Kill•Spill

Integrated Biotechnological Solutions for Combating Marine Oil Spills

Deliverable D2.5

Community metabolic network analysis performed on reference samples

This project is supported by the European Union under the Food, Agriculture and Fisheries and Biotechnology theme of the 7th Framework Programme for Research and Technological Development under GA no. 312139

Work package	WP2 Development of biosensors and in-situ monitoring tools to determine biodegradation efficiency
Deliverable no	D2.5
Deliverable title	Report on community metabolic network analysis performed on reference samples
Due date:	M30 (2015-06-30)
Actual submission date:	
Start date of project:	2013-01-01
Deliverable Lead Beneficiary (Organisation name)	IAMC-CNR
Participant(s) (Partner short names)	Bangor, IAMC-CNR, ICP-CSIC
Author(s) in alphabetic order:	Simone Cappello, Renata Denaro, Manuel Ferrer, Maria Genovese, Peter Golyshin, Michail M. Yakimov
Contact for queries:	Michail M. Yakimov Sp.ta S. Raineri 86 98122 Messina T +39 0906015437 E michail.yakimov@iamc.cnr.it
Dissemination Level: (PUblic, Restricted to other Programmes Participants, REstricted to a group specified by the consortium, COnfidential only for members of the consortium)	PU
Deliverable Status:	V1 2015-07-15 sent for QA Final Version 2016-01-28

Table of Content

1 Introduction to deliverable D2.5 1
2 Materials and Methods 1
2.1 Study locations 1
2.2 Sampling and Characterization of samples 1
2.3 Protein extraction 2
2.4 Mass spectrometry and data analysis 2
2.5 Metabolomic fingerprint analysis of sediment samples 3
3 Results 4
3.1 Study sites: physical-chemical characteristics 4
3.2 Metaproteomic analysis: general comments 4
3.3 Identities of expressed proteins 5
3.4 Differences at the metabolic and organismal levels as revealed through metaproteomics 6
3.5 Community metabolic activity associated with biodegradation 10
4 Discussion and conclusion 10

List of Figures

Figure 1 Relative abundances and distributions of archaeal (A) and bacterial (B, C) proteins in the metaproteomes of marine sediment samples based on taxonomic bins for proteome-derived proteins that could be assigned a taxonomic annotation. 5
Figure 2 Most prevalent pathways of metabolic network revealed in the analysed marine sediment samples (A-C) and a protein expression patterns that correlate with metabolic coupling of CH_4 and CH_3OH metabolism as well as the Wood-Ljungdahl pathway (D) 8
Figure 3 Sulfur metabolism patterns in microbial communities inhabiting petroleum-polluted marine sediments based on proteome analysis. 9

List of Tables

Table 1 Overall physical-chemical characteristics of the investigated sediment samples 4

1 Introduction to deliverable D2.5

The aim of this deliverable was to analyse the structure and the function of natural microbial populations present in three reference samples developed and studied during Kill•Spill: i) tar collected near the sunken tanker *Haven* at the Gulf of Genoa (code sample MPS-HAV); ii) chronically polluted sediments collected from the harbour of Messina (MPS-MES) and iii) chronically polluted sediments collected from the harbour of Priolo Gargallo (code sample MPS-PRI).

Data presented here are of particular importance as the perspective for designing efficient and cost-effective tools and methods to allow the *in-situ* monitoring of hydrocarbon pollutants during and after the event of pollution. The ultimate goals of D2.5 are the analysis of microbial metabolic networks, and determination of the activity of the petroleum-degrading bacterial fraction of microbial populations. These can give important information to evaluate the efficiency of decontamination treatments, and allow the biodegradation potential of the microbial communities present at the contaminated site to be monitored. We foresee that in the future this approach would provide the authorities with a toolbox of methods for documenting biodegradation in oil spills.

To achieve the goals of D2.5, IAMC-CNR of Messina was involved together with Bangor University and ICP-CSIC partners in the sampling of environmental samples, the characterization of the main chemical-physical parameters and chemical characterization of pollutant hydrocarbons. Moreover, all these partners were involved in metabolomic and proteomic analysis of the samples used in the study.

2 Materials and Methods

2.1 Study locations

Three areas, located in the Ligurian and Ionian Seas along the Northern and Southern Italian coast, were investigated in this activity. In particular

(1) the harbour of Priolo Gargallo (Siracusa, Italy; 37°10'27.462"N, 15°12'7.505"E), which is characterized by heavy industrialization and heavy tanker traffic charging-discharging both crude and refined oil;
(2) the harbour of Messina -MES (Sicily, Italy; 38°11'42.267"N, 15°34'25.014"E), which is exposed to chronic petroleum pollution because of intensive maritime traffic and a limited hydrodynamic regimen; and
(3) the Gulf of Genoa in the northernmost part of the Ligurian Sea (Genoa, Italy; 44° 22'25.75N, 8° 41'59.58E), where the M/T *Haven* tanker (*HAV*) has been incidentally sunk in 1991 and released into the sea of about 40,000 tons of crude oil.

The samples were named as follows, based on the site of sampling adding the code 'MPS', which refers to MetaProteome Source: MPS-HAV (*Haven* tanker at the Gulf of Genoa); MPS-MES (the harbour of Messina); MPS-PRI (the harbour of Priolo Gargallo).

2.2 Sampling and Characterization of samples

Sediment samples were collected at water depths of 1, 6 and 78 m in MES, PRI and HAV, respectively. Conductivity, temperature (T), redox potential (Eh), pH, and dissolved oxygen (DO) concentration measurements were measured *in situ* using a Multi 340i multi-parameter meter (WTW, Germany) at the sediment surface immediately after sampling. Conductivity calibration was carried out with a control solution of 0.01 M KCl. Microelements (metal) concentrations were determined through

Inductively Coupled Plasma-Mass Spectrometry (ICP/MS). Ammonium was determined using the indophenol blue technique. The dissolved organic carbon (DOC) content was determined by the dichromate wet oxidation method; total organic matter content was calculated by multiplying the values of the organic carbon by 1.8. Hydrocarbon amounts were chromatographically determined using a Master GC DANI Instruments (Development Analytical Instruments), equipped with a split-splitless (SSL) injector and FID detector.

2.3 Protein extraction

Sediment samples were subjected to protein isolation in a two-step protocol. First, microbial cells were separated from the sediment matrix and after re-suspension, the samples were kept in an ice water bath and sonicated using a water bath ultra-sonicator (60 W output) for 30 min at 4ºC.The samples were then centrifuged to remove the sediment material. The resulting supernatant was then centrifuged at 13000 g for 15 min at 4ºC to pellet microbial cells. Immediately after that, the whole-cell protein extraction was performed by mixing one volume of disruption buffer and one volume of the microbial cell pellet which was heated at 80ºC for 1 h, with 2 min treatments in an ultrasound bath every 10 min. The disruption buffer contained 150 mM NaCl, 2% sodium dodecyl sulfate (SDS), 100 mM ethylenediaminetetraacetic acid (EDTA), 1M Tris HCl, pH 8.0, 100 mM 1,4-dithio-D-threitol (DTT) and a ¼ tablet of Complete protease inhibitors (Roche Applied Science, Germany) added to 1 mL of buffer. The above procedure was followed by 7 cycles of 5 s sonication, 5 min centrifugation at 15,000 x *g* and spin filtering of the supernatant for 7 hours using Vivaspin filters (Sartorius, Germany) with a molecular weight cut off of 10,000 Da, after 100-fold dilution, to reduce the SDS concentration, using 20 mM triethylammonium bicarbonate buffer (TEAB). Urea was added to a final concentration of 4 M prior to spin filtering to accomplish a better recovery of proteins. Quantitation of the extracted protein was carried out by means of the Bradford Protein Assay (Bio-Rad).

2.4 Mass spectrometry and data analysis

Peptides obtained from the protein pellets were measured by mass spectrometry using a nanoAcquity UPLC (Waters, Milford, MA, USA) coupled to an LTQ-Orbitrap Velos (Thermo Fisher Scientific, Waltham, MA, USA). Samples were injected using an autosampler and concentrated on a trapping column (nanoAcquity UPLC column, C18, 180 µm × 2 cm, 5 µm, Waters) with water containing 0.1% formic acid at a flow rate of 15 µL/min. After 6 min, the peptides were eluted into a separation column (nanoAcquity UPLC column, C18, 75 µm × 15 cm, 1.75 µm, Waters). Chromatography was performed with 0.1% formic acid in solvent A (100% water) and B (100% acetonitrile). The gradient was 2 to 15% B (0-10 min), 15 to 40% B (10-77 min), 85 % B (77-87 min), followed by reequilibration at 2% B for 13 min. For an unbiased analysis, continuous scanning of eluted peptide ions was carried out between 300 and 2000 m/z, automatically switching to MS/MS collision induced dissociation (CID) mode on ions exceeding an intensity of >2000, and ten MS/MS events per survey scan. For MS/MS CID measurements, a dynamic precursor exclusion of 120 s was enabled.

To increase protein identification rates and to decrease the false negative rate, the MetaGenomic Sequences (BioProject IDs PRJNA222659 [for MGS-HAV], PRJNA222657 [for MGS-MES], and PRJNA222658 [for MGS-PRI] at NCBI; AZIB00000000 [for MGS-HAV], AZIC00000000 [for MGS-MES], and AZIF00000000 [for MGS-PRI] at DDBJ/EMBL/GenBank) containing a total number of 54,323 unique/non-redundant sequence entries, were complemented by a 2-step database approach. The final database was used for protein identification in MaxQuant (v. 1.5.1.0). Oxidation of methionine was defined as variable modification and carbamido-methylation of cysteine as fixed modification. The three different sample sets were divided into different parameter groups and matching was

disabled. All remaining standard settings were kept. These included a peptide and protein false discovery rate (FDR) below 1%, at least 1 peptide, a precursor mass tolerance of 4.5 ppm after mass recalibration and a fragment ion mass tolerance of 0.5 Da. The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium (http://www.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD001490. Predicted protein sequences were aligned against the National Center for Biotechnology Information non-redundant (NCBI nr) database by using BLASTP. Taxonomic binning of the sequences was performed by summarizing the top significant BLASTP hits with e-values ≤ 0.00001. Additionally, composition-based binning of contigs containing the genes encoding the identified proteins was performed using the GOHTAM Web-server, to determine taxonomic affiliations.

2.5 Metabolomic fingerprint analysis of sediment samples

Metabolites were extracted from sediment samples as follows. In 100 mL Erlenmeyer flasks 5 g of sediments were mixed with 10 mL of cold (-80ºC) HPLC-grade methanol (MeOH). The samples were sonicated for 120 sec (at 15 W) on an ice water bath. This procedure was repeated 4 times, and the samples were kept on ice for at least 2 min between each step. After sonication, the supernatant was removed by centrifugation at 10,000 x *g* for 30 min at 4ºC, and the supernatant was stored in 20 mL serum vials at -80ºC. MeOH extracts were filtered using 0.2 µm nylon syringe filters (Thermo Scientific, Rockwood, USA) and analyzed by LC-Q-TOF-MS. Briefly, samples for LC-MS analysis were prepared by filtering the methanol extracts using 0.2 µm nylon syringe filters (Thermo Scientific, Rockwood, USA). The analytical run began with the analysis of Quality Controls (QCs), followed by the samples in random order; a QC sample was injected in between blocks of four samples until the end of the run. All vials were maintained at 4ºC throughout the run. Each metabolic fingerprint was determined using a liquid chromatography system consisting of a degasser, binary pump, and autosampler (1290 infinity, Agilent). A total of 0.5 µL of extract was applied to a reverse-phase column (Zorbax Extend C_{18} 50 × 2.1 mm, 3 µm; Agilent), which was maintained at 60°C during the analysis. The system was operated in positive and negative ion mode at a flow rate of 0.6 mL/min with solvent A (water with 0.1% formic acid) and solvent B (acetonitrile with 0.1% formic acid). The gradient was 5% B (0-1 min), 5 to 80% B (1-7 min), 80 to 100% B (7-11.5 min), and 100 to 5% B (11.5-12 min), followed by reequilibration at 5% B for 3 min (15 min of total analysis time). Data were collected in positive and negative ESI mode in separate runs on a Q-TOF (Agilent 6550 iFunnel). Both ion modes were operated in full scan mode (*m/z* 50 to 1,000 in positive and *m/z* 50 to 1,100 in negative ion mode). For each mode, the capillary voltage was 3,000 V, the scan rate was 1.0 spectrum/s, the gas temperature was 250°C, the drying gas flow was 12 L/min, and the nebulizer was 52 psi. The MS-TOF parameters for positive ion mode were as follows: fragmentor 175 V, skimmer 65 V, and octopole radio frequency voltage (OCT RF Vpp) 750 V. The same MS-TOF parameters were used in negative ion mode, except the fragmentor, which was set to 250 V. Two reference masses were used for each mode: *m/z* 121.0509 ($[C_5H_4N_4+H]^+$) and *m/z* 922.0098 ($[C_{18}H_{18}O_6N_3P_3F_{24}+H]^+$) during positive analysis and *m/z* 112.9855 ($[C_2O_2F_3-H]^-$) and *m/z* 1033.9881 ($[C_{18}H_{18}O_6N_3P_3F_{24}+TFA-H]^-$) during negative analysis. The reference masses were continuously infused into the system to permit constant mass correction. Compound identification and peak integration were performed using Mass Hunter Qualitative Analysis (B.05.00, Agilent).

3 Results

3.1 Study sites: physical-chemical characteristics

As shown in Table 1, the three studied chronically contaminated sediment sites exhibited the following characteristics at the sampling time: i) a temperature ranging from 15°C (for HAV) to 23°C (for MES); ii) pH from 6.85 (for PRI) to 8.05 (for HAV); iii) oxygen concentration ranging from oxygen depleted conditions (<0.1mg L^{-1}) (for PRI) to 6.0-6.50 mg L^{-1} (for HAV); iv) conductivity ranging from 49.0 (for HAV and PRI) to 70.0 (for MES) ms cm^{-1}; and v) from a total concentration of 1,000 mg petroleum hydrocarbons kg^{-1} sediment (for PRI) to tar sample (for HAV).

Table 1 Overall physical-chemical characteristics of the investigated sediment samples

	HAV	PRI	MES
GPS coordinates	44° 22'25.75"N 8° 41'59.58"E	37°10'27.462"N 15°12'7.505"E	38°11'42.267"N 15°34'25.014"E
Water Depth (m)	78.0	6.0	1.0
Petroleum Hydrocarbons (ppm)[1]	Tar	4,000	1,000
Temperature (ºC)	15.0	19.0	23.0
Dissolved O_2 (mg L^{-1})[2]	6.0-6.5	0	1.0-2.2
pH	8.05	6.85	7.37
Conductivity (mS cm^{-1})	49.0	49.0	70.0
Ammonium (µmol L^{-1})	0.6-0.7	420	7
Dissolved organic carbon (DOC) (mg L^{-1})	5.00	125.00	50.00
Open reading frames in metagenome[3] / Total number of scaffolds	8,388 / 4,412	5,858 / 1,855	40,077 / 20,103

[1] Total petroleum hydrocarbon concentration mg kg^{-1} (ppm)
[2] PRI is an anoxic site; MES is a micro-aerophilic environment.
[3] Gene prediction results from sequencing data obtained by Illumina HiSeq and Roche 454 sequencing of metagenomic DNA from the microbial communities in the three polluted sediments collected in the Mediterranean Sea.

3.2 Metaproteomic analysis: general comments

We selected a shotgun metaproteomic approach to analyse the active populations thriving in chosen chronically contaminated sites. We were aware that identification could strongly depend on protein abundance and a large part of the proteome, represented by minor quantities thus may remain hidden. Nevertheless, it can be assumed that the identified proteins represent the dominant pathways driven in each analysed ecosystem. In total, 651 non-redundant proteins were unambiguously identified (with total numbers of proteins in MPS-HAV: 310; MPS-PRI: 388 and MPS-MES: 333). Since the proteome analysis could be correlated with the corresponding reference metagenome datasets, the metaproteome sizes were within the range commonly observed for other communities and, as it is often the case, a few times smaller than those observed for cultured organisms. Although distinct environmental sites were investigated, 106 of 651 proteins (or 16.2%) comprised a subset of proteins common to all three samples. A total of 90, 143 and 144 proteins were specifically identified in HAV, MES and PRI metaproteomes respectively. This indicated that the HAV, MES and PRI communities displayed overlaps, but also some considerable heterogeneity at the level of proteome composition and at the level of functional categories. The metaproteomic approach applied here allowed us to compare taxonomic annotations and to evaluate the differences

between the contributions of particular groups of organisms in the overall communities, as well as to predict the importance of particular sets of proteins for the overall functioning of the community.

3.3 Identities of expressed proteins

Taxonomic classifications revealed that in all three samples the proteins assigned to Bacteria were predominant compared to archaeal proteins from the microbial communities (HAV: 94%; MES: 91.6%; PRI: 90.3% of the total proteins). Remarkably, the percentage of proteins binned to Archaea negatively correlated with O_2 concentration *in situ* with the highest percentage obtained at the anoxic site of Priolo Gargallo (r^2 = 0.99; *p*-value = 0.067). By contrast, the relative percentage of archaeal proteins did not correlate with either petroleum hydrocarbon concentration, *in situ* temperature or dissolved organic carbon. Evidently, the decrease in O_2 concentrations may stimulate the growth of such strictly anaerobic archaea, as methanogens, and as a result, in the enhanced expression of their proteins. Indeed, among all archaeal proteins those associated with methanogenic *Methanosarcinales* were most abundant in Priolo Gargallo sediments. Apart from proteins assigned to *Methanosarcinales*, other archaeal proteins most likely originated from *Thermococcales*, *Thermoplasmatales*, *Halobacteriales*, *Archaeoglobales*, *Halobacteriales* and *Aciduliprofundum*, were additionally observed in PRI but were absent from the HAV and MES metaproteomes (Figure 1A).

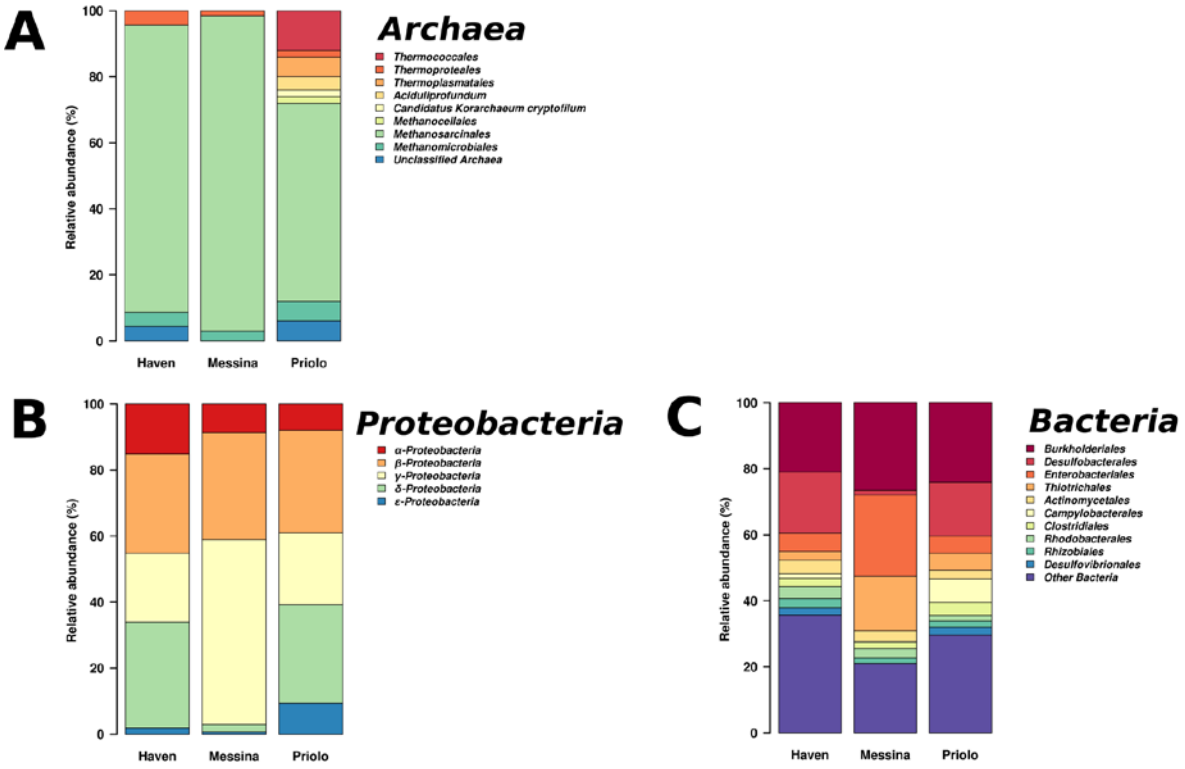

Figure 1 **Relative abundances and distributions of archaeal (A) and bacterial (B, C) proteins in the metaproteomes of marine sediment samples based on taxonomic bins for proteome-derived proteins that could be assigned a taxonomic annotation.**

Within the bacterial sub-proteomes, proteins assigned to the phylum *Proteobacteria* were predominant in all three studied samples (HAV: 66%; MES: 57%; PRI: 81%, with respect to total protein numbers). The distribution at the level of *Proteobacteria* classes (Figure 1B) revealed only

significant differences in the number of proteins assigned to *Deltaproteobacteria*. Consistent with the physiological constraints, the expressed proteins belonging to these mostly anaerobic bacteria were detected at significantly lower levels in oxygen-saturated MES sample than in microaerobic HAV (8.4-fold, in terms of numbers identified) and in anaerobic PRI (9.5-fold).

Proteins assigned to microaerophilic and anaerobic *Epsilonproteobacteria* were by far more abundant in PRI than in HAV (7.2-fold) and MES (11.6-fold) (Figure 1B). Since PRI is the anaerobic environment (Table 1), the higher numbers of proteins from *Epsilonproteobacteria* most likely reflect their association with anoxic sites; notably, members of this class have also been found to be abundant in pollutant-degrading microbial consortia operating under anaerobic conditions.

By contrast, the proteins assigned to *Gammaproteobacteria*, were expressed at higher levels in MES than in HAV (2.8-fold) and PRI (2.6-fold) (Figure 1B).

In addition to proteins from the phylum *Proteobacteria*, the proteins most likely derived from *Firmicutes* (HAV: 7.8%; MES: 6.2%; PRI: 8.1% of total proteomes) and *Acidobacteria* (HAV: 5.8%; MES: 4.4%; PRI: 4.8%), formed the second and the third predominant groups. Additional groups of Bacteria that expressed proteins at quantities allowing taxonomic binning/identification under our assay conditions are summarized in Figure 1C. These findings were verified by phylogenetic analysis of NCBInr derived peptides using the web-based Unipept tool.

Among all proteins, the ten functional groups of the most abundant proteins were identified:

i. ABC transporters or outer membrane proteins (116, or 18% of the total),
ii. hypothetical proteins (109, or 17%);
iii. ribosomal proteins (32, or 5%);
iv. AprAB adenosine-5-phosphosulfate reductases (22, or 3.4%);
v. DsrAB sulfite reductases (18, or 2.8%);
vi. TonB-dependent receptors (14, or 2.2%);
vii. chaperones (14 or 2.2%);
viii. glutamate decarboxylases/dehydrogenases/synthases (13, or 2%);
ix. ATP synthases (9, or 1.4%); and
x. proteins for the C1-compounds uptake and methanogenesis [9 methyl-coenzyme M reductases, 7 methanol dehydrogenases and 7 methanol-5-hydroxybenzimidazolylcobamide methyltransferases; or 3.5%]).

This suggests that apart from transport and energy production systems, sulfur, C_1-compounds metabolism and glutamate metabolism are among the most active ones within all communities, albeit at different levels. Relative enrichment of iv (AprAB), v (DsrAB), and (McrA) in PRI ecosystem compared to MES and HAV, correspond to the environmental constraints, because sulphate reduction and methanogenesis have been commonly found to be markedly stimulated under anaerobic conditions. Note, that proteins involved in glutamate conversion were only found in HAV and PRI (11 and 8 non-redundant proteins, respectively), but not in MES; the fact that glutamate metabolism is generally activated in response to low temperature stress, and that both HAV and PRI were characterized by a much lower seawater temperature compared to MES (Table 1).

3.4 Differences at the metabolic and organismal levels as revealed through metaproteomics

As is well established, the proteomics approach is able to determine the active metabolic pathways operating in microbial communities and thus, can significantly facilitate the study of the metabolic network present. Here, on the basis of their probable functions, at the level of metabolic pathways we were able to capture overall functional differences of the three chronically polluted sites. Specifically, in relation with sulphur and C1 metabolism, and hypothesized different pathway

organizations at the organism level we detected remarkable distinction. Below these differences are detailed.

C1 metabolism: A total number of 41 unique proteins (or 6.1% of the total proteome) potentially involved in the CH_4, CH_3OH and CO metabolism (HAV: 16; MES: 20; PRI: 24) were detected (Figure. 2). This corresponds to a 31% of the theoretical number of proteins (*in silico* proteome) presumptively involved in C1 metabolism, as determined after examination of potential protein-coding genes (≥ 20 amino acids long) from the meta-sequences. Protein signatures for the initial step of CH_4 conversion to CH_3OH were only found in HAV; this conversion can be performed by methane monooxygenase from bacterial members of the order *Methylococcales* (the top hit BAH22843, beta-subunit, PmoB). Enzymatic pathways for C1-compounds uptake and (methylotrophic) methanogenesis were detected in all three communities, although to different extent (Figure 2). The evidence for methylotrophic methanogenesis were the presence of: *i*) methanol corrinoid proteins, methanol-5-hydroxybenzimidazolylcobamide methyltransferases and methylcobalamin:coenzyme M methyltransferases converting CH_3OH to methyl-CoM; and *ii*) methyl-coenzyme M reductases and heterodisulfide reductases. In addition, trimethylamine:corrinoid methyltransferases converting trimethylamine to methyl-CoM and dimethylamine, were also found in the proteomes. Notably, all these enzymes were unambiguously attributed to *Methanosarcinales*.

Signatures for methanol-utilizing bacterial methylotrophs were also identified in the proteomes. Proteins for subsequent conversion of CH_3OH to formaldehyde by methanol dehydrogenase were only detected in MES and PRI (Figure 2). *Proteobacteria*, especially those associated to members of the orders *Thiotrichales* (in MES and PRI), *Vibrionales* (in MES) and *Rhizobiales* (in MES), were the ones capable of presumptive utilization of CH_3OH (Figure 3). Methanol conversion to formaldehyde was not evident from HAV proteome data. Enzymes for the subsequent metabolism of formaldehyde were found in MES and PRI, namely formaldehyde-activating enzymes that transform formaldehyde to tetrahydromethanopterin (THMPT). Notably, the occurrence of the latter conversion was supported by proteins assigned to *Methanosarcinales* (in MES and PRI) and *Thiotrichales* (in PRI). No signatures for the conversion of formaldehyde to formate were observed, although a formate dehydrogenase (binned to *Rhodospirillales*) converting formate to CO_2 was found in MES and PRI.

A carbon monoxide dehydrogenase (CODH catalytic subunit; top hit WP_031450384.1) (binned to *Desulfobacterales*) for the carbonyl branch of the Wood-Ljungdahl pathway was found in PRI (Figure 2). Finally, a carbon dioxide concentrating mechanism/carboxysome shell protein (top hit WP_015107688.1), a structural protein involved in the storage of enzymes participating in CO_2 metabolism, such as carbonic anhydrase and RubisCO was found in both MES and PRI.

Of course, we are aware that some caution is required in the interpretation of suggested networks. Based on profiles of detected and taxonomically affiliated proteins, the implication seems to be that the different organisms provide different functions in a whole network. Nevertheless, this does not exclude the complete pathways present in a single organism. These signatures are possibly not seen due to relatively low abundance (i.e. detection) of the relevant proteins in that organism. In any cases, observed protein profiles were used to design an integrative model of metabolic network whereby the different organisms can share common metabolites present in the environment. Obviously, further studies are required to elucidate robustness of the interpretations. Moreover, special emphasis on proportion of the proteome that is sampled from each organism should be done.

Taken all proteome data together, the presence of two active processes is suggested: methanogenesis and methylamine utilization mediated by Archaea (*Methanosarcinales*) and methanol/formaldehyde detoxification by a set of bacteria. In addition, all studied communities

possess highly developed trophic networks based on assimilatory and dissimilatory metabolism of C1-compounds.

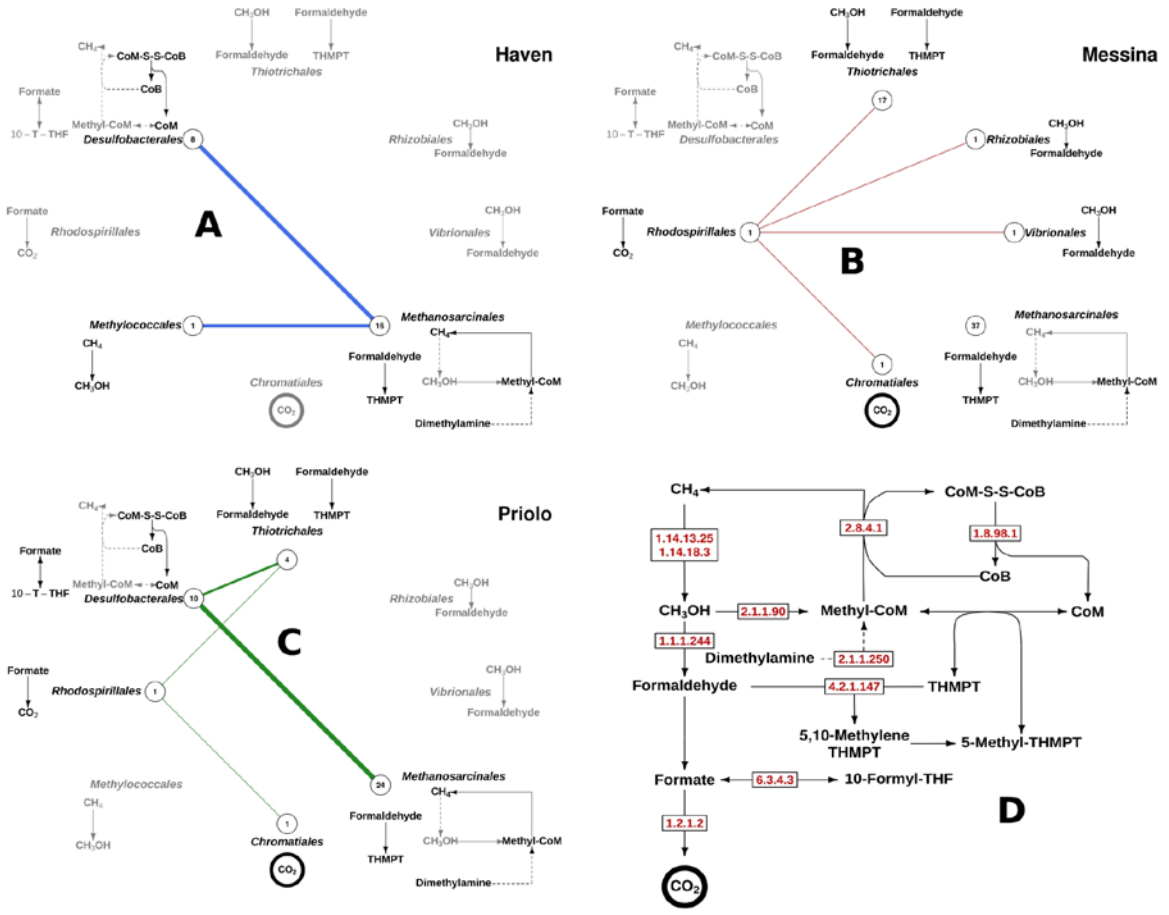

Figure 2 **Most prevalent reactions of metabolic network proposed in marine sediment samples as revealed by detection of various proteins in HAV(A), MES (B) and PRI (C) proteomes. Proposed correlation of detected proteins coupled with C1 metabolism (D).**
Circles represent the relative abundance of proteins in a sample assigned to each taxonomic group; the total number of proteins for each taxonomic group is indicated in circles. Solid lines represent eventual connections of the substrates and products of contiguous reactions as revealed by detection and taxonomical binning of proteins. The connecting line thickness reflects the relative abundance of proteins assigned to each of the reactions and taxonomic group. Grey colour indicates transformations for which no proteins in the proteome were identified, black colour represents transformations for which proteins were found (putatively active pathways). Transformations in the CH_4 and CH_3OH metabolism for which no proteome evidences were found (e.g. glutathione (GSH) pathway connecting methanol with CO_2) are not indicated.

The taxonomy-guided pathway reconstruction emphasized differences at the organism level in relation to the CH_4 and CH_3OH metabolism. Thus, Figure 2 shows the contribution of major organisms in all three sites, in which eventual interconnection and/or pathway partitioning can be suggested. As we mentioned above, applied technology does not exclude the existence of complete pathways in a single organism and what we detected reflect the metabolic coupling between community members based on capabilities to transform same compounds.It can be seen how the metabolism and activation of CH_4 seems to be supported by a bacterial member of the *Methylococcales* order. Finally, the further methylotrophic conversion of CH_3OH was dominated by bacterial proteins (from

members of the orders *Thiotrichales*, *Vibrionales*, *Rhizobiales*, *Rhodospirillales* and *Desulfobacterales*) and, to some extent, by archaeal proteins from methylotrophic methanogens of the *Methanosarcinales*. These organisms possess different respiration requirements and likely spatially separated. Their co-existence in proteome could be explained by the sediment sampling methodology which does not exclude the mixing from oxic and anoxic parts of the sediment leading to representation of pathways typically associated with either aerobes or anaerobes.

Sulfur metabolism: A total number of 44 unique proteins (or 6.8% of the total) potentially involved in the metabolism of sulphur compounds were detected. Notably, while these proteins were abundant in HAV (23) and PRI (40), they were absent in MES (Figure 3). This is particularly noticeable given that the metagenome sequence data of sample MES was 4.8 and 6.9-fold bigger as compared to that of samples HAV and PRI, respectively (Table 1). The absence of these proteins as revealed in the MPS-MES proteome must reflect their low biological significance in MES. In fact, 27 potential protein-coding genes (≥ 20 amino acids long) presumptively involved in sulphur metabolism were found in the meta-sequences of MES), none of which were detected at expression levels in corresponding proteome (Figure 3).

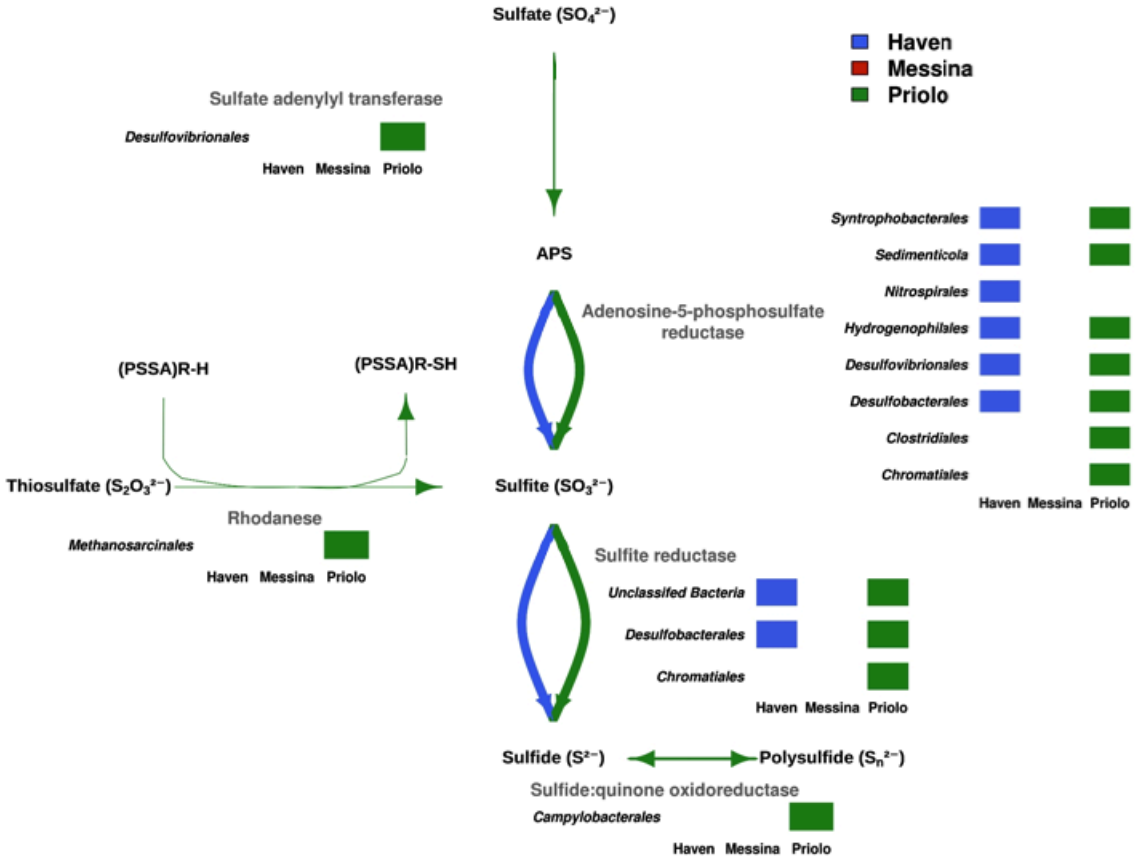

Figure 3 **Sulfur metabolism patterns in microbial communities inhabiting petroleum-polluted marine sediments based on proteome analysis.**
The presence of enzymes for each transformation reaction (linked by solid lines) and the taxonomic affiliation of polypeptides are shown. The thickness of solid lines represents the abundance level of enzymes (referred to the total number proteins assigned to these pathways) associated to each of the transformations in the pathway. Transformations in the sulphur metabolism for which no proteome evidences were found (e.g. sulfur assimilation metabolism via ATP sulfurylase) are not indicated.

We detected AprAB adenosine-5-phosphosulfate reductases and DsrAB sulfite reductases converting sulfite (SO_3^{2-}) to adenylyl sulfate (or facilitating the reverse reaction) and sulfide (H_2S), both in HAV (23 proteins in total) and PRI (36 proteins in total) (Figure 3). A total of 15 AprAB proteins conformed to the common set. Proteins in HAV were assigned to anaerobic bacteria belonging to *Desulfobacterales* (15), *Desulfovibrionales* (2), *Nitrospirales* (1), *Sedimenticola* (1) and *Syntrophobacterales* (2). Proteins in PRI assigned to *Desulfobacterales* (21), *Chromatiales* (5), *Syntrophobacterales* (3), *Clostridiales* (2), *Desulfovibrionales* (2), *Hydrogenophilales* (2), and *Sedimenticola* (1). However, one sulfate adenylyltransferases converting sulfate (SO_4^{2-}) to adenylyl sulfate (binned to *Desulfovibrionales*), two sulfide:quinone oxidoreductases converting *n* sulfide (*n* HS^-) to polysulfide (a defensive HS-oxidation pathway, implying to dump the excess of electrons from the cytoplasm/membrane) binned to *Campylobacterales*), and a rhodanese involved in the detoxification of cyanide (CN^-) binned to *Methanosarcinales* (note that rhodaneses may have in euryarchaea another function as sulfurtransferases, involved in elemental sulphur respiration), were only found in PRI. Taken together, the PRI microbial community contained abundant proteins involved in various metabolic pathways of both oxidized and reduced sulfur intermediates. None of these metabolic capacities were detected in MES sample, suggesting a lower gross importance of sulfur cycle in biogeochemical functioning of MES ecosystem.

3.5 Community metabolic activity associated with biodegradation

Only one relevant protein, a carboxymuconolactone decarboxylase (EC: 4.1.1.4), involved in the protocatechuate catabolism, was identified in the studied metaproteomes. This protein, identical to EGD01850 from *Burkholderiales* (100% sequence identity), was only found in HAV and PRI. This suggests that enzymes potentially involved in biodegradation steps are produced at much lower levels as compared to other proteins. Accordingly, the presence of chemical intermediates in the aerobic and anaerobic degradation of pollutants in the oil-polluted sites investigated might be significantly low. To demonstrate this, we used a metabolomics approach to identify chemical signatures indicating biodegradation capabilities by microbes inhabiting HAV, MES and PRI sediments. Our protocol comprised the isolation of metabolites from microorganisms obtained from sediment material followed by a metabolome-wide scan via a combination of mass spectrometry (MS) with liquid chromatography (LC) separation. In such a way, we could identify both pollutants being transported inside or attached to the microbial cells, as well as chemical intermediates produced during their degradation.

A total of 1,485 (LC-MS negative mode) and 3,390 (LC-MS positive mode) mass features were found in each analysis after deconvolution. Empirical formulae were assigned to accurate masses with a maximum error of 5 ppm using CEU Mass Mediator (http://biolab.uspceu.com/mediator) and putative chemical species participating in degradation (either initial substrates or intermediates) were specifically identified. Surprsngly, only 24 out of 4,776 metabolite mass features (or 0.5% of the total) were tentatively attributed as pollutants or chemical intermediates.

4 Discussion and conclusion

The Mediterranean Sea has a number of sites subjected to the chronic pollution by petroleum. The total concentration of hydrocarbons in the studied sediments (Table 1) exceeded that in clean seawater (15 ppm) by more than 70-fold. These sites host finely tuned microbial communities, with a phylotype richness highly dependent on the environmental conditions, the type of pollutant and its load. We performed a multiphasic analysis using metaproteomic and meta-metabolomics approaches supported by metagenome sequencing data produced earlier, to establish protein and metabolite compositions from microbial communities inhabiting contaminated marine sediments

from three sites located in the Mediterranean. These proteins and the source microorganisms may play important roles in the cycling of prevalent and persistent pollutants, as well as in the total carbon cycling.

We detected the presence of methanotrophic bacteria that can oxidize methane through sequential reactions catalyzed by a series of enzymes including methane monooxygenase, methanol dehydrogenase, formaldehyde dehydrogenase, and formate dehydrogenase. This full conversion was suggested to be active only in the HAV, which is consistent with the highest concentration of oxygen in this site. PRI and MES did not show evidence for active bacteria utilizing methane but rather a high diversity of methanol-consuming bacteria, which was also observed in HAV. C1 compounds such as methanol, which may be produced as an intermediate of organic matter biodegradation, can thus be used by communities inhabiting all three sites as another source of carbon and energy. Our results have revealed the presence of microbes presumably active in methane oxidation to be uniquely present on M/T *Haven* site. The fact that the first step in the aerobic methane oxidation was only found in HAV, agrees with the higher O_2 concentration in this site, as compared to MES and PRI. In addition to that, all three sediments did contain microbes that utilise biogenic C_1 compounds, such as methanol. Accordingly, the ability to utilize methanol and some other single carbon compounds, but not methane, was found in MES and PRI, whereas PRI and HAV were populated by trimethylamine-utilizing microbes. Figure 2 shows the simplified metaproteomics-based reconstruction of methane and methanol metabolism. Results evidenced the contributions of different archaeal and bacterial groups were linked to active methane and methanol cycling (5 groups in HAV and 8 in MES and PRI). Among the 10 groups involved in those metabolisms, only 3 conformed to the common set. Particularly noticeable is the absence in MES of proteins from *Desulfobacterales* whose relevance in in HAV and PRI was quite pronounced.

As it revealed by proteomic analysis, Figure 2 and Figure 3 depict possible microbial metabolic networks between the members of microbial communities. We hypothesize that individual members could run C1 metabolic pathway only partially and the identified proteins do represent predominant reactions. The corresponding transformations may be complementary to those supported by other community members, so that the whole pathway becomes complete. Although we are aware, that due to specificity of applied approach, it is important to highlight the alternative explanation(s) e.g. that a number of different organisms are conducting the same process concomitantly. In other words, as many molecular approaches, the proteome analysis is not bias-free so the observed data could result from all organisms present being able to carry out the full process, but in all cases proteins covering the full pathway have not been sampled/detected in the analysis.

The examined metaproteomes contained few signatures of active biodegradation pathways, possibly due to the low rates of biodegradation processes in chronically petroleum-polluted sediments that are limited by low oxygen (or presence of other forms of organic matter). This has led to the prevalence of proteins involved in metabolism of C_1–compounds, suggesting the latter segment of the carbon cycle is most active under *in situ* conditions. The metabolome profiling was more sensitive in the detection of putative biodegradation pathways by identification of substrates and intermediates of degradation pathways of petroleum constituents. However, the fact that only a minor proportion of known initial and intermediate metabolites for the degradation of alkanes and polyaromatic hydrocarbons were identified suggests that the accumulation of those occurred at a low level reflecting low degradation rates. Another possible explanation is that absence of degradation intermediates could equally indicate rapid turnover resulting in low standing concentrations of the metabolites.

The combination of metaproteomics (to reveal major active pathways) and meta-metabolomics (for a sensitive and targeted detection of pathways' intermediates) used in present work seems to be

useful to reconstruct and understand the functionality and interactions of microbial community members at different scales and to identify the main contributors to the processes of different carbon and energy sources. Furthermore, the results of this study could draw some attention to the yet untapped diversity of proteins from archaea and bacteria operating in the Mediterranean Sea sediments under conditions of limited oxygen concentrations.