MIBIREM – Toolbox for Microbiome based Remediation

Deliverable 2.3

Metagenome and transcriptome of the six most promising microbiomes

Due Date:	31 December 2024
Submission Date:	20 December 2024
Dissemination Level:	PUBLIC
Lead Beneficiary:	CNRS
Lead Author:	Aurélie Cébron, CNRS aurelie.cebron@univ-lorraine.fr
Project Acronym: MIBIREM	**Project Number**: 101059260
Start Date of Project: 01/10/2022	**End Date of Project**: 31/03/2027

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Document Control Information	
Title	*Metagenome and transcriptome of the six most promising microbiomes*
Editor	*Aurélie Cébron (CNRS)*
Contributors	*AIT, UNIPI, SENSA, UHAS*
Reviewer/s	*Thomas Reichenauer (AIT), Sofie Thijs (UHAS), Simona di Gregorio (UNIPI)*
Dissemination Level (please choose just one according to the DoA)	☐ **SEN** - Sensitive (please provide one-page Publishable Summary) ☒ **PU** - Public
Approved by (please tick the boxes after all the partners have provided their written approval of the deliverable)	X RTDS (COO) X AIT (SCO) X UHAS X CNRS X UGENT X SENSA X ALTAR X DND X UU X UNIPI X TAUW X ESC
Underlying IPRs	*Are there intellectual property rights included in this deliverable? If yes, please specify.* **No**
Underlying Datasets	*Are there relevant datasets included in this deliverable? If yes, please specify.* **Yes, List of microbiomes of interest**

Version/Date	***Change/Comment***
Version 1, 02/12/2024	*Initial draft by CNRS*
Version 2, 13/12/2024	*Final draft by CNRS*
Version 3, 20/12/2024	*Final, approved version by project consortium*

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Table of Contents

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

1 Publishable summary

This deliverable describes the sequencing of metagenomes and metatranscriptomes of the most promising microbiomes with efficient degrading activities for the three use cases. These data (currently under acquisition) will be analysed to: i) reconstruct the genomes of each individual bacteria from the consortia (MAGs), ii) predict their biochemical pathways and potential role in the consortia, and iii) identify the genes and pathways that are over-expressed or repressed when the consortia are grown on contaminants and during degradation. The metagenome and transcriptome were planned on the six most promising microbiomes, but were actually performed on ten.

This information will be used to select the best microbiomes for the next steps, and mainly for directed evolution. In addition, these data will be further used to target functional genes in bioaugmentation and bioremediation assays during pilot and field testing of our microbiomes.

As a complement, ongoing actions will result in SIP-metabarcoding and metagenomic datasets that will be included in an updated version of the deliverable prior to M36.

2 Introduction

From the enriched microbiomes, the most promising ones with efficient degradation activities for the three use cases, were selected for metatranscriptomic analysis to identify pathways and the genes over-expressed during contaminant biodegradation under different conditions. In order to efficiently analyse the metatranscriptomic data obtained by sequencing, from a functional point of view, metagenomic sequencing and analysis was also performed to reconstruct the genomes of each of the strains forming the microbiome consortia. Metagenomics will give us insight into the full functional pathways held by the strains that make up the microbiomes. It will also identify metagenome-assembled genomes (MAGs): microbial genomes reconstructed from metagenomic data.

It was also planned to perform omic analysis on stable isotope probing (SIP) experiments. SIP uses ^{13}C- or ^{15}N-labelled contaminants added to soil or groundwater samples in laboratory microcosms, and allows the $^{13}C/^{15}N$-labelling of the microbiomes involved in the degradation (use of carbon or nitrogen for their growth). This method will provide insight into the diversity and the functions of the active microbiomes. Many steps of development and preliminary experiments were needed to evaluate the kinetic parameters of biodegradation of each use case, and SIP experiments were only performed on one contaminant (hexadecane). Due to difficulties, we decided to perform DNA-SIP instead of RNA-SIP. Thus, SIP-metabarcoding and metagenomic datasets will be added in an updated version of the deliverable prior to M36.

3 Metagenomes and metatranscriptomes of microbiomes degrading the three use-cases

3.1 List of microbiomes

Six microbiomes were planned to be analysed for metagenomic and metatranscriptomic analyses, but we ended up analysing ten microbiomes in more detail (list in **Table 1**).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Table 1. List of the microbiomes, their site of origin, the contaminant degraded (use-case) used for metagenomic and metatranscriptomic analyses

Microbiomes	Site of origin	Use case[1]	Partner	# Metagenomes (Mg)/Metatranscriptomes (Mt)
PLO_05_A_7.2a	Ploufragan, France	PHC-diesel	CNRS	3 Mg, 9 Mt
PLO_05_A_7.2b	Ploufragan, France	PHC-oil	CNRS	3 Mg, 9 Mt
NL_UTR_SC2	Utrecht, Netherlands	PHCs, PAHs	UHAS	5 Mg
NL_UTR_SC4	Utrecht, Netherlands	PHCs, PAHs	UHAS	8 Mg
AT_SIE_01A_BT	Siebenhirten, Austria	Cyanide	AIT	4 Mg, 10 Mt, 20 16S amplicon
DE_STO_04_GW	Stockach, Germany	Cyanide	AIT	4 Mg, 10 Mt, 20 16S amplicon
AT_SIM_04_BT	Simmering, Austria	Cyanide	AIT	4 Mg, 10 Mt, 20 16S amplicon
IT_COL_03_C_S_EC002_T02	Colleferro, Italy	HCH	UNIPI	1Mg/3Mt
IT_COL_03_C_S_EC002_T02 L001	Colleferro, Italy	HCH	UNIPI	1Mg/3Mt
IT_COL_03_C_S_EC002_T02 l002	Colleferro, Italy	HCH	UNIPI	1Mg/3Mt

[1] PHC: petroleum hydrocarbon, PAH: polycyclic aromatic hydrocarbon, HCH: hexachlorocyclohexane

3.2 Microbiomes degrading petroleum hydrocarbons (PHC)

3.2.1 Microbiome from Ploufragan site degrading diesel and oil

A microbiome (PLO_05_A_7.2) was selected from the ones enriched using diesel as the sole carbon source (bushnell Hass medium + 1% diesel) from the Ploufragan soil samples. Since metatranscriptomic analyses require at least 2 conditions (a test compared to a control), it was decided to grow this microbiome on 2 different substrates (diesel and oil sampled at Ploufragan) and to evaluate the transcriptome at 3 different times during the biodegradation (15, 24 and 48h) and in triplicate (18 samples). This experimental design will allow us to identify deregulated genes over time on each growth substrate, and also to compare the specific genes involved in diesel and oil degradation. As changing the growth substrate will change the diversity (relative proportion of the different strains) in the consortium, the metagenomic analysis was performed for both conditions at one time point (24h), in triplicate (6 samples). In parallel, for both culture conditions (diesel and Ploufragan oil), the measurement of the degradation of the hydrocarbons (compound specific) was followed through liquid:liquid extraction and GC-MS analyses (method that required further development) over time.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

RNA and DNA extracts of good quantity and quality (analysed done using Qubit) were sent to Novogene for sequencing. For transcriptomics, 4 Gb of raw data were sequenced per sample. For metagenomics, 8 Gb of raw data were sequenced per sample. Novogene then performed the first steps of bioinformatics analysis: Metagenome Assembly using MEGAHIT to find contigs in the dataset, perform taxonomic annotation to assign each contig to the corresponding microorganism using the NR database (NCBI), and perform Differential Expression Gene (DEG) analysis to calculate Foldchange for each gene and significantly over and under-expressed genes.

The sequencing data will be available in early 2025 and will be deposited at SRA under MIBIREM Bioproject.

3.2.2 Microbiome from Griftpark site (UHAS)

For the PHC/PAH contaminated soil core samples from Griftpark (Utrecht), enrichment cultures were started from soil samples at different depths, on 7/6/2023 (**Figure 1**). The cultures were grown in Bushnell-Haas medium and every 2 weeks, optical density was measured and the cultures were spiked with an acetone-extract of the soil to add initial contaminants, or with 10 ppm naphthalene, 10 ppm benzene and 1 ppm phenanthrene. The most promising degrading consortia from SC2 and SC4 were selected for sequencing.

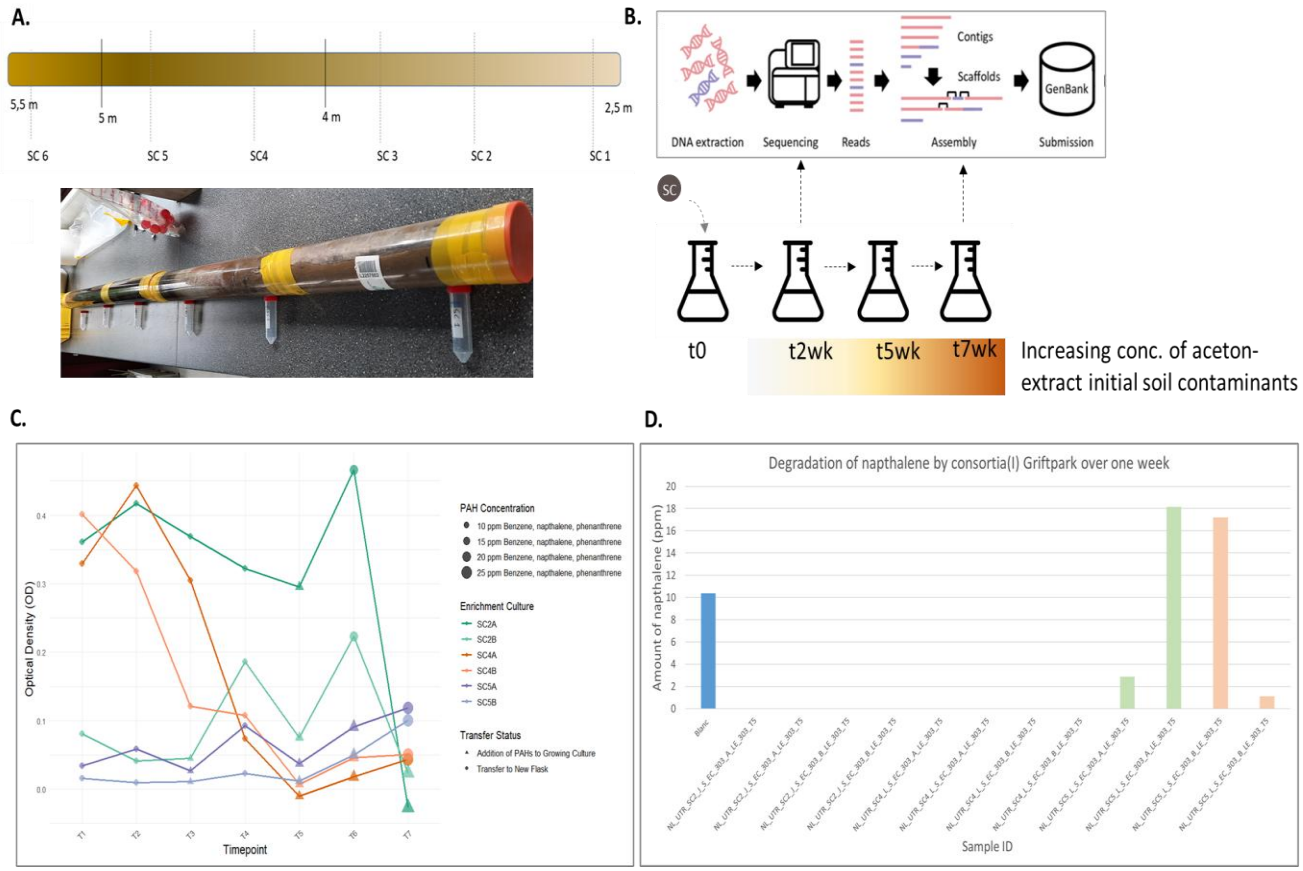

Figure 1: A. Picture of the soil core. B. Enrichment cultures of each soil core, transfer to new medium and increasing concentrations of pollutants. C. Optical density of the cultures in time. D. Degradation of naphthalene by the best consortia.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

At time points of 2 and 5 weeks, sub-samples were collected, centrifuged and pellets subjected to DNA-extraction. This was followed by amplicon sequencing and subsequently shotgun sequencing using the Illumina DNA-prep kit, and sequenced at 250 PE, Novoseq (Novogene).

In total, 254 million raw reads were generated with > 90 % of high quality (**Table 2**). Raw sequences were submitted to NCBI, SRA under number: SUB14924112.

Table 2: Raw sequencing reads

Sample	Raw reads	Raw data	Effective(%)	Error(%)	Q20(%)	Q30(%)	GC(%)
SC2_A_A_1_2wk	16142810	4.04	100	0.03	96.29	90.72	59.24
SC2_A_A_1_5wk	16328296	4.08	100	0.03	96.35	91.85	61.17
SC2_B_A_1_2wk	18650748	4.66	100	0.03	96.44	91.92	62.85
SC2_B_A_1_5wk	10045868	2.51	100	0.03	96.19	90.67	62.43
SC2_B_i_1_2wk	33137526	8.28	100	0.03	96.96	92.76	63.22
SC4_A_A_1_2wk	14925282	3.73	100	0.03	96.19	91.8	63.22
SC4_A_A_1_5wk	23607550	5.9	100	0.03	96.21	91.6	63.41
SC4_A_i_1_2wk	10065312	2.52	100	0.03	96.09	91.44	65.03
SC4_A_i_1_5wk	12661328	3.17	100	0.03	97.14	92.94	64.73
SC4_B_A_1_2wk	31500526	7.88	100	0.03	95.85	90.89	64.1
SC4_B_A_1_5wk	34478756	8.62	100	0.03	96	90.53	62.89
SC4_B_i_1_2wk	15987300	4	100	0.03	96.59	91.6	63.92
SC4_B_i_1_5wk	17360200	4.34	100	0.03	96.99	92.58	64.76

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

3.3 Microbiomes degrading Hexachlorocyclohexane (HCH)

For the HCH degrading microbiomes, a comparative approach was adopted. The comparison was related to the effect of different growth conditions and of the cryopreservation on the composition of the enriched microbiomes and the associated capacity to maintain the HCH degradation capacity.

To the scope the metagenome of three microbiomes were sequenced: one microbiome (IT_COL_03_C_S_EC002_T02) that was the best performing microbiomes in HCH depletion (with reference to the 4 main HCH isomers) enriched from the Colleferro site; the same microbiome after one month of cryopreservation in 30% glycerol and incubated for 1 month in mineral salt medium with HCH (100 ppm) as sole carbon source (IT_COL_03_C_S_EC002_T02_LE001); the selected enriched microbiome grown in Luria Bertani medium for 3 days (IT_COL_03_C_S_EC002_T02_LE002).

We obtained more than 2x 10^9 raw reads with > 90 % of high quality. The reason for the very high sequencing depth is related to the need to explore the metagenomic modification that might be involved in the adaptation to different growth conditions, especially with the perspective to select a growth medium for a massive growth. The possibility to evaluate a possible contribution of eukaryotes, even though the use of cycloheximide during the enrichment phase, has been also evaluated in relation to the sequencing depth.

The comparison between the three microbiomes will be analysed to depict genomic organisation essential to the growth of the latter in different growth conditions and the one that are essential to determine the HCH depletion. The whole set of data will be compared to the one associated with the effect of a routinary cryopreservation protocol on the composition of the microbiota in taxonomic terms and of the associated genomic organisation. A metatranscriptomic analysis of the three metagenomes has been also performed to assess the role of specific functions in HCH transformation. The metatranscriptomic analysis was performed by sequencing three biological replicates per microbiome.

This approach of metagenomic and metatranscriptomic analysis of the HCH degrading microbiomes with reference to the microbiomes degrading the other class of contaminants is due to the evident instability of the genomic traits responsible for the catabolic traits of interest. The evaluation of the differential genome organisation and transcription of genes of interest in different growth conditions was mandatory. The analysis of the WGS and of the metatranscriptomic data are ongoing and will be. Actually, the contribution of transposases and insertion sequences that might be responsible for the instability of the HCH depletion capacity genomic organisation has been evaluated in detail. Results of the analysis will be deposited in the SRA archive (https://www.ncbi.nlm.nih.gov/sra/) in early 2025.

3.4 Microbiomes degrading cyanide

Transcription profiles of three well-performing cyanide degrading enrichment cultures, each seeded from native microbial communities in unrelated groundwater bodies were investigated. The initial conversion of cyanide to ammonia was leveraged in a differential expression experimental design. Cyanide-degrading enrichments were fed with either ammonia or cyanide as the sole nitrogen sources. Each treatment was replicated 3-4 times. RNA was extracted from enrichments at four time points, for two the cyanide was still present in all or most samples, and cyanide-supplemented enrichments were sampled more often than those with ammonia as nitrogen source.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

After sequencing the mRNA, cyanide-specific molecular signatures in cyanide treatments, but not in enrichments supplemented with ammonia are expected. Data analysis is ongoing. Next to differential expression analysis of genes mapped to MAGs (metagenome-assembled genomes), gene expression levels will be compared to the mean gene expression levels.

In the ongoing data evaluation, particular attention will be paid to the differential expression of genes not associated with cyanide assimilation, as they might be important for ecological interactions (e.g., cross-feeding). Alternatively, they might indicate so far unknown cyanide degradation mechanisms.

The challenge of this metatranscriptomics approach is the additional complexity of mixed species cultures. Not only do we expect the mRNA signature to differ between the treatments, but also the ratios between community members. This poses a challenge for the differential expression analysis performed with the generated data, as we will need to correct the molecular responses to the cyanide for the species composition in complex microbial communities.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)]. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union