Metaproteogenomic profiling of chemosynthetic microbial biofilms reveals metabolic flexibility during colonization of a shallow-water gas vent

Sushmita Patwardhan\textsuperscript{1}, Francesco Smedile\textsuperscript{1,2}, Donato Giovannelli\textsuperscript{1,3,4,5} and Costantino Vetriani\textsuperscript{1,6}\textsuperscript{*}

\textsuperscript{1} Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA
\textsuperscript{2} CNR, Institute for Coastal Marine Environment, Messina, Italy
\textsuperscript{3} Department of Biology, University of Naples "Federico II", Naples, Italy
\textsuperscript{4} CNR, Institute for Marine Biological and Biotechnological Resources, Ancona, Italy
\textsuperscript{5} Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan
\textsuperscript{6} Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, USA

Correspondence:
Costantino Vetriani
Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, USA
vetriani@marine.rutgers.edu

Running title: Metaproteome of chemosynthetic microbial biofilms

Keywords: shallow-water gas vent, Mediterranean Sea, Tor Caldara, microbial filaments, biofilms, metaproteome, metagenome, metabolic profile, Epsilonproteobacteria, Gammaproteobacteria
Abstract

Tor Caldara is a shallow-water gas vent located in the Mediterranean Sea, with active venting of CO₂, H₂S. At Tor Caldara, filamentous microbial biofilms, mainly composed of *Epsilon*- and *Gammaproteobacteria*, grow on substrates exposed to the gas venting. In this study, we took a metaproteogenomic approach to identify the metabolic potential and *in situ* expression of central metabolic pathways at two stages of biofilm maturation. Our findings indicate that inorganic reduced sulfur species are the main electron donors and CO₂ the main carbon source for the filamentous biofilms, which conserve energy by oxygen and nitrate respiration, fix dinitrogen gas and detoxify heavy metals. Three metagenome-assembled genomes (MAGs), representative of key members in the biofilm community, were also recovered. Metaproteomic data show that metabolically active chemoautotrophic sulfide-oxidizing members of the *Epsilonproteobacteria* dominated the young microbial biofilms, while *Gammaproteobacteria* become prevalent in the established community. The co-expression of different pathways for sulfide oxidation by these two classes of bacteria suggests exposure to different sulfide concentrations within the biofilms, as well as fine-tuned adaptations of the enzymatic complexes. Taken together, our findings demonstrate a shift in the taxonomic composition and associated metabolic activity of these biofilms in the course of the colonization process.
Introduction

Marine gas vents transport volatile elements and compounds from the geosphere to the hydrosphere by seepage through sediments and bedrocks (Suess, 2014). At these sites, geological processes, biogeochemical reactions and the activity of microorganisms act together to alter the composition of volatile products. The release of such reduced volatiles into the oxygenated water column generates a redox disequilibrium that can be harnessed by benthic prokaryotic communities to convert chemical energy into ATP. Hydrothermal and gas vents occur both in shallow water (depth < 200 meters) and in the deep-sea (depth > 200 meters; Tarasov et al., 2005). Shallow-water hydrothermal and gas vents are associated with submarine volcanoes, arc and back-arc volcanoes and occur globally in the proximity of active plate margins and intraplate hotspots (reviewed in Price and Giovannelli, 2017).

The subduction of the African plate below Europe has resulted in the formation of the Mediterranean Ridge and deep subduction basins as well as active volcanic arcs in the Tyrrhenian and Aegean Seas (Dando et al., 1999). The terrestrial volcanic systems in these areas have been well-studied, both because of the long history of devastation caused by their eruptions and because of their geothermal potential. The submarine parts of the system have received relatively little attention until the last decade, and are still poorly studied compared to some of the mid-ocean ridge systems (Price and Giovannelli, 2017). Volcanic arc hydrothermal systems release large volumes of volatiles (Barry et al., 2019), because of the both degassing of the subducted slab and the mantle, and the decomposition of carbonates in the overlying marine sediments (Fisher, 2008). Since most of the known venting in the Mediterranean is from shallow vents, the majority of the outlets are of the gasohydrothermal type, emitting large volumes of carbon dioxide.

The microbiology of the shallow-water vent systems of the Aegean and Tyrrhenian basins has been mainly investigated in sediments (Giovannelli et al., 2012; Gugliandolo and Maugeri, 1993; Kerfahi et al.; Maugeri et al., 2013; Price et al., 2013; Sievert et al., 1999, 2000). However, besides sediment communities, substrate-attached chemosynthetic microbial biofilms are widespread and, in their role of primary producers, relevant in these ecosystems (Kalanetra et al.,...
Here, we investigate two types of biofilm communities at a shallow-water gas vent located at Tor Caldara, Italy.

Tor Caldara is a natural reserve located near the town of Anzio, about 60 km south of Rome. Active venting of gases, including CO$_2$ originating from a deep magma source, occurs at Tor Caldara in association with the quiescent volcanic complex of Colli Albani, located around 40 km north-east of the venting site (Carapezza and Tarchini, 2007). The absence of a thermal anomaly at Tor Caldara is attributed to the infiltration of cold meteoric water (Carapezza et al., 2012). In a previous study, we investigated the chemical composition of the gases at Tor Caldara and the taxonomic diversity of the associated filamentous biofilm communities (Patwardhan et al., 2018). The gases at Tor Caldara are mainly composed of CO$_2$ (avg. of 76.67 mol%) and H$_2$S (avg. of 23.13 mol%) with minor contribution of CH$_4$ (avg. of 0.18 mol%), CO (avg. of 0.0080 mol%) and H$_2$ (avg. of 0.00072 mol%). Young and established filamentous biofilms - the former collected on glass slides deployed at the venting site at Tor Caldara and the latter attached to the native rocky substrates - were dominated by members of the classes Epsilonproteobacteria and Gammaproteobacteria, respectively, albeit in different proportions. On average, sulfur-oxidizing Epsilonproteobacteria of the genus Sulfurovum accounted for 57.6% of the active young biofilms, while sulfur-oxidizing Gammaproteobacteria of the genus Thiomicrospira as well as sequences related to the Thiothrix CF-26 group constituted more than 60% of the active established biofilm community (Patwardhan et al., 2018). This observed transition from Epsilon- to Gammaproteobacteria during the maturation of the biofilms revealed an ecological succession between the young and established communities. Previous studies of sulfidic environments, including caves, shallow-water and deep-sea hydrothermal vents, reported a spatial segregation between these two classes of bacteria and correlated this distribution pattern with geochemical data, leading to the hypothesis that Epsilonproteobacteria can cope with higher concentrations of hydrogen sulfide than Gammaproteobacteria (Engel et al., 2004; Giovannelli et al., 2013; Macalady et al., 2008; Meier et al., 2017; Miranda et al., 2016; O’Brien et al., 2015; Reigstad et al., 2011). Based on these studies, and on the observed succession observed at Tor Caldara, we tested the hypothesis that Epsilon- and Gammaproteobacteria are adapted to different concentrations of sulfide. Sulfide gradient experiments with laboratory strains showed that, conservatively, Epsilonproteobacteria can tolerate sulfide concentrations up to 20 times higher.
than *Gammaproteobacteria*, possibly facilitating their role as pioneer colonizers of sulfidic environments (Patwardhan *et al.*, 2018).

To conserve energy, prokaryotes oxidize reduced inorganic sulfur compounds via a number of different pathways. The Sox pathway is widespread and occurs in both anaerobic as well as aerobic phototrophic and chemolithoautotrophic microorganisms. The well-characterized Sox enzyme complex in the alphaproteobacterium, *Paracoccus pantotrophus*, has four periplasmic proteins: SoxYZ, SoxXA, SoxB and SoxCD. These proteins work in concert to oxidize thiosulfate, hydrogen sulfide, sulfur as well as sulfite all the way to sulfate. In some sulfur-oxidizing bacteria (SOB) such as *Rhodobacter capsulatus*, sulfide is oxidized to sulfur by the membrane bound sulfide quinone oxidoreductase (SQR) enzyme. Additionally, in few SOBs, sulfide is also oxidized to sulfur by the periplasmic sulfide dehydrogenase (FccAB) (Visser *et al.*, 1997; Müßmann *et al.*, 2007).

In this study, we used an integrated metagenomic/metaproteomic approach to compare the metabolic potential and the *in situ* expression of central metabolic pathways of the established and young filaments at Tor Caldara. We identified and quantified key genes and proteins involved in the carbon, sulfur and nitrogen cycles and compared them between the young and established biofilms to establish how different metabolic pathways were expressed in the two communities. We also reconstructed three metagenome-assembled genomes (MAGs) belonging to the classes *Epsilonproteobacteria* and *Gammaproteobacteria*.

**Materials and Methods**

**Study Site**

Tor Caldara is a natural reserve located approximately 60 km south of Rome and 40 km southwest of Colli Albani, Italy (Patwardhan *et al.*, 2018). The study site (41°29′ 9″ N 12°35′ 23″ E) is a coastal submarine gas vent located at a depth of approximately 3 meters. The site is characterized by vigorous venting of gases that escape from the sandy seabed. The sediment in the venting area is distinctly darker in color than the control sediment because of possible sulfide
deposition. Conspicuous growth of white microbial filaments is seen on the rocks directly exposed to the gas venting.

**Sample Collection**

Sampling of filaments was performed as described previously (Patwardhan *et al.*, 2018). Briefly, established filaments (EF from here on) growing on rocks exposed to the gas venting were collected by a SCUBA diver in August 2016. To study the early stages of colonization, young biofilms (YF from here on) were collected on sterile glass slides mounted on an aluminum rod and exposed to the venting area for four days. Biofilms samples were then stored in RNA Later at -80°C for further nucleic acid and protein extraction. Representative samples of the two communities (one EF and one YF) were further used for downstream applications.

**Nucleic acid extraction, metagenomic library preparation and analysis**

DNA was extracted from YF and EF biofilm biomass stored in RNA Later following a phenol:chloroform extraction protocol. Briefly, 0.8 g of sample were added to 850 µl of extraction buffer (50 mM Tris-HCl, 20 mM EDTA, 100 mM NaCl; pH 8.0) supplemented with 100 µl of lysozyme (100 mg/ml) and incubated at 37°C for 30 min. This mix was then supplemented with 5 µl of proteinase K (20 mg/ml), incubated at 37 °C for 30 min, and subsequently supplemented with 50 µl SDS (20%) and further incubated at 65 °C for 1 hr. Nucleic acids were extracted by performing a series of phenol:chloroform:isoamylalcohol (25:24:1) and chloroform:isoamyl alcohol (24:1) extractions. Multiple samples were co-extracted to reduce potential bias. The final supernatant was precipitated in 3 M sodium-acetate and isopropanol, washed twice with 70% ice-cold ethanol and re-suspended in ultra-pure sterile water. The library was prepared at Molecular Research LP (Shallowater, TX), using the Nextera DNA Sample preparation kit (Illumina) following the manufacturer's user guide. The initial concentration of DNA was evaluated using the Qubit® dsDNA HS Assay Kit (Life Technologies). Initial library preparation resulted in very large size library (>2500 bp), therefore, library preparation was repeated after DNA inhibitor removal using the DNEasy PowerClean Pro Cleanup Kit (Qiagen). 50 ng DNA was used to prepare the library. The sample underwent the simultaneous fragmentation and addition of adapter sequences. These adapters are utilized during a limited-cycle (5 cycles) PCR in which unique indices were added to the sample. Following the
library preparation, the final concentration of the library was measured using the Qubit® dsDNA HS Assay Kit (Life Technologies), and the average library size was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies). The library was diluted (to 10 pM) and sequenced paired end for 500 cycles using the HiSeq platform (Illumina). Following the removal of the adapters, at the sequencing facility, the sequences were quality checked using FastQC v.0.11.5 (Andrews S., 2010. FastQC: a quality control tool for high throughput sequence data: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). A total of 6,698,798 and 10,125,188 paired end sequences were used for downstream analysis for the EF and YF samples respectively. Sequences encoding rRNA genes were extracted using Metaxa2 v2.1 using default parameters (Bengtsson-Palme Johan et al., 2015) and annotated using Silva ngs v.1.3.9 (Quast et al., 2013). The paired end sequences were then assembled using Megahit v1.1.3 (Li et al., 2015) with default parameters, and assembly was checked using QUAST v3.4 (Gurevich et al., 2013). The EF assembly had a total of 99,137 contigs with the largest contig being 212,020 bp long, whereas the YF assembly had a total of 87,836 contigs with the largest contig being 261,286 bp long. Assembled metagenomes were submitted to DOE Joint Genome Institute’s Integrated Microbial Genome Metagenomic Expert Review (IMG/ER) pipeline for Open Reading Frame (ORF) identification as well as functional and taxonomic annotation (Markowitz et al., 2012). The EF and YF metagenomic assemblies had a total of 308,874 and 362,063 sequences, respectively. The sequences of both metagenomes had COG (51%), KO (41%) and KEGG (25%) annotations. Post annotation, using default parameters for bowtie2 v.2.3.3.1, the reads were mapped back to the assembly (Langmead et al., 2009). Using HTSeq v0.9.1, the number of reads mapping to each gene was counted (Anders et al., 2015). Custom scripts were used to normalize absolute read counts mapped to each gene for varying gene lengths and difference in sequencing depth between samples. These normalized gene abundances were expressed as TPM (Transcripts Per Million; (Wagner et al., 2012). Sequences are available through the NCBI Short Read Archive database with accession number PRJNA498803.

**Protein extraction and metaproteomic analysis**

Total proteins were extracted from the YF and EF biofilms by addition of Laemmli buffer, followed by sonication and incubation at 95 °C. Proteins were then precipitated by addition of urea, purified and digested with trypsin on a SDS-PAGE gel, and then label-free samples were
run on the LC-MSMS. The resulting tandem mass spectra were searched against the predicted peptide sequences encoded by the combined metagenomes of the two filament samples, using the open-source software X!Tandem (Craig and Beavis, 2004). A total of 138,387 and 97,607 peptides were obtained in the metaproteomes of EF and YF, respectively. The proteins identified via X!Tandem were functionally and taxonomically annotated using the corresponding annotated metagenomes. Protein counts were normalized to the total proteins obtained from each sample. Normalized counts for proteins of interest were manually compared between the two samples. These normalized counts were expressed as a percent of total proteins observed in each sample. Raw data was uploaded to the ProteomeXchange database.

Whole genome reconstruction from metagenomes
Quality checked reads from both samples were co-assembled using Megahit v1.1.3 with default parameters and QUAST v3.4 was used to check the quality of the co-assembly. The co-assembled contigs from the two metagenomes were then binned using MaxBin2.0 (Wu et al., 2014) to recover individual genomes. Completeness, taxonomic affiliation and contamination of recovered bins was checked using CheckM v1.0.7 (Parks et al., 2015).

Phylogenetic Analyses
Sulfide quinone oxidoreductase (SQR) amino acid sequences obtained in this study and from GenBank were aligned with ClustalX v 2.0 (Thompson et al, 1997) and manually adjusted using Seaview (Galtier et al, 1996). Phylogenetic distances were calculated using the LG amino acid replacement matrix (Le and Gascuel, 2008) and the maximum likelihood method was used to evaluate tree topologies in PhyML 3.0 (Guindon et al., 2010). Branch support was estimated using the approximate likelihood ratio test (aLRT; (Anisimova and Gascuel, 2006) and by bootstrap analysis with 1,000 resamplings. Both estimates gave comparable results.

Results

Microbial community composition of the established and young filaments
The bacterial diversity of EF and YF was evaluated following in-silico extraction of the 16S rRNA gene sequences from the two metagenomes. The EF biofilm was dominated by Gammaproteobacteria (54%), followed by Epsilonproteobacteria (29%) and small proportions
of Bacteroidia (4%) and Alphaproteobacteria (3%; Fig. 1a). Within the Gammaproteobacteria, the genus Thiomicrospira (40%; now reclassified as Thiomicrorhabdus; Boden et al., 2017) and uncultured members belonging to the Thiotrichales (11%) were dominant (Fig. 1b). A reverse trend was observed in the YF biofilm, with Epsilonproteobacteria (67%) dominating the community, followed by Gammaproteobacteria (23%), Bacteroidia (2%) and Alphaproteobacteria (1%; Fig. 1a). In the YF biofilm, Sulfurovum (66%) was the most abundant genus within the Epsilonproteobacteria (Fig. 1b).

**Comparative metagenomics and metaproteomics of the established and young filaments**

We compared the normalized abundances of key genes and proteins involved in central metabolic pathways, including carbon fixation, nitrogen, sulfur metabolism and (micro)aerobic respiration (Suppl. Tables 1 and 2). Since shallow-water vents are enriched in heavy metals, genes and proteins involved in their detoxification were also identified and compared (Suppl. Tables 1 and 2). In line with the 16S rRNA gene-inferred community composition, the majority of functional genes and proteins were affiliated with Gamma- and Epsilonproteobacteria, (Figs. 2, 3 and 4). The metagenomes provided a snapshot of the metabolic potential of the microbial communities, and were also used to annotate the metaproteomes, which revealed the *in situ* expression of central metabolic pathways of the communities. The metagenomes were also used to obtain metagenome-assembled genomes (MAGs) of the most abundant members of the filamentous biofilms.

**Carbon fixation.** The metagenome and the metaproteome of both filamentous communities included genes and proteins involved in the reductive tricarboxylic acid (rTCA) and Calvin-Benson-Bassham (CBB) cycles for carbon fixation (Fig. 2). The YF biofilm community had three times as many genes for the rTCA than the EF biofilm, while EF had twice as many genes for the CBB cycle than YF (Fig. 2a, columns 1 and 3, and Suppl. Table 1). While proteins of the rTCA cycle contributed only to 2.2% of the total proteins in EF, they contributed to 5% of the total in YF (Fig. 2b, columns 1 and 3, and Suppl. Table 2). Taxonomically, the rTCA proteins were all affiliated with *Sulfurovum* and *Sulfurimonas* within the Epsilonproteobacteria (Fig. 2b, columns 2 and 4). In contrast, proteins of the CBB cycle comprised 1.1% and 0.7% of the total proteins in EF and YF, respectively (Fig 2b, columns 1 and 3, and Suppl. Table 2), with most of them taxonomically identified as Gammaproteobacteria and specifically affiliated to the genera.
Thiomicrospira, Thiothrix and Thioalkalivibrio (Fig. 2b, columns 2 and 4). These bacteria are chemolithotrophs that derive energy from the oxidation of reduced sulfur species and are key players in marine geothermal environments.

**Nitrogen and sulfur metabolism.** The metagenomes of the two biofilm communities encoded for genes involved in nitrate reduction, denitrification, nitrogen fixation, sulfur oxidation/reduction, oxygen respiration and heavy metal detoxification (Fig. 3a). However, the abundance of some of these genes differed between the two biofilm communities. For instance, the napAB (encoding two components of the periplasmic nitrate reductase complex) and nifDHK (nitrogenase) genes, involved in nitrate reduction and nitrogen fixation, respectively, were found to be enriched in the YF biofilm (Fig. 3a and Suppl. Table 1) and were taxonomically affiliated with the Epsilonproteobacteria (Fig. 3b). The narGHI (membrane bound nitrate reductase) and nirS (nitrite reductase) genes, as well as those for sulfur oxidation, fccAB (sulfide dehydrogenase), and sulfate reduction/sulfide oxidation, dsrAB (dissimilatory bisulfite reductase) were enriched in EF (Fig. 3a and Suppl. Table 1). The sor gene (sulfur oxygenase/reductase) was found in EF but was absent in YF. Taxonomically, the nar, nir and sor genes were mostly affiliated with the Gammaproteobacteria (Fig. 3b). The sqr (sulfide quinone oxidoreductase) gene was also enriched in the EF biofilm community and predominantly affiliated with the Gammaproteobacteria (Figs. 3a and b). However, the proportion of Epsilonproteobacteria-affiliated sqr genes increased in the YF community (Fig. 3b).

In the metaproteome, the Epsilonproteobacteria-affiliated nitrogenase (Nif) accounted for 0.35% of the total proteins in EF as opposed to 0.001% in YF (Fig. 4a and Suppl. Table 2). The periplasmic nitrate reductase (Nap) was undetectable in both the samples, while both the membrane bound nitrate reductase (Nar) and the nitrite reductase (Nir), affiliated with the Gammaproteobacteria, were more abundant in EF (0.02% and 0.3%; Fig. 4a). Proteins involved in the dissimilatory sulfite reductase (Dsr) pathway for either elemental sulfur oxidation or sulfite reduction and the APS pathway for indirect sulfite oxidation (Apr, SAT) were twice as abundant in the EF biofilm compared to the YF one (Fig. 4a and Suppl. Table 2). Dsr and Apr in both metaproteomes were mostly affiliated with Gammaproteobacteria, while SAT was mostly affiliated with Gammaproteobacteria in the EF biofilm and with both Gammaproteobacteria and Epsilonproteobacteria in the YF community (Fig. 4b). The sulfide dehydrogenase (Fcc) was twice more abundant in the EF than in the YF biofilm (Fig. 4a) and it was affiliated with
Gammaproteobacteria in both samples (Fig. 4b and Suppl. Table 2). In contrast, the abundance of the sulfide quinone oxidoreductase (SQR) was comparable between the two biofilms (Fig. 4a), but its taxonomic affiliation changed from comparably epsilon- and gammaproteobacterial in the EF biofilm to predominantly epsilonproteobacterial in the YF community (Fig. 4b and Suppl. Table 2). The Sor protein, involved in the disproportionation of elemental sulfur (Kletzin, 1989), was classified as Thioalkalivibrio spp. and was enriched in the EF biofilm (Fig. 4a and Suppl. Table 2). A putative NADPH-dependent sulfur reductase (NSR) affiliated with Sulfurovum spp. was six times more abundant in the total proteins of YF than EF (Fig. 4a and Suppl. Table 2).

The sox genes and the Sox proteins were almost equally abundant in both the YF and EF biofilms (Figs. 3a and 4a); however, in the YF biofilm, both sox genes and Sox proteins were predominantly affiliated with the Epsilonproteobacteria, while in the EF biofilm they were mostly affiliated with the Gammaproteobacteria (Figs. 3b and 4b). We did not find any hydrogen oxidation-related genes or proteins in the metagenomes and metaproteomes of the two biofilm communities.

Oxygen reduction. Three cytochrome oxidase gene clusters (coxABCD, ccoNOP and cydAB) were abundant and comparable between the two metagenomes, indicating that oxygen respiration played an important role in both biofilm communities (Fig. 3a and Suppl. Table 1). In the YF biofilm, the majority of all three genes were taxonomically classified as Epsilonproteobacteria, while in the EF biofilm the taxonomic affiliation of coxABCD and cydAB revealed a higher diversity of the oxygen respiring community, except for ccoNOP, which was mostly affiliated with the Gammaproteobacteria (Fig. 3b). Compared to proteins involved in nitrate/nitrite respiration, the aa3-type and cbb3-type cytochromes, mediating oxygen respiration, were highly enriched (seven times as much) in YF (Fig. 4a and Suppl. Table 2). The majority of these proteins were classified as Epsilonproteobacteria (Fig. 4b) and were specifically affiliated with Sulfurovum spp., with the exception of cytochrome cbb3 in EF, which was equally affiliated with Epsilon- and Gammaproteobacteria (Fig. 4b). The cytochrome d (CydAB) was mostly affiliated with Gammaproteobacteria (Fig. 4b).

Heavy metal detoxification. Elevated concentrations of potentially toxic heavy metals such as mercury, arsenic, selenium etc. are present at hydrothermal vents, leading to heavy metal resistance in the resident microbiota (Rathgeber et al., 2002; Vetriani et al., 2005; Price et al., 2013). Genes involved in the detoxification of arsenate (arsC), mercury (merA) and selenate
(dedA) were recovered from both metagenomes (Fig. 3a). The abundance of the arsC gene was comparable in both samples (Fig. 3a and Suppl. Table 1), while this gene was affiliated mostly with *Gammaproteobacteria* and *Epsilonproteobacteria* in the EF the YF, respectively (Fig. 3b). The merA gene (mercury reduction) was more abundant in the YF than the EF metagenome (Fig. 3a and Suppl. Table 1), and was affiliated with the *Gammaproteobacteria*, while the dedA gene (selenate detoxification) was more abundant in the EF and it was mostly affiliated with the *Epsilonproteobacteria* (Figs. 3a and b). In the metaproteome, only DedA, which has been shown to be involved in selenite resistance (Ledgham *et al.*, 2005), was expressed in the YF biofilm, as opposed to both DedA and arsenate reductase (Ars) being expressed in EF (Fig. 4a and Suppl. Table 2). Both enzymes were classified as *Epsilonproteobacteria* (Fig. 4b). While present in both metagenomes, the mercuric reductase gene (*merA*) was not found in the metaproteomes of either biofilm communities (Figs. 3a and b).

**Metagenome-assembled genomes (MAGs)** Despite the fundamental role as primary producers of Gamma and Epsilonproteobacteria in marine geothermal habitats, there is a dearth of representative cultures and their genomes. A total of 31 MAG bins were obtained from the co-assemble of the two metagenomes. Out of these, three (TCMF1, TCMF2 and TCMF9) were ≥ 85% complete with contamination of ≤ 8% and complied with the proposed criteria that define medium to high quality draft MAGs (Bowers *et al.*, 2017). The 16S rRNA genes were recovered from bins TCMF2 and TCMF9 (Table 1). The 16S rRNA gene of MAG TCMF2 showed >97% sequence similarity to 16S rRNA genes related to the *Thiotrix*-related CF-26 group recovered from microbial mats at shallow-water vents located on the South Tonga Arc (HQ153886) (Murdock *et al.* 2010) and at the Kuieshan Island (JQ611107). In our previous study of the Tor Caldara vents, dominant OTUs annotated as *Thiotrix*-related CF-26 from EF showed similar sequence similarity to the OTUs recovered from the same microbial mats, indicating that the TCMF2 MAG likely belongs to this group of uncultured *Thiotrichales*. Most of the genes involved in the CBB cycle, including RuBisCo (*rbcSL*), were present in MAG TCMF2. Further, the potential for thiosulfate oxidation was evidenced by the presence of *sox* genes, in addition to the *sqr* and *fccAB* genes involved in sulfide oxidation. This MAG also contained the genes for cbb3-type cytochrome oxidase and *nirBD*, involved in microaerobic and nitrite respiration, respectively (Table 1).
MAG TCMF9 showed 93% 16S rRNA gene similarity to a gammaproteobacterial endosymbiont clone of the gutless oligochaete worm, *Olavius ilvae*. *O. ilvae* is known to harbor sulfur-oxidizing gammaproteobacterial endosymbionts that fix CO$_2$ via the CBB cycle (Ruehland et al., 2008). This MAG contained most of the genes involved in the CBB cycle, including RuBisCo ($rbcSL$) and phosphoribulokinase ($prk$). A whole suite of genes required for thiosulfate and sulfide oxidation such as the $sox$ system (except $soxC$), $sqr$ and $fccAB$ were present. This MAG also contained all three genes ($dsrAB$, $aprAB$, $sat$) involved in the reverse $dsr$ pathway, indicating versatility for the oxidation of sulfur compounds. Three out of the four genes encoding the aa3-type cytochrome oxidase ($cox$) and all genes encoding the microaerobic $cbb3$-type cytochrome oxidase ($cco$) were present (Table 1). The majority of the genes involved in bacterial chemotaxis and flagellar motility were also present (data not shown).

Since the 16S rRNA gene was not recovered in MAG TCMF1, other single-copy ribosomal proteins were used to constrain its lineage. Both the 30S ribosomal protein S8 and S9 were greater than 90% identical to a *Sulfurovum* species. Additionally, the average amino acid identity (AAI) shared between TCMF1 and *Sulfurovum* strain NBC37 was 58%, which is well within the range that defines the genus boundary (Rodriguez-R and Konstantinidis, 2014). Based on this data, TCMF1 belonged to the *Sulfurovum* lineage within the *Epsilonproteobacteria*, which dominated the YF community (Fig. 1b). *Sulfurovum* spp. are sulfur/hydrogen oxidizers that fix CO$_2$ using the rTCA cycle (Inagaki, 2004; Mino et al., 2014; Giovannelli et al., 2016). Out of the three diagnostic genes for the rTCA cycle, $kor$ and $frd$ were present. *Sox* and *sqr* genes were also present, while hydrogenases were absent. Interestingly, this MAG contained the $nifDHK$ gene essential for nitrogen fixation, while the periplasmic nitrate reductase encoding gene, *napAB*, which is widespread in the *Epsilonproteobacteria* (Vetriani et al., 2014), was missing. Three out of the four genes encoding the aa3-type cytochrome oxidase ($cox$) and all genes encoding the micro-aerobic $cbb3$-type ($cco$) and bd cytochrome oxidase ($cyd$) were present, indicating the ability to respire oxygen within a broad range of concentrations. The $luxS$ gene, involved in quorum sensing and conserved across all members of the *Epsilonproteobacteria* (Pérez-Rodríguez et al., 2015), was also present in TCMF1.

Given the presence of genes coding for *sqr* both in the metagenomes and in the recovered MAGs, we carried out phylogenetic analyses of representative SQR enzymes from the Tor Caldara biofilms and close relatives obtained from GenBank. This analysis placed the sequences
into three discrete clusters separated from the *Chlorobium* lineage, from which SQR was originally characterized: one cluster was related to *Sulfurovum* spp., one was related to the *Thiotricales* (*Thiomicrospira* and *Thiothrix* spp) and a third cluster included two groups of sequences: one related to *Sulfurovum* spp. and one related to the *Thiotricales* (Fig. 5). In line with the taxonomic annotation derived from other genes, the SQR from MAG TCMF1 was related to *Sulfurovum* sequences, while the enzymes from MAGs TCMF2 and TCMF9 were related to *Thiotricales*-derived sequences (Fig. 5).

**Discussion**

Chemosynthetic microbial biofilm are commonly found in geothermal and sulfidic environments (Sievert and Vetriani, 2012), including cold seeps (Ristova et al., 2015), sulfidic caves (Macalady et al., 2006), thermal springs (Beam et al., 2016), mud volcanoes (Heijs et al., 2005) as well as deep-sea (Gulmann et al., 2015, O’Brien et al., 2015) and shallow-water hydrothermal vents (Miranda et al., 2016; Patwardhan et al., 2018). Members of *Gammaproteobacteria* and *Epsilonproteobacteria* dominate most of these biofilms and a spatial segregation between the two classes of bacteria has been correlated with the *in situ* sulfide concentration (Engel et al., 2004; Crépeau et al., 2011; Gulmann et al., 2015; O’Brien et al., 2015; Miranda et al., 2016; Macalady et al., 2008; Meier et al., 2017).

At Tor Caldara, a submarine gas vent in the Tyrrhenian Sea characterized by vigorous venting of CO$_2$ and unusually a high concentration of H$_2$S (avg. of 23.13 mol%; Patwardhan et al., 2018), white filamentous microbial communities grow profusely on the rocks near the gas venting. To investigate the composition and function of established and young biofilms, we collected native substrates (EF) as well as glass slides (YF) deployed in the vicinity of a vent. Within 4 days, filamentous biofilms colonized the slides.

Our survey of 16S rRNA genes and transcripts revealed a shift in the two main taxonomic groups during colonization, whereby sulfur-oxidizing *Epsilonproteobacteria* dominated the YF biofilms, while *Gammaproteobacteria* became prevalent in the EF community. Further, we showed that representative species of *Epsilon*- and *Gammaproteobacteria* are adapted to different sulfide concentrations (Patwardhan et al., 2018). In this study, we integrated metagenomic and metaproteomic approaches to investigate the *in situ* expression of central metabolic pathways related to carbon fixation and energy conservation in the YF and EF biofilms. In line with our
previous 16S rRNA amplicon survey, the metagenome-extracted 16S rRNA genes showed that *Epsilonproteobacteria* and *Gammaproteobacteria* dominated the YF and EF biofilms, respectively (Figs. 1a and b).

**The relative expression of the rTCA and CBB cycles for carbon fixation reflects the taxonomic composition of the YF and EF biofilms.** Chemoautotrophs are the primary producers at hydrothermal vents, forming the basis of the food web by fixing inorganic CO$_2$ into organic material (Jannasch and Wirsen, 1979; Karl *et al.*, 1980, Sievert and Vetriani, 2012). Till date, six pathways for carbon fixation are known: the Calvin-Benson-Bassham (CBB) reductive pentose phosphate cycle, the reductive tricarboxylic acid cycle (rTCA), the reductive acetyl coenzyme A pathway(Wood-Ljungdahl), the 3-hydroxypropionate bi-cycle, the 3-hydroxypropionate/4-hydroxybutyrate cycle, and the dicarboxylate/4-hydroxybutyrate cycle (Nakagawa and Takai, 2008; Berg, 2011). The rTCA and CBB cycles are most prevalent at diffuse-flow hydrothermal vents and, in these habitats, they are generally diagnostic of *Epsilon-* and *Gammaproteobacteria*, respectively (Nakagawa and Takai, 2008). Enrichment of proteins of the rTCA cycle affiliated with *Epsilonproteobacteria*, in addition to a similar signature at the gene level, confirmed that autotrophy using the rTCA cycle in *Epsilonproteobacteria* was the predominant mechanism for carbon fixation in the YF communities (Figs. 2, 6 and Suppl. Tables 1, 2). The majority of the rTCA genes and proteins were classified as *Sulfurovum* spp. (Fig. 2), which is known to be ubiquitous in vent environments (Inagaki, 2004; Dahle *et al.*, 2013; Mino *et al.*, 2014; Giovannelli *et al.*, 2016). These findings underline the importance of *Sulfurovum* spp. in the YF community and are in accordance with data from the 16S rRNA diversity analysis (Patwardhan *et al.*, 2018; Fig. 1). While the EF community also encoded genes and expressed enzymes of the rTCA cycle, the CBB cycle was overrepresented in the EF biofilm at both the gene and protein levels, with contributions from various genera within the *Gammaproteobacteria*, in particular *Thiomicrospira, Thioalkovibrio* and *Thiothrix*-related spp. (Figs. 2, 6 and Suppl. Tables 1, 2).

This is also in line with our 16S rRNA survey, which showed these genera as dominant in the EF biofilm (Patwardhan *et al.*, 2018; Fig. 1). Thus, we observed a shift from rTCA-based carbon fixation (*Epsilonproteobacteria*) to CBB-based carbon fixation (*Gammaproteobacteria*) during biofilm maturation. The lower energy demand of the rTCA cycle, compared to the CBB cycle
(Nakagawa and Takai, 2008; Hügler and Sievert, 2011), might provide an advantage to the
*Epsilonproteobacteria* during early substrate colonization.

The expression of different pathways for sulfide/sulfur oxidation in the YF and EF biofilms
may reflect adaptations to fluctuating redox regimes. In shallow-water geothermal systems,
the concentration of reduced sulfur compounds and other electron donors for microbial
oxidations, along with the availability of oxidants, vary considerably over time (Yücel et al.,
2013; Price and Giovannelli, 2017). Hence, chemolithoautotrophic microorganisms have adapted
to such fluctuating redox regimes. In particular, reduced sulfur compounds represent an
important energy source that drives energy-yielding reactions in geothermal environments
(Nakagawa and Takai, 2008). The abundance and range of various genes and proteins involved
in sulfur oxidation pathways presented here validated our initial findings about the importance of
sulfur-oxidizing bacteria at Tor Caldara (Patwardhan et al., 2018). For instance, the Sox pathway
for sulfur oxidation was one of the most abundant energy-yielding pathway recovered both in the
metagenomes and metaproteomes (Figs. 3, 4, 6 and Suppl. Tables 1, 2).

Thiosulfate oxidation catalyzed by the multienzyme complex Sox is widespread in sulfur-
oxidizing chemoautotrophs (Friedrich et al., 2001, 2005; Yamamoto and Takai, 2011) . This
enzyme complex, consisting of four protein components, soxYZ, soxXA, soxB and soxCD,
catalyzes the oxidation of thiosulfate and sulfide to sulfate. The abundance of the Sox complex in
both filament communities at the gene as well as the protein (YF:1.8%; EF:1.6%) level indicates
that this is the predominant pathway used by the biofilm bacteria to conserve energy (Figs. 3a, 4a
and Suppl. Tables 1, 2). This is consistent with a metaproteogenomic study of chimney-
associated microbial communities from deep-sea hydrothermal vents, which revealed that Sox
represented one of the most highly expressed energy-yielding pathways (Pjevac et al., 2018).
However, the taxonomic affiliation of the sox reads (Fig. 3b) and Sox enzymes (Fig. 4b) from the
Tor Caldara biofilms shows a finer resolution, indicating that *Epsilonproteobacteria* are mostly
expressing the Sox complex in the YF biofilm, while *Gammaproteobacteria* take over that same
role in the EF biofilms (Fig. 6).

A similar trend is evident in the profile of the sulfide:quinone oxidoreductase enzyme complex
(SQR) involved in oxidation of sulfide to elemental sulfur, which is predominantly expressed by
*Epsilonproteobacteria* in the YF filaments, while its contribution by *Gammaproteobacteria*
increases in the EF filament community (Fig. 4b and 6). In addition to SQR, in some
chemolithotrophic bacteria oxidation of hydrogen sulfide to elemental sulfur is catalyzed by the
flavocytochrome c-sulfide dehydrogenase (FccAB; Visser et al., 1997; Mußmann et al., 2007).
The metaproteomic data presented here show that the FccAB enzyme is affiliated only with
\textit{Gammaproteobacteria}, and that its expression is higher in the EF biofilm than in the YF
community (Figs. 4a and 6). The bioenergetics of microbial sulfide oxidation might provide a
framework to interpret the SQR and Fcc expression profiles in the Tor Caldara biofilms. Since
the midpoint potential of the $\text{NAD}^+/\text{NADH}$ couple at pH 7 is approximately 50 mV more
negative than the midpoint potential of the $\text{S}^0/\text{H}_2\text{S}$ couple, chemolithotrophic SOBs require
energy to transport electrons from hydrogen sulfide upwards to $\text{NAD}^+$ by reverse electron flow.
In SQR-catalyzed sulfide oxidation, the electrons from sulfide enter the transport chain at the
level of quinones, generating the necessary electrochemical proton potential across the
membrane for reverse electron transfer (Griesbeck et al. 2000). However, when Fcc is involved,
the electrons enter the transport chain at the level of $c$-type cytochromes, whose midpoint
potential is more electropositive than that of quinones. That increases the energetic burden for
the reverse transfer of electrons upwards to $\text{NAD}^+$ and implies that, in chemolithotrophic SOBs,
sulfide oxidation by SQR provides more energy than sulfide oxidation by Fcc (Griesbeck et al.
2000). However, due to its high affinity for sulfide (Brune, 1995), Fcc is hypothesized to be
expressed at low sulfide concentrations, supplementing the energetically more efficient SQR.
Previous work on the Tor Caldara biofilms demonstrated that \textit{Epsilonproteobacteria} are adapted
to higher sulfide concentrations than \textit{Gammaproteobacteria} and that they are the pioneer
colonizers at this site (Patwardhan \textit{et al.}, 2018). We hypothesize that, early in the colonization
process (YF biofilm), \textit{Epsilonproteobacteria} might be driving down sulfide levels within the
biofilms and that the increased expression of the high affinity FccAB in EF might provide an
advantage to the \textit{Gammaproteobacteria} in the subsequent stages of colonization (EF biofilm;
Fig. 6). The hypothesis that Fcc is more prevalent under low sulfide and more oxidized
conditions (Griesbeck \textit{et al.}; Brune, 1995) supports our model. Overall, the co-expression of the
energy efficient SQR and the high affinity Fcc may provide metabolic flexibility in response to
fluctuating sulfide concentrations within the biofilm communities.
The reverse dissimilatory sulfate reductase pathway (rDSR) oxidizes stored intracellular sulfur to
sulfate via sulfite (Friedrich \textit{et al.}, 2005). Consistent with findings in other filamentous
Gammaproteobacteria (Mußmann et al., 2007; Chernousova et al., 2009; Sharrar et al., 2017), all three proteins involved in this pathway, namely, DsrAB, AprAB and SAT, were affiliated with Gammaproteobacteria and were enriched in the EF biofilm (Figs. 4, 6 and Suppl. Table 2). The expression of the rDSR pathway is in accordance with the stored intracellular elemental sulfur observed in electron micrographs of EF (Patwardhan et al., 2018), while its higher expression in the EF biofilm suggests that this community may experience transient depletions of hydrogen sulfide. This stored sulfur might also be disproportionated into sulfite, thiosulfate and sulfide as evidenced by the expression of gammaproteobacterial Sor (Fig. 4) in the EF biofilm, and might provide a competitive advantage to the secondary colonizers (Kletzin, 1989; Veith et al., 2012). Since the Dsr pathway for S-oxidation is not encoded in the currently available genomes of Epsilonproteobacteria (Yamamoto and Takai, 2011), the epsiloproteobacterial SAT found in the metaproteome is probably involved in sulfate assimilation.

Sulfur oxidation was complemented with sulfur reduction evidenced by high expression of the a putative Sulfovum-affiliated NADH-dependent sulfur reductase (NSR; Schut et al., 2007) in the YF (Figs. 4, 6 and Suppl. Table 2). Despite sulfur respiration is known to be coupled to hydrogen oxidation in two Sulfovum spp. (Yamamoto Masahiro et al., 2010; Mino et al., 2014), we did not detect hydrogenases in the metaproteomes. Since hydrogen concentrations are low at Tor Caldara, the Sulfovum spp. found there may be adapted to oxidize alternative energy sources, such as malate and/or formate (Campbell et al., 2006; Yamamoto Masahiro et al., 2010; Keller et al., 2015).

Epsilonproteobacteria-mediated nitrogen fixation and Gammaproteobacteria-mediated nitrate reduction are prevalent in the EF biofilm. Microbial nitrate respiration occurs via two main pathways, dissimilatory nitrate reduction to ammonia (DNRA) and denitrification. Both processes have been investigated in marine geothermal environments (Bourbonnais et al., 2012; Bourbonnais et al., 2012; Pérez-Rodríguez et al., 2013; Vetriani et al., 2014). The first step in nitrate reduction is carried out by the enzyme nitrate reductase. Currently, two different types of respiratory nitrate reductases, the membrane-bound Nar and the periplasmic Nap, are known to occur in bacteria (Richardson et al., 2001). Nar is a low-affinity enzyme expressed under nitrate-rich conditions, whereas Nap is a high-affinity enzyme whose expression is favored under low nitrate conditions (Potter et al., 1999; Mintmier et al., 2018). The genes encoding both Nar and...
Nap were present in the EF and YF communities, as well as the genes encoding various nitrite reductases (NirBD, NirK, NirS and NrfA), the nitric oxide reductase (NorB) and the nitrous oxide reductase (NosZ; (Fig. 3a and Suppl. Table 1). However, metaproteomic analyses showed that only the Gammaproteobacteria-affiliated Nar and Nir enzymes were expressed, while Nap and the enzymes involved in the denitrification pathway could not be detected (Figs. 4b). The coastal area at Tor Caldara is likely enriched in nitrate (possibly from runoff of lawn fertilizers), which might explain the expression of NarGHI, but not of NapAB (Fig 4a and Suppl. Table 2). The NirBD nitrite reductase, often involved in the reduction of nitrite to ammonia in DNRA (Reyes et al., 2017), was also found to be abundant. Both the Nar and Nir enzymes were enriched in the EF biofilm, indicating that nitrate respiration is prevalent in the established community.

Apart from nitrate respiration, nitrogen fixation also plays an important role in nitrogen cycling at vents (Rau, 1981; Mehta et al., 2003). Expression of epsilonproteobacterial NifDHK was observed in both filament communities but was highly enriched in the EF (Figs. 4, 6 and Suppl. Table 2). While nitrogen fixation is not common in Epsilonproteobacteria, recent studies have shown that some members of this class have the genomic potential for it (Keller et al., 2015; Waite et al., 2017), and the ability to fix nitrogen was demonstrated experimentally in Lebetimonas spp. (Meyer and Huber, 2013). Nitrogen fixation is an energy expensive reaction and nitrogenases are usually expressed in nitrogen limiting conditions (Olivares et al., 2013). We hypothesize that, when the filamentous community transitions from a young to an established state, nitrate respiration carried out by Gammaproteobacteria might result in a depletion in nitrogen within the biofilm microenvironment. Hence, the ability to fix nitrogen might give the Epsilonproteobacteria an advantage while competing for a fixed nitrogen source against the secondary colonizers in the established biofilm.

Enzymes for aerobic and microaerobic respiration are differentially expressed in the YF and EF biofilms. Most of the well-characterized bacterial oxidases belong to the heme-copper superoxidase family. There are two types of oxidases that use cytochrome c as a substrate: cytochrome c oxidase aa3-type and cytochrome c oxidase cbb3-type. The former is expressed in aerobic (high oxygen) conditions while the latter is expressed in microaerobic (low oxygen) conditions. Additionally, cytochrome bd, a ubiquinol oxidase, is also expressed in microaerobic
conditions (Garcia-Horsman et al., 1994; Pitcher and Watmough, 2004). At Tor Caldara, the microaerobic, high affinity cbb3-type and bd-type cytochrome oxidase were enriched in EF biofilm and were mostly classified as *Gammaproteobacteria* (Figs. 4, 6 and Suppl. Table 2), suggesting that the biofilms experience microaerobic/anoxic conditions as they mature. On the other hand, primary colonizers in the YF might experience oxic to micro-aerobic conditions as suggested by equal abundance of aerobic aa3 type and micro-aerobic cbb3 type cytochrome c oxidase classified dominantly as *Epsilon proteobacteria*.

**Overall view: A metaproteome-based model of microbial colonization at the Tor Caldara gas vents.** Based on the metaproteomic data, we reconstructed the central metabolic pathways expressed by the Tor Caldara young and established biofilm communities (Fig. 6). According to this model, substrates exposed to the gas emissions are initially colonized by a population dominated by sulfide-tolerant *Epsilon proteobacteria* (Fig. 1; Patwardhan et al., 2018). Members of this class oxidize various reduced sulfur compounds via the Sox and SQR pathways coupled to oxygen reduction (both aerobic and micro-aerobic), while fixation of carbon dioxide occurs via the rTCA cycle. In the early stages of colonization, *Gammaproteobacteria* are present, but underrepresented (Figs. 1 and 6). Over time, sulfide oxidation by *Epsilon proteobacteria* may lower the sulfide concentration within the biofilm community, conditioning the environment for the growth of the less sulfide-tolerant *Gammaproteobacteria* (Patwardhan et al., 2018), which dominate the established filaments. *Gammaproteobacteria* also oxidize various sulfur compounds using the Sox and SQR pathways, but this reaction is now coupled to either microaerobic and/or anaerobic respiration using oxygen and nitrate as terminal electron acceptors, respectively. The high affinity, periplasmic FccAB enzyme is co-expressed with SQR in response to fluctuating concentrations of sulfide, while elemental sulfur is stored in the cytoplasm. Energy thus conserved is used to fix carbon dioxide via the CBB cycle (Fig. 6). When sulfide availability becomes limiting within the biofilm community, certain *Gammaproteobacteria* further oxidize the stored sulfur to sulfate via the reverse dissimilatory sulfate reductase pathway, thus making full use of their metabolic repertoire. As the filamentous community transitions from a young to a more established stage, anaerobic respiration by the *Gammaproteobacteria* may lead to loss of nitrate from the system. This triggers increased
nitrogen fixation by the *Epsilonproteobacteria* as a potential strategy to compete with the *Gammaproteobacteria*.

**Conclusion**

Our study revealed a complex pattern of protein expression in chemosynthetic biofilm communities that colonize the gas vent system at Tor Caldara. Functions common to chemosynthetic microbial communities, such as carbon fixation and sulfide oxidation, are catalyzed via the expression of different enzymatic complexes encoded by two predominant groups of SOBs belonging to the *Epsilon* - and *Gammaproteobacteria*. We hypothesize that the expression of these enzymatic complexes is finely tuned to the chemical and physical conditions within the biofilm communities, and reflect the physiological characteristics of the two groups of SOBs. Overall, the enzyme expression profiles obtained in this work reveal metabolic flexibility and adaptations to fluctuating redox regimes during colonization of the shallow-water gas vents of Tor Caldara. These results are consistent with the previously reported transition from *Epsilonproteobacteria* - the pioneer colonists - to *Gammaproteobacteria* at Tor Caldara.

**Acknowledgements**

We thank the crew of the R/V *Antares* for their support during the sampling operations, and Satyajit Rajapurkar for help and guidance with the bioinformatics analyses. This work was partially supported by the following grants to CV: NSF OCE 11-24141, NSF OCE 11-36451, NSF MCB 15-17567 and NASA NNX15AM18G. This paper is C-DEBI contribution xx.

The authors declare that they have no conflict of interest.
References


DOI=10.3389/fmicb.2016.00025


Figure Legends

**Fig. 1.** Class (a) and genus (b) level distribution of 16S rRNA gene and transcript sequences recovered from established and young filaments.

**Fig. 2.** Profiles of genes and proteins for carbon fixation in established and young filamentous biofilms. (a) Transcripts per million (TPM; columns 1 and 3) and taxonomic affiliation of reads annotated to genes (columns 2 and 4); (b) Percentage (columns 1 and 3) and taxonomic affiliation of reads annotated to proteins (columns 2 and 4).

**Fig. 3.** (a) Transcripts per million (TPM) and (b) taxonomic affiliation of reads annotated to key genes involved in nitrogen, sulfur, oxygen and heavy metal detoxification pathways for established and young filaments.

**Fig. 4.** (a) Percentage and (b) taxonomic affiliation of key proteins involved in nitrogen, sulfur, oxygen and heavy metal detoxification pathways for established and young filaments.

**Fig. 5.** Maximum likelihood phylogenetic tree showing the position of the sulfide quinone oxireductase amino acid sequences, SQR, obtained in this study (indicated in boldface). Closely related sequences were obtained from GenBank. Approximate likelihood ratio test (aLRT) values for branch support are indicated. Bar, 1% estimated substitutions.

**Fig. 6.** Metaproteomic-based reconstruction of the central metabolic pathways expressed by the Tor Caldara young and established biofilm communities. The size of the enzymatic complexes is proportional to their abundance in the metaproteome. The numbers indicate the normalized protein abundance (%). The colors indicate the taxonomic affiliation of the enzymes and reflect the relative proportions of the two predominant bacterial classes. Blue: *Epsilonproteobacteria*; Pink: *Gammaproteobacteria*. 


Table 1. Genes involved in main metabolic pathways recovered from MAGs

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Gene</th>
<th>TCMF1</th>
<th>TCMF2</th>
<th>TCMF9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><em>Sulfurovum</em></td>
<td><em>Thiotrix</em>-related</td>
<td>Gammaproteobacteria</td>
</tr>
<tr>
<td>16S rRNA</td>
<td></td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Calvin cycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>rbc</em></td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>prk</em></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reductive TCA cycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>kor</em></td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td><em>frd</em></td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sulfur oxidation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>sox</em></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>sqr</em></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>fcc</em></td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>dsr</em></td>
<td>-</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>apr</em></td>
<td>-</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>sat</em></td>
<td>-</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>Oxygen respiration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>cox</em></td>
<td>Yes</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>cco</em></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td><em>cyd</em></td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>nir</em></td>
<td>-</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td><em>nif</em></td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Quorum sensing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>lux</em></td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 1

(a) Abundance (%) of established and young filaments
- Alphaproteobacteria
- Bacteroidia
- Epsilonproteobacteria
- Gammaproteobacteria
- Others

(b) Abundance (%) of established and young filaments
- Genus:
  - Others
  - Sulfurovum
  - Thiomicrothrix
  - Thiobacillus
  - Uncultured
  - Uncultured Thiothrix

Figure 2
Figure 3

Pathway
- Heavy Metal
- Nitrogen
- Oxygen
- Sulfur

TPM
- 500
- 1000
- 1500
- 2000
- 2500

Class
- Other
- Azinobacteria
- Alphaproteobacteria
- Aquificae
- Bacteroidia
- Betaproteobacteria
- Cytophaga
- Deltaproteobacteria
- Epsilonproteobacteria
- Flavobacteria
- Gammaproteobacteria
- Saprospir
Figure 4
Figure 6
Supplementary Table 1. Metagenomes: Normalized abundances of key genes involved in carbon fixation, nitrogen & sulfur metabolism, oxygen respiration, and heavy metal detoxification pathways expressed as transcripts (reads) per million (TPM).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Young Filaments (YF)</th>
<th>Established Filaments (EF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP-citrate lyase (aclAB)</td>
<td>681.5</td>
<td>210.34</td>
</tr>
<tr>
<td>Fumarate reductase (frdAB)</td>
<td>1331.01</td>
<td>115.24</td>
</tr>
<tr>
<td>Pyruvate ferredoxin oxidoreductase (porABDG)</td>
<td>1141.07</td>
<td>550.12</td>
</tr>
<tr>
<td>Phosphoribulokinase (prk)</td>
<td>99.47</td>
<td>207.22</td>
</tr>
<tr>
<td>Ribulose-bisphosphate carboxylase (rbcSL)</td>
<td>182.77</td>
<td>408.48</td>
</tr>
<tr>
<td>Nitrate reductase (napAB)</td>
<td>381.01</td>
<td>153.62</td>
</tr>
<tr>
<td>Nitrate reductase (narGHI)</td>
<td>21.55</td>
<td>86.98</td>
</tr>
<tr>
<td>Nitric oxide reductase (norB)</td>
<td>1.35</td>
<td>3.72</td>
</tr>
<tr>
<td>Nitrite reductase (nirBD)</td>
<td>312.33</td>
<td>498.28</td>
</tr>
<tr>
<td>Nitrite reductase, cytochrome c-552 (nrfA)</td>
<td>0.37</td>
<td>4.92</td>
</tr>
<tr>
<td>Nitrite reductase, NO-forming (nirK)</td>
<td>0.77</td>
<td>7.69</td>
</tr>
<tr>
<td>Nitrite reductase, NO-forming (nirS)</td>
<td>2.23</td>
<td>4.81</td>
</tr>
<tr>
<td>Nitrogenase (nifDHK)</td>
<td>499.47</td>
<td>325.34</td>
</tr>
<tr>
<td>Nitrous-oxide reductase (nosZ)</td>
<td>6.71</td>
<td>12.34</td>
</tr>
<tr>
<td>Adenylylsulfate reductase (aprAB)</td>
<td>17.5</td>
<td>26.49</td>
</tr>
<tr>
<td>Polysulfide reductase (psrA)</td>
<td>2.61</td>
<td>11.83</td>
</tr>
<tr>
<td>Sox system (soxXYZABCD)</td>
<td>2553.8</td>
<td>2308.16</td>
</tr>
<tr>
<td>Sulfate adenylytransferase (sat)</td>
<td>237.44</td>
<td>255.85</td>
</tr>
<tr>
<td>Sulfide dehydrogenase (fccAB)</td>
<td>635.6</td>
<td>1402.89</td>
</tr>
<tr>
<td>Sulfide:quinone oxidoreductase (sqr)</td>
<td>375.84</td>
<td>843.84</td>
</tr>
<tr>
<td>Sulfide dehydrogenase</td>
<td>52.65</td>
<td>2.22</td>
</tr>
<tr>
<td>Sulfito reductase (dsrAB)</td>
<td>76.72</td>
<td>210.2</td>
</tr>
<tr>
<td>Sulfur oxygenase/reductase (sor)</td>
<td>0</td>
<td>38.43</td>
</tr>
<tr>
<td>Putative Sulfur reductase (nsr)</td>
<td>323.84</td>
<td>394.22</td>
</tr>
<tr>
<td>Cytochrome c oxidase, aa3-type (coxABCD)</td>
<td>834.02</td>
<td>665.43</td>
</tr>
<tr>
<td>Cytochrome c oxidase, cbb3-type (ccoNOPQ)</td>
<td>944.63</td>
<td>1026.12</td>
</tr>
<tr>
<td>Cytochrome bd ubiquinol oxidase (cydAB)</td>
<td>284.87</td>
<td>392.7</td>
</tr>
<tr>
<td>Arsenate reductase (arsC)</td>
<td>527.49</td>
<td>543.82</td>
</tr>
<tr>
<td>Mercuric reductase (merA)</td>
<td>502.49</td>
<td>174.16</td>
</tr>
<tr>
<td>Selenate detoxification (dedA)</td>
<td>0.69</td>
<td>15.14</td>
</tr>
</tbody>
</table>
Supplementary Table 2. Metaproteomes: Normalized abundances of key proteins involved in carbon fixation, nitrogen & sulfur metabolism, oxygen respiration, and heavy metal detoxification pathways expressed as a percentage of total proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Young Filaments (YF)</th>
<th>Established Filaments (EF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP-citrate lyase (AclAB)</td>
<td>1.202</td>
<td>0.545</td>
</tr>
<tr>
<td>Fumarate reductase (FrdAB)</td>
<td>0.008</td>
<td>0.003</td>
</tr>
<tr>
<td>Pyruvate ferredoxin oxidoreductase (PorABDG)</td>
<td>3.825</td>
<td>1.677</td>
</tr>
<tr>
<td>Phosphoribulokinase (PRK)</td>
<td>0.081</td>
<td>0.267</td>
</tr>
<tr>
<td>Ribulose-bisphosphate carboxylase (RbcSL)</td>
<td>0.623</td>
<td>0.846</td>
</tr>
<tr>
<td>Nitrate reductase (NarGHI)</td>
<td>0.014</td>
<td>0.025</td>
</tr>
<tr>
<td>Nitrite reductase (NirBD)</td>
<td>0.103</td>
<td>0.329</td>
</tr>
<tr>
<td>Nitrogenase (NifDHK)</td>
<td>0.002</td>
<td>0.354</td>
</tr>
<tr>
<td>Adenylylsulfate reductase (AprAB)</td>
<td>0.006</td>
<td>0.015</td>
</tr>
<tr>
<td>Sox system (SoxXYZABCD)</td>
<td>1.85</td>
<td>1.635</td>
</tr>
<tr>
<td>Sulfate adenylyltransferase (SAT)</td>
<td>0.085</td>
<td>0.18</td>
</tr>
<tr>
<td>Sulfide dehydrogenase (FccAB)</td>
<td>0.351</td>
<td>0.72</td>
</tr>
<tr>
<td>Sulfide:quinone oxidoreductase (Sqr)</td>
<td>0.336</td>
<td>0.42</td>
</tr>
<tr>
<td>Sulfite reductase (DsrAB)</td>
<td>0.399</td>
<td>0.639</td>
</tr>
<tr>
<td>Sulfur oxygenase/reductase (Sor)</td>
<td>0.013</td>
<td>0.06</td>
</tr>
<tr>
<td>Putative Sulfur reductase (Nsr)</td>
<td>1.799</td>
<td>0.301</td>
</tr>
<tr>
<td>Cytochrome c oxidase, aa3-type (CoxABCD)</td>
<td>0.397</td>
<td>0.083</td>
</tr>
<tr>
<td>Cytochrome c oxidase, cbb3-type (CcoNOPQ)</td>
<td>0.357</td>
<td>0.469</td>
</tr>
<tr>
<td>Cytochrome bd ubiquinol oxidase (CydAB)</td>
<td>0.014</td>
<td>0.073</td>
</tr>
<tr>
<td>Arsenate reductase (ArsC)</td>
<td>0</td>
<td>0.003</td>
</tr>
<tr>
<td>Selenate detoxification (DedA)</td>
<td>0.008</td>
<td>0.004</td>
</tr>
</tbody>
</table>