contiguous sequences that were analyzed with BLASTx.

In the present survey BLASTx assignments were performed based on the similarity of the inferred translation products to aminoacid sequences of proteins stored in the GenBank protein database. Assigning a function based on sequence similarity is a matter of prediction; proofs for function can only be brought using functional tests, e.g. expression tests (Riesenfeld et al., 2004). BLASTx gave in some cases results in different frames for fragments which were assigned to neighboring regions in the same protein and it might occur due to sequencing errors, knowing that deletion or misreading of only one nucleotide from the chromatogram can induce a modification in the inferred translation.

The existence of five phylotypes in the endosymbiotic community of *O. algarvensis* makes difficult the link between the functional information encoded in the community genome to any of the symbionts without having a 16 S rRNA phylogenetic anchor in the same clone.

The similarities are given taking in consideration both E-values (see Annex 1) and percentages of identical and positive aminoacids. The positive amino acids were considered since it is known that there are groups of aminoacids with comparable biochemical traits and these amino acids are considered as producing a positive change, maintaining the folding of the protein hereby keeping its function. In fact the function of a protein can be preserved also if only functional domains are conserved, which would allow a lot of sequence variation in unconserved parts of the protein.

Classification of the potentially encoded protein information was done into five categories: carbon metabolism, sulfur metabolism, nitrogen metabolism, transport
proteins and pathogenic/symbiotic relevant traits. This classification is maybe helpful for this study since we did not deal with such a high amount of sequence information like in a complete genome sequencing project were more categories are considered (e.g. Riley. 1993, Kaneko et al. 2000).

4.1. Carbon metabolism

The ability of O. algarvensis symbionts to grow with CO\textsubscript{2} was demonstrated earlier (Dubilier et al., 2001) by showing the occurrence of ribulose bisphosphate carboxylase, the key enzyme of the CO\textsubscript{2} fixation pathway. Immunocytochemical analyses showed the presence of this enzyme in the large morphotype, now considered as being the Gamma 1 phylotype (Rühland et al., In Prep.). Previously five genes encoding for RubisCO form I were determined by PCR screening and in the present study they were sequenced and showed sequence equivalence (Annex 2) facilitating the assumption that they might belong to the same symbiont, namely the Gamma 1. Autotrophy of the symbionts via Calvin cycle is nourishing the worm with reduced carbon substrates.

Using the sequence driven metagenomic approach were discovered encoded proteins which might be involved in one carbon metabolism: CO dehydrogenase, formate dehydrogenase, trimethylamine methyltransferase, opine oxidase and sarcosine oxidase. It seem that O. algarvensis symbiotic community might use one carbon compounds like CO, formic acid but also compounds that are known as methyl-group-donors like trimethylamines, sarcosine (N-methyl-alanine) or opines. The reductive-acetyl-CoA pathway for autotrophic CO\textsubscript{2} fixation can be used also for assimilation of various one carbon compounds (Lengeler et al., 1999). Methyloptrophs use CO, formate, formaldehyde, methanol, methylamine, methylmercaptane, or methane as sole sources of carbon. Some methyloptrophs are able to use one-carbon units linked via a hetero-atom (N, S, O) to the rest of an organic molecule(Lengeler et al., 1999). Methyloptrophs are known as symbionts for some marine invertebrates (Fisher, 1990) This kind of metabolism with synthesis of macromolecules starting from one carbon compounds, is certainly beneficial for a host which has no ability to use them but is incorporating the symbiotically synthesized macromolecules. The finding of encoded CO dehydrogenase and formate
dehydrogenase show that substrates like CO and formate might also support symbiotic growth in *O. algarvensis*.

The host also can nourish its symbionts with products of its metabolism like trimethylamine, opines, sarcosine or urea. A proof for such way of cycling carbon compounds is the finding of potentially encoded trimethylamine methyltransferase, opine oxidase and sarcosine oxidase.

### 4.1.1. CO dehydrogenase

CO dehydrogenases are present in physiologically and phylogenetically diverse microbes where the enzyme functions to either oxidize CO, synthesize acetyl-CoA, or cleave acetyl-CoA. Aerobic microbes utilize Mo-Fe-flavin CO dehydrogenases to oxidize CO in respiratory pathways. Phototrophic anaerobes converts CO and water to CO₂ and H₂, process catalyzed by Ni-(Fe-S) CO dehydrogenase. Acetate-producing anaerobes employ a Ni-(Fe-S) CO dehydrogenase to synthesize acetyl-CoA from a methyl group, CO and CoA. A similar enzyme is responsible for the cleavage of acetyl-CoA by methanogens that obtain energy by fermenting acetate to CH₄ and CO₂. Acetotrophic sulfate reducers also utilize CO dehydrogenase to cleave acetyl-CoA yielding methyl and carbonyl groups and they obtain energy for growth via a respiratory pathway in which the methyl and carbonyl groups are oxidized to CO₂ and sulfate is reduced to sulfide (Ferry, 1995). Autotrophic sulfate reducers fix CO₂ into their cell carbon through the non cyclic reductive acetyl-CoA pathway which has as key enzyme the same CO dehydrogenase/acetyl-CoA synthase. The reductive acetyl-CoA pathway for autotrophic CO₂ fixation can be used also for assimilation of various one carbon compounds (Lengeler et al., 1999)

The conceptual translation products of two contigs from the studied BAC library were found related to CO dehydrogenases from the methanogen *Methanosarcina mazei* and from the anaerobic phototroph *Rugeria* sp.

Knowing in which kind of organisms the CO dehydrogenases could occur and relating with the members of *O. algarvensis* symbiotic community, could be speculated that any of the sulfate-reducing delta symbionts might utilize the CO dehydrogenase/acetyl-CoA
synthase pathway, either just for gaining energy by complete oxidation to CO\textsubscript{2} being a complete oxidizing sulfate-reducer but might be that they perform CO\textsubscript{2} fixation being autotrophs, by using the reductive CO dehydrogenase pathway.

### 4.1.2. Formate dehydrogenase

Formate dehydrogenase oxidizes formate to carbon dioxide with the release of a proton and two electrons. In some anaerobic bacteria formate is a fermentation product, and formate dehydrogenase is a component of anaerobic formate hydrogen lyase complex. The ability to grow with formate has been observed in most genera of sulfate reducing eubacteria. Furthermore, spirilloid sulfur reducers and Archaeoglobus fulgidus are known to grow with formate. Formate dehydrogenase has been found in Desulfovibrio and in acetate oxidizing sulfate reducers, except Desulfobacter. Formate dehydrogenase in acetate oxidizers is part of the carbon monoxide pathway for acetyl CoA oxidation. In autotrophic sulfate reducers formate dehydrogenase is the first carbon fixing enzyme of the reductive acetyl-CoA pathway. Formate dehydrogenase in the sulfur reducer, Wolinella succinogenes, is membrane bound and involved in the oxidation of formate to CO\textsubscript{2} resulting in electron transport to sulfur reductase which is in close contact to formate dehydrogenase. (Widdel and Hansen, 1992a; Lengeler et al., 1999).

In this study the highest similarity of the potential encoded formate dehydrogenase is to the alpha and iron-sulfur subunits of formate dehydrogenase form the marine isolate Silicibacter pomeroyi. The genes for these two subunits in Silicibacter are placed in the following context: \textit{fdhβ}, \textit{fdhα}, followed by a large genomic sequence and than \textit{fdhγ}, \textit{fdh} for Fe-S subunit and \textit{fdhα} are encountered (NCBI complete genome for S. pomeroyi, CP000031). S pomeroyi seems to rely upon a lithoheterotrophic strategy using inorganic compounds (CO and sulfide) to supplement heterotrophy (Moran et al., 2004).

Taking in consideration the types of symbionts that \textit{O. algarvensis} bears could be therefore suggested that formate dehydrogenase might belong to a sulfate reducing member from the endosymbiotic community, given that in acetate oxidizers and autotrophic SRB this enzyme occurs. Another possibility is that the spirochete ferments
carbohydrates to formate with formate dehydrogenase as part of the formate hydrogen lyase complex.

4.1.3. Trimethylamine methyltransferase

Trimethylamine is a degradation product of trimethylammonium compounds like choline, carnitine, or lecithine which are present in all animal tissues. Trimethylamine in eukaryotes is oxidized to trimethylamine oxide which is an osmoregulators in muscle tissues of marine organisms protecting proteins by counteracting destabilizing forces like urea, ammonia, hydrostatic pressure, temperature stress, and salt-stress. (Lengeler et al., 1999; Seibel and Walsh, 2001).

Trimethylamine methyltransferase is catalyzing the transfer of methyl groups either to a corrinoid protein in methanogens or to tetrahydrofolate in methylotrophs (Lengeler et al., 1999).

Similarities of the conceptual translations of two contigs to trimethylamine methyltransferase from *Silicibacter pomeroyi*, and *Methanosarcina mazei* represent an indicator that *O. algarvensis* symbionts could utilize trimethylamines.

Trimethylamine represents a precursor for worm’s osmolytes e.g. trimethylamine oxide, but in the same time trimethylamines could represent a new carbon source for its symbionts. Trimethylamines are known substrates for some sulfate reducers which have the ability to degrade N-methylated compounds (Widdel and Hansen, 1992a). In *Olavius algarvensis* one possibility would be that the sulfate reducing symbionts take use of the break down of the worm osmoregulators, being able to use trimethylamines as methyl group donors for the reductive acetyl-CoA pathway. Symbiotic utilization of the precursors of worm’s osmoregulators might not be valuable from the host point of view.

4.1.4. Sarcosine oxidase

Sarcosine or N-methylglycine is a eukaryotic metabolite found in eukaryotic tissues which can be used as sole carbon and energy source by microorganisms. Sarcosine oxidase is a flavoprotein that catalyzes the oxidative demethylation of sarcosine yielding
glycine, H₂O₂, 5,10-methylenetetrahydrofolate in a reaction requiring tetrahydrofolate and oxygen. In the absence of tetrahydrofolate, the methyl group from sarcosine is released as formaldehyde. Bacterial sarcosine oxidases have been isolated from different organisms and fall into two classes: heterotetrameric (α β γ and δ subunits) and monomeric, the last ones being similar to β subunit (SoxB) from the heterotetrameric enzymes. (Chlumsky et al., 1995)

One contig was found during this study as potentially encoding for a protein similar to sarcosine oxidase subunit α from *S. pomeroyi*. In *S. pomeroyi* the gene cluster that encodes for sarcosine oxidase shows a similar organization like the described operon which encodes for the heterotetrameric enzyme in *Corynebacterium sp* (Chlumsky et al., 1995) which is soxBDAG with soxA encoding for subunit α, soxB for subunit β, soxC for γ whether soxD encodes for the δ subunit (Moran et al., 2004).

Sarcosine which results from the host metabolism could be used by the endosymbionts most probably by the sulfate reducers which are known for their ability to use methyl groups bounded to a heteroatom like nitrogen, in the case of sarcosine.

4.1.5. **Opine oxidase**

Opines are the products of the NAD(P)H-dependent reductive condensation between α-keto acid and the α- or ω-NH₂ group of an amino acid. This kind of compounds have been isolated from eukaryotic cells, plant tumors, bacteria and marine invertebrates (Britton et al., 1998). In marine invertebrates, muscular metabolism has its own characteristics with glycolytic end products different than lactate. In invertebrates opines are formed instead of lactate and they represent condensation products of pyruvate with amino acids (e.g. alanine for alanopine, glycine for strombine, arginine for octopine). (Portner, 2002). Opine oxidase catalyze the degradation of opines and this enzyme is met in syntrophic or pathogenic bacteria that can use this kind of metabolites from their hosts, like the plant pathogen *A. tumefaciens* or the leguminous plant symbiont, *S. meliloti* which is presumed that can parasitize plant tumors produced by the first one (Barnett et al., 2001).
The translated contiguous sequences of one contig was considered similar in the first part to subunit B from *S. meliloti* and the neighboring region seemed similar to subunit A of opine oxidase from *B. japonicum*. Might be possible that *O. algarvensis* has an advantage if it has endosymbionts which might consume the acidifying end products of its anaerobic glycolysis.

4.2. **Sulfur metabolism**

The existence of both sulfur oxidizing and sulfate reducing bacteria in *O. algarvensis* makes possible the cycling of reduced and oxidized sulfur species between symbionts. The idea of this kind of syntrophic sulfur cycle was supported initially by phylogenetic analysis, showing the relations of *O. algarvensis* symbionts to other sulfur oxidizing and sulfate reducing bacteria (Dubilier et al., 2001). Evidences for the reductive side of this syntrophic sulfur cycle beside 16S rRNA phylogeny were the determination of sulfate reduction rates at levels similar to free living sulfate reducers and the PCR amplification of dissimilatory sulfite reductase (Dubilier et al., 2001). Evidences for the oxidative side of this syntrophic sulfur cycle beside 16S rRNA phylogenetic analyses was the occurrence of intracytoplasmatic sulfur globules and immunohistochemical analyses with an antiserum directed against RubisCO (Dubilier et al., 2001). For these endosymbionts the enzymology of oxidative sulfur metabolism and reductive sulfur metabolism is unknown. Furthermore it is important to establish which kind of sulfur substrates are used, and how they are metabolized, which could be possible only by looking for specific enzymes.

The present study found similarities for inferred translation products to proteins involved in sulfur assimilation, dissimilatory sulfur reduction and oxidation but also in polysulfide respiration. During this survey was found encoded a sulfite reductase which is involved in assimilation of sulfur into cysteine. Encoded APS reductase and dissimilatory sulfite reductase involved in both sulfur reduction and oxidation were also revealed. The genes encoding for these two enzymes are considered good “clocks” for the evolution of the dissimilatory sulfur metabolism for both SRB and sulfur oxidizing bacteria (Hipp et al.,
Another finding of this study was an encoded polysulfide reductase which shows the potential of \textit{O. algarvensis} symbionts to utilize polysulfide as final electron acceptor.

4.2.1. Sulfite reductase

Reductive sulfate assimilation is a pathway widely occurring in plants, algae, fungi and bacteria. The reduction of sulfite to sulfide is catalyzed by sirohaem sulfite reductases and is one of the final steps in cysteine biosynthesis. The electron donor for this the reduction of sulfite is NAD(P)H. Sulfide is immediately incorporated into O-acetyl-L-serine yielding cysteine (Lengeler et al. 1999, Neumann et al., 2000).

The inferred translation of one contig was similar to sulfite reductase hemoprotein beta subunit, from \textit{Chromobacterium violaceum}. This assimilatory sulfite reductase could belong to any member from \textit{O. algarvensis} endosymbiotic community, given that reductive sulfur assimilation occurs in any bacterial group.

4.2.2. APS reductase

APS reductase is found in sulfate-reducing bacteria and archaea, and in some chemoautotrophic and phototrophic sulfur oxidizing bacteria (Hipp et al. 1997). The reduction of APS and its conversion to sulfite and AMP is catalyzed by APS reductase in sulfate reducers, whether the same enzyme in sulfur oxidizers catalyzes the reverse reaction. In sulfur oxidizing bacteria two pathways for sulfite oxidation were described: direct oxidation via a sulfite acceptor-oxidoreductase and indirect AMP-dependent oxidation known as the APS pathway. The direct pathway for sulfite oxidation is considered to be far more widespread between sulfur oxidizers (Robertson & Kuenen 1992, Kappler & Dahl 2001) but for the sulfur-oxidizing endosymbionts of invertebrates so far the APS reductase pathway has been demonstrated (Nelson & Hagen, 1995).

The inferred translation of one contig was found similar to ATP sulfurylase from the phototrophic sulfur oxidizer \textit{A. vinosum}. In \textit{A. vinosum} an operon is formed by the gene for ATP sulfurylase (\textit{sat}) and APS reductase (\textit{apRMBA}) with \textit{aprA} and \textit{aprB} encoding for APS reductase α and β subunits and \textit{aprM} encoding a putative membrane anchor. The
two gamma endosymbiotic phylotypes are also closely related to *A. vinosum* at the 16S rRNA gene level (Rühland et al. In Prep), and corroborating with the fact that APS reductase is used as an evolutionary marker for dissimilatory sulfur metabolism (Hipp et al., 1997) seems reasonably to assign APS reductase to the gamma sulfur oxidizing symbionts.

### 4.2.3. Dissimilatory sulfite reductase

Dissimilatory sulfite reductase (DSR) in sulfate reducing bacteria catalyzes the reduction of sulfite to sulfide involving the transfer of six electrons whether in sulfur oxidizing bacteria the enzyme catalyze the reverse reaction.

The similarity between translation product of one contig to DsrF and DsrE and the end of the β subunit of DSR form *A. vinosum* corroborated to the 16S rRNA relationships to *A. vinosum* for the gamma symbionts from *O. algarvensis* ensures that this sequence might only belong to the sulfur oxidizing symbionts. In *A. vinosum* the *dsr* gene cluster has 15 genes *dsrABEFHCMKLJOPNSR*, and with interposon mutagenesis in *dsrB* and *dsrL* the was demonstrated the essential role of the Dsr proteins in the oxidation of intracellular sulfur. *dsrA* and *dsrB* encode the α and β subunits of dissimilatory sulfite reductase. DsrEFH together with DsrK, DsrAB, DsrC,DsrO and DsrJ are considered as being associated in a super complex involved in the oxidation of intracellular sulfur. The formation of all Dsr gene products is enhanced by sulfide (Dahl et al., 2005). Hereby inside *O. algarvensis* sulfide is currently produced by the sulfate reducers and might enhance the growth rates for the sulfur oxidizers.

The comparison of the nucleotide sequence of this contig against the sequence of a *dsr* locus obtained by sequencing the entire clone from the same BAC library (unpublished data Anna Blazejack) revealed a high sequence difference (Annex 3). This might be due to the fact that the sequences belong to a different gamma phylotype from the two known in *O. algarvensis*. 

42
4.2.4. Polysulfide reductase

Polysulfides are formed abiotically form sulfur and sulfide and its reduction to sulfide is catalyzed by polysulfide reductase. This process is typical for the catabolism of extremely thermophilic archaebacteria, of the thermophilic eubacterium *Thermotoga*, and of certain proteobacteria including *Desulfuromonas*, *Sulfurospirillum deleyianum* and *Wolinella succinogenes*. The function of polysulfide reductase is determined in *W. succinogenes* where this enzyme is a component of the phosphorylative transport system with polysulfide as terminal electron acceptor (Widdel and Hansen, 1992a; Krafft et al., 1995). By similarity search the inferred translation of one contig was related to polysulfide reductase from the green sulfur bacterium *C. tepidum*. The genome sequence of this organism revealed a *psrA* gene followed by *psrB* and *psrC*. The encoded function was not yet determined experimentally but it is assumed that the products of these genes might function in sulfide oxidation to form polysulfides (Eisen et al., 2002).

In *O. algarvensis* which is like a “closed system” accumulation of sulfur and sulfide could occur as a result of fast growth rates of bacterial cells, resulting in enough substrates for the chemical formation of polysulfides. It is possible that the sulfate-reducing endosymbionts could reduce polysulfides to sulfide but their energy yield would be rather low.

4.3. Nitrogen metabolism

Enzymes involved in ammonia release from urea (urease), synthesis of storage nitrogenous compounds (cyanophycin synthetase) or energy metabolism (dissimilatory nitrate reductase) were found encoded on clones from the BAC library of *O. algarvensis* symbionts. This shows the potential of these symbionts to use ammonia detoxification products, like urea released from the host metabolism but also the ability to synthesize storage compounds like cyanophycin, which is not a common trait among bacteria. The ability to respire nitrate for gaining energy seems also possible. Some other compounds like opines or sarcosine are also considered potential nitrogen sources besides their role as carbon sources for bacteria, and both are discussed in the carbon metabolism section.
4.3.1. **Urease**

Animals can either excrete ammonia directly, or synthesize urea or uric acid. Hydrolysis of urea to ammonia and carbon dioxide is catalyzed by the nickel containing enzyme, urease. Many eukaryotes synthesize urease, but urease activity is widely observed among prokaryotes, including many eubacteria and at least one archaeabacterium, *Methanobacterium thermoautotrophicum*. Microbial urease is associated in general with pathogenesis that causes direct damage to the host due to ammonium or alkali production (Mobley and Hausinger, 1989; Toffanin et al., 2002). Analysis of urease from *Klebsiella aerogenes*, the model organism for this enzyme, has revealed that the three catalytic subunits: \( \alpha \), \( \beta \) and \( \gamma \) associate in a \((\alpha\beta\gamma)_3\) structure containing three bi-nickel catalytic sites. Assembly of the bi nickel metallocenter requires accessory proteins, UreD, UreE, UreF and UreG (Lee et al., 1992).

By similarity search the inferred translations of two contigs were related to \( \alpha \) subunit of urease from the alphaproteobacterial plant pathogen *A. tumefaciens*, and to the urease accessory proteins UreF and UreG from the cyanobacterium *Nostoc* sp.(see Fig. 9).

A source of nitrogen for *O. algarvensis* endosymbionts could be the ammonia detoxification products from the worm, like urea; The problem remains unsolved since it is not possible to assign this enzyme to a symbiont in the community, but it shows that the community adapted to use the resources from its host.

4.3.2. **Periplasmic nitrate reductase**

There are two pathways for nitrate respiration and many organisms are able to use nitrate as electron acceptor, which is reduced to nitrite and further either via nitric oxide and nitrous oxide to molecular nitrogen (denitrification) or more directly reduced with NAD(P)H to ammonia (ammonification). The dissimilatory ammonification pathway differs from the assimilatory pathway by having the enzymatic systems integrated in the membranes or located in the periplasm (Lengeler et al., 1999). Hereby a periplasmic located nitrate reductase is considered as being involved in the dissimilatory reduction of nitrate to nitrite.
During this study was identified potential encoded periplasmic nitrate reductase with similarity to NapA from the epsilonproteobacterium *W. succinogenes*, NapG from the gammaproteobacterial pathogen *Salmonella enterica* and to NapH also from *W. succinogenes*. In *W. succinogenes* the gene cluster encoding for periplasmic nitrate reductase is *napAGHBFLD* (Simon et al., 2003) whether in *S. enterica* is *napFAGHBC*. Studies on respiratory nitrate reductase were done on *W. succinogenes* and it was proved with interposon mutation in napA that the only respiratory nitrate reductase is NapA. NapG and NapF might be involved in electron transfer while NapH in menaquinol oxidation (Simon et al., 2003)

This encoded enzyme shows the potential of *O. algarvensis* symbionts to reduce nitrate, and could be suggested that it belongs to the sulfate reducing symbionts forasmuch neither spirochete or sulfur reducing symbionts seem likely to perform nitrate respiration.

### 4.3.3. Cyanophycin synthetase

Cyanophycin is a branched, non-ribosomally synthesized polypeptide consisting of aspartate in the backbone and a roughly equimolar amount of arginine in the side chain. The capability for cyanophycin synthesis is not only restricted to cyanobacteria, as it was initially supposed, but it occurs among other bacterial groups representing a temporary nitrogen, energy and possibly carbon source (e.g. *Accinetobacter*, *Clostridium*, *Nitrosomonas*). Cyanophycin was described as accumulating under conditions of intense light at high carbon dioxide concentrations, and phosphate or sulfur starvation but also a high N/C ratio promotes cyanophycin biosynthesis. Cyanophycin synthetase is encoded by *cphA* and is involved in synthesis of cyanophycin via an ATP dependent mechanism (Krehenbrinck et al., 2002; Krehenbrinck and Steinbüchel, 2004).

The encoded cyanophycin synthetase from *O. algarvensis* symbionts was found similar to its counterpart in the betaproteobacterial human pathogen *Francisella tularensis* and it is intricate to make an assumption on the symbiont that might present cyanophycin synthetase capacity but it seems interesting to determine the reason for synthesizing storage compounds. Could be that compounds that are synthesized by such a
metabolically diverse community accumulate and instead of eliminating them in the environment they are stored by the microorganisms.

4.4. Transport proteins

Transport is an essential aspect of all life-endowing processes: metabolism, communication, reproduction and both cooperative and antagonistic inter-organismal behavior (Busch and Saier, 2004). In conclusion the transporters are appropriate for studying the interactions of the symbiotic community from O. algarvensis. The translation products for three sequences were related to proteins like ferrous iron transport protein, natrium driven multidrug efflux pump and spermidine/putresceine transporter.

4.4.1. Ferrous iron transport protein B

Most bacteria that live in oxic environments produce iron (III) chelators, called siderophores, to satisfy their needs for iron, which forms highly insoluble ferric hydroxides at neutral pH. The ferrous iron (II) uptake proteins was proved in E. coli to be encoded by the feoAB (Kammler et al., 1993) and now in almost 50 % of the bacterial genomes a feoB-like gene was found. The FeoB protein from E. coli was identified as having GTPase activity that is necessary for ferrous iron (II) transport whether the uptake is directly accomplished by the FeoB protein. The iron transport function of Feo appears to be of particular importance during low oxygen conditions when ferrous iron (II) predominates over ferric iron (III) (Andrews et al., 2003; Hantke, 2003). The ability to utilize ferrous iron efficiently is considered to be a major factor in pathogenesis for those microbes that try to colonize essentially anaerobic environments (Guerinot, 1994). During the present study the determined translation of one contig had high similarity to FeoB from the pathogen Porphyromonas gingivalis. This ferrous iron transporter shows that the bacteria adapted to the oligochaete host, while living in conditions of low oxygen
concentration, inside the worm, are able to take up their essential micronutrients, iron being one of them.

4.4.2. Na+ driven multidrug efflux pump

Drug efflux mechanism is one of the major mechanisms of drug resistance. Multidrug efflux pumps are integral membrane transporters that have the ability to extrude a variety of structural unrelated drugs from the cell. NorM is a multidrug efflux pump driven by an electrochemical potential of Na+ instead of a proton motive force or ATP dependency like in other categories of transporters (Brown et al., 1999). A Na+ driven multidrug efflux pump was described for the marine bacterium *Vibrio parahaemolyticus* with a homologue in *E. coli* (Morita et al., 1998; Morita et al., 2000) and in the causative agent of cholera, *Vibrio cholerae* (Huda et al., 2003).

One clone from *O. algarvensis* BAC library after contig assembly was identified as bearing the encoded information for NorM similar to its correspondent from *Rhodopseudomonas palustris*, showing the possible ability of these symbionts to extrude drugs.

4.4.3. Spermidine/putrescine transporter

Polyamines (putrescine, spermidine and spermine) are necessary for cell growth. Polyamines are among the major polycations in cells together with Ca$^{2+}$ and Mg$^{2+}$ and their content in cells is mediated by biosynthesis, degradation and transport. The putrescine-specific and the spermidine-preferential uptake systems are ABC transporters. Polyamine transport systems were identified in most bacteria for which whole genome were sequenced and it even exists in archaea. (Igarashi and Kashiwagi, 1999)

During this study was found a PotF encoded enzyme similar to PotF from *Fusobacterium nucleatum*. PotF is the substrate binding protein and its presence shows that polyamines might be used as metabolites by the symbionts of *O. algarvensis*.
4.5. **Pathogenic/symbiotic relevant traits**

Pathogenic associated traits correlate with the expression of disease related factors that are present in pathogenic organisms and lack from the other non-pathogenic species. Symbiotic associated traits are represented by various factors which can increase the adaptability and versatility of the bacterium. Some pathogenic associated traits are also met in symbiotic organisms, offering to the last ones a better fitness, and several of these pathogenic associated traits were found in the genome on *O. algarvensis* symbionts: ferrous iron transporter, urease and virulence associated proteins but also opine oxidase.

4.5.1. **Virulence associated proteins BC**

*vapABCD* genomic region was first characterized in the pathogen *Dichelobacter nodosus* (Katz et al., 1992) and was associated with virulence. Novel studies on *Leptospira interrogans* associated the vapBC to physiological regulation and not to virulence, being characterized as a toxin-antitoxin module (Zhang et al., 2004). The same study stated that vapBC is widely present and expressed in both non pathogenic and pathogenic leptospires. Therefore it was considered more related to physiological regulation than to virulence. It was demonstrated in the same study that the product of vapC gene is toxic for *E.coli* cells, while vapB gene product counteracts VapC toxicity.

A region potentially encoding for vapBC was found in this project by similarity search and was related to VapB from *Leptospira interrogans* and VapC from *Geobacter sulfurreducens*. It is possible that in *O. algarvensis* symbionts its presence is associated to physiological regulation than to virulence, enhancing the fitness of the symbionts the oligochaete host.
## Annex 1

List of result obtained by similarity search against protein data bases; The E-values represents the number of hits that can be seen just by chance searching a data base of a particular size.

<table>
<thead>
<tr>
<th>Type of metabolism</th>
<th>BAC</th>
<th>Vector Primer</th>
<th>Encoded protein 1</th>
<th>E-value</th>
<th>Encoded protein 2</th>
<th>E-value</th>
<th>Encoded protein 3</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon metabolism</td>
<td>9B5</td>
<td>SP6</td>
<td>CO dehydrogenase (<em>M. mazei</em>)</td>
<td>1E-07</td>
<td>Protein 3 Com E</td>
<td>1</td>
<td>Sensory transduction histidine kinase (uncultured archeon)</td>
<td>8E-14</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>5C6</td>
<td>SP6</td>
<td>CO dehydrogenase (<em>Rugeria</em> sp.)</td>
<td>9E-70</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>1B4</td>
<td>T7</td>
<td>Formate dehydrogenase Subunit α (<em>S. pomeryyi</em>)</td>
<td>1E-52</td>
<td>Formate dehydrogenase Subunit Fe-S (<em>S. pomeryyi</em>)</td>
<td>1E-76</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>3B6</td>
<td>SP6</td>
<td>Trimethylamine methyltransferase Mtb1 (<em>M. loti</em>)</td>
<td>E-132</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>9A11</td>
<td>T7</td>
<td>Trimethylamine methyltransferase Mtb1 (<em>M.acetivorans</em>)</td>
<td>6E-10</td>
<td>-</td>
<td>-</td>
<td>Spermidine/putrescine ABC transporter (<em>F. nucleatum</em>)</td>
<td>1E-06</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>8E6</td>
<td>T7</td>
<td>Sarcoine oxidase (S. pomeryyi)</td>
<td>E-122</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>8G3</td>
<td>T7</td>
<td>Opine oxidase subunitβ (S. meliloti)</td>
<td>1E-24</td>
<td>NAD/FAD dehydrogenase (<em>B. fungorum</em>)</td>
<td>1E-17</td>
<td>Opine oxidase subunitA (B. japonicum)</td>
<td>9E-38</td>
</tr>
<tr>
<td>Sulfur metabolism</td>
<td>9C10</td>
<td>T7</td>
<td>Sulfite reductase β subunit Cysl (C. violaceum)</td>
<td>3E-39</td>
<td>Protein chain releasing Factor B (M. degradans)</td>
<td>3E-45</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sulfur metabolism</td>
<td>2F5</td>
<td>SP6</td>
<td>Adenylylsulfate reductase α subunit AprA (<em>A. vinosum</em>)</td>
<td>4E-78</td>
<td>-</td>
<td>-</td>
<td>FKB type peptidyl prolyl cis-trans Isomerase (<em>T. denitrificans</em>)</td>
<td>3E-01</td>
</tr>
<tr>
<td>Sulfur metabolism</td>
<td>1H4</td>
<td>T7</td>
<td>DsrF (<em>A. vinosum</em>)</td>
<td>1E-48</td>
<td>DsrE (<em>A. vinosum</em>)</td>
<td>1E-63</td>
<td>Dissimilatory sulfite reductase β subunit, DsrB (<em>A. vinosum</em>)</td>
<td>3E-03</td>
</tr>
<tr>
<td>Sulfur metabolism</td>
<td>9C10</td>
<td>T7</td>
<td>Polysulfide reductase subunit A (C. tepidum)</td>
<td>1E-77</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>8D1</td>
<td>T7</td>
<td>UreF (Nostoc sp.)</td>
<td>2E-15</td>
<td>UreE (Nostoc sp.)</td>
<td>3E-69</td>
<td>Transposase (Polaromonas)</td>
<td>9E-21</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>9D3</td>
<td>T7</td>
<td>Urease α subunit, UreC (A. tumefaciens)</td>
<td>2E-15</td>
<td>-</td>
<td>-</td>
<td>Anthranilate synthase component II (C. efficiens)</td>
<td>6.5</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>2D12</td>
<td>SP6</td>
<td>Periplasmatic nitrate reductase Subunit A, NapA (<em>W. succinogenes</em>)</td>
<td>8E-40</td>
<td>NapG (<em>S. enterica</em>)</td>
<td>2E-42</td>
<td>NapH (<em>W. succinogenes</em>)</td>
<td>1.4</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>10H10</td>
<td>T7</td>
<td>Cyanophycin synthetase (F. tularensis)</td>
<td>1E-35</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Transports</td>
<td>2D4</td>
<td>SP6</td>
<td>Ferrous iron transport protein FeoB (<em>P. gingivalis</em>)</td>
<td>3E-63</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Transports</td>
<td>9H6</td>
<td>SP6</td>
<td>Na⁺ driven multidrug efflux pump (<em>R. palustris</em>)</td>
<td>4E-29</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pathogenic/symbiotic Traits</td>
<td>10H7</td>
<td>SP6</td>
<td>Virulence associated protein B VapB (<em>L. interrogans</em>)</td>
<td>3E-07</td>
<td>Virulence associated protein C VapC (<em>L. interrogans</em>)</td>
<td>2E-24</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Annex 2
BioEdit alignment of the five clones found as bearing a gene encoding for RubisCO. Boxed nucleotides represent the conserved regions whether unboxed parts shows the differences between the two DNA fragments.
Annex 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>GA</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AC</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>CT</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>CG</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>CA</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>AC</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>CG</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>CA</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>AC</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>CG</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>CA</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>AC</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>CG</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>CA</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>AC</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>CG</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>CA</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>GT</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>AC</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>AT</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>TG</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>
Annex 3
BioEdit alignment of the contig assembled on BAC clone 1H4 against the region that contains a *dsr* operon, from a completely sequenced BAC from *O. algarvensis* library. Boxed nucleotides represent the conserved regions whether unboxed parts shows the differences between the two DNA fragments.


homologues to cyanophycin synthetase and cloning of an active cyanophycin synthetase from Acinetobacter sp. strain DSM 587. Archives of Microbiology 177: 371-380.


Aknowledgements

I want to thank to Nicole Dubilier for supervision, to Anna Blazejak for good advices during my work in the Molecular Ecology Department, to Silke Wetzel, Anke Meyerdierks and Michael Richter and not the last to Niculina Musat for her support and advices.

I thank to the entire Molecular Ecology group for the great working atmosphere and finally to Max Planck Geselshaft for the financial support.