symbiorem

STRAINS AND METAGENOMICS OF THE CONTAMINATED SOIL MICROBIOME

Marco Candela, Silvia Turroni, Simone Rampelli,
Giorgia Palladino, Elena Radaelli

17/05/2024

PROJECT INFORMATION

Project Acronym	SYMBIOREM
Project Title	Symbiotic, circular bioremediation systems and biotechnology solutions for improved environmental, economic and social sustainability in pollution control
Grant Number	101060361
Project Coordinators	Dr. Leire Ruiz Rubio, Dr. José Luis Vilas Vilela, Dr. Mónica Loyo-Menoyo University of the Basque Country - UPV/EHU
Project Duration	1 September 2022 - 31 August 2026

DELIVERABLE INFORMATION

Deliverable No.	3.1
Dissemination Level [1]	PU
Work Package	3
Task	3.1
Lead Beneficiary	GAIKER
Contributing beneficiary(ies)	UNIBO
Due date of deliverable	29/02/2024
Actual submission date	29/02/2024

[1] PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)

DOCUMENT HISTORY

V	Date	Beneficiary	Author/Reviewer
1.0	14/02/2024	UNIBO	Marco Candela, Silvia Turroni, Simone Rampelli, Giorgia Palladino, Elena Radaelli
1.1	15/02/2024	GAIKER	Pilar Brettes, Josu Berganza
2.0	16/02/2024	UNIBO	Marco Candela, Silvia Turroni, Simone Rampelli, Giorgia Palladino, Elena Radaelli
2.1	23/02/2024	UPWR	Anna Kancelista, Zbigniew Lazar, Wieslaw Fialkiewicz
3	28/02/2024	EHU	Leire Ruiz

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

4	10/05/2024	GAIKER	Pilar Brettes, Josu Berganza
5	13/05/2024	UNIBO	Marco Candela, Silvia Turroni, Simone Rampelli, Giorgia Palladino, Elena Radaelli
5.1	17/05/2024	GAIKER	Pilar Brettes

TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Deliverable D3.1 is presented in this document as a public report within work package 3, task 3.1. UNIBO as the partner responsible for this deliverable is represented by Marco Candela, Silvia Turroni, Simone Rampelli, Giorgia Palladino, and Elena Radaelli. The aim of deliverable D3.1 is to present the genomic characterization of isolated bacterial strains from hydrocarbon contaminated soils from Azkoitia (Spain) and the shotgun metagenomics carried out with the same soils from Azkoitia (Spain) and additionally with soils from Pola Osobowickie (Poland) for the taxonomic and functional characterization of the associated microbial community, in order to identify potential hydrocarbons degrading microorganisms.

DISCLAIMER

The SYMBIOREM project is 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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

1 Introduction

Work package 3 is dedicated to the development of bioremediation and recovery strategies for contaminated soils. In this context, UNIBO is responsible of performing genomic analysis of isolated strains

and metagenomic characterization of native soil microbial communities by shotgun sequencing using Illumina technology in order to predict potential strains for hydrocarbon degradation.

2 Aim and Scope – Strain identification

Sequencing of isolated bacterial strains from hydrocarbon contaminated soil was performed in order to characterize the culturable fraction in soil samples from Azkoitia site (Spain).

The microorganisms present in the contaminated soil capable of using hydrocarbons as a source of carbon and energy were isolated. To isolate these potential hydrocarbon-biodegrading strains, Azkoitia soil samples were incubated in a minimal salt liquid medium (Bushnell Haas Broth) with petroleum hydrocarbons (mineral oil) added as the sole carbon source. These samples were incubated for 7 days at 30°C in a primary incubation, and 4 successive re-incubations were performed in the same culture medium with the aim of selecting strains capable of growing with hydrocarbons as the sole carbon source. After the successive incubations in the liquid medium, the resulting samples were plated on Bushnell Haas agar medium with light mineral oil to isolate colonies.

Of all the isolated strains, 14 different strains were selected due to their different morphology and different metabolic profiles. **Table 1** show the colonial morphology of selected colonies.

Table 1 – Colonial morphology of selected colonies

ID	Colour	Morphology
CA-1	Whitish colonies	Bacillus bacteria
CA-2	White colonies	Actinomycetes-like bacteria, few colonies
CA-3	Yellow colonies	Motile rod bacteria
CA-5	White colonies	Bacillus bacteria
CA-6	White colonies	Cocci bacteria
CA-7	Yellow colonies	Rod bacteria, mimosas-like
CA-10	White colonies	Cocci-like bacteria
CA-13	Orange colonies	Cocci-like bacteria, very high density
CA-14	White colonies	Circular bacteria sometimes arranged in filaments
CA-15	White colonies	Bacillus bacteria
CA-16	White colonies	Motile rod bacteria
CA-19	White colonies	Cocci bacteria
CA-22	White colonies	Rod bacteria
CA-23	White colonies	Scattered cocci-like bacteria

The metabolic profiles were determined by biochemical tests. **Table 2** show the results of the biochemical tests of the selected strains:

Table 2 – Results of biochemical tests of selected strains from Azkoitia soil

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Biochemical tests	CA-1	CA-2	CA-3	CA-5	CA-6	CA-7	CA-10
2-nitrophenyl-ßD-galactopyranoside	-	+	-	-	-	-	-
L-Arginin	-	-	+/-	+	-	-	-
L-Lysin	-	-	+	-	-	-	-
L-Ornithin	-	-	-	+	-	-	-
Trisodium citrate	-	-	+	+	+	-	-
Sodium thiosulfate	-	-	-	-	-	-	-
Urea	-	-	+	+	-	-	+
Gelatin	+	+	+	-	-	+	-
D-glucose	-	-	-	-	-	-	-
D-mannitol	-	-	-	-	-	-	-
Inositol	-	-	-	-	-	-	-
D-sorbitol	-	-	-	-	-	-	-
L-Rhamnose	-	-	-	-	+	-	-
D-saccharose	-	-	-	-	-	-	-
D-melibiose	-	-	-	-	-	-	-
Amygdalin	-	-	-	-	-	-	-
L-arabinose	-	-	-	-	-	+	-
Oxidase	-	-	+	+	+	-	-
Catalase	+	-	-	+	-	-	+
Gram	-	+	-	-	-	-	-

Biochemical tests	CA-13	CA-14	CA-15	CA-16	CA-19	CA-22	CA-23
2-nitrophenyl-ßD-galactopyranoside	-	-	-	-	-	-	-
L-Arginin	-	-	-	+/-	-	+	-
L-Lysin	-	-	-	+/-	-	+	-
L-Ornithin	-	-	-	+/-	-	+	-
Trisodium citrate	-	-	-	+	+	+	+
Sodium thiosulfate	-	-	-	-	-	-	-
Urea	-	-	-	+/-	-	+	-
Gelatin	-	-	-	-	-	-	-
D-glucose	-	-	-	-	-	-	-
D-mannitol	-	-	-	-	-	-	-
Inositol	-	-	-	-	-	-	-
D-sorbitol	-	-	-	-	-	-	-
L-Rhamnose	-	-	-	-	+	-	-
D-saccharose	-	-	-	-	-	-	-
D-melibiose	-	-	-	-	-	-	-
Amygdalin	-	-	-	-	-	-	-
L-arabinose	-	-	-	-	+/-	-	+/-
Oxidase	-	+	+	+	+	+	+
Catalase	+	-	+	+	-	+	-
Gram	+	-	+	-	-	-	-

From this site, 14 isolated bacterial strains were processed for identification through Sanger sequencing of the 16S rRNA gene using the universal bacterial primers 27F and 1492R (Dos Santos et al., 2019). Sequencing outputs were then processed for taxonomic assignment using the nucleotide BLAST alignment tool (NCBI) (Altschul et al., 1990). **Table 3** shows the BLAST outputs for each isolated bacterial strain, considering both the forward and the reverse strands. Low-quality sequences with less than 100

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

high-quality nucleotides were removed from the analysis and indicated as “<100 nt”. Query coverage, e-value and percent identity are reported for the best BLAST hit. Where the alignment of the forward and reverse strands in the database resulted in different best hits, the alignment values of each strand on both best hits are reported.

For readability purposes, high-quality full sequences of isolated strains are reported in **Annex 1**.

Table 3 – Alignment outputs of sequenced strains from the Azkoitia site on the NCBI 16S rRNA non-redundant database using nucleotide BLAST. Low-quality sequences with less than 100 high-quality nucleotides were removed from the analysis and indicated as “<100 nt”.

Sample ID	Strand	BLAST best hit	Query coverage	E-value	% ID	Values for the other strand	Query coverage	E-value	% ID
CA-1	Forward (1F)	< 100 nt							
	Reverse (1R)	Pseudarthrobacter phenanthrenivorans Pseudarthrobacter siccitolerans Pseudarthrobacter polychromogenes	100%	0	100%				
CA-2	Forward (2F)	Nocardioides flavus Nocardioides luteus	100%	0	99.54%	For strand 2R	100%	0	98.89%
	Reverse (2R)	Nocardioides albus	100%	0	100.00%	For strand 2F	100%	0	99.23%
CA-3	Forward (3F)	[Pseudomonas] hibiscicola Stenotrophomonas geniculata Stenotrophomonas cyclobalanopsidis	100%	0	97.96%	For strand 3R	100%	0	95.54% 95.39%
		Stenotrophomonas lactitubi			97.97%				
	Reverse (3R)	Stenotrophomonas cyclobalanopsidis Stenotrophomonas lactitubi	100%	0	95.69%				
CA-5	Forward (4F)	Achromobacter pulmonis Achromobacter xylosoxidans Achromobacter denitrificans	100%	2.00E-113	99.12%	For strand 4R	96%	1.00E-89 1.00E-89 3.00E-90	97.91%
	Reverse (4R)	Achromobacter agilis	100%	2.00E-91	98.93%	For strand 4F	100%	3.00E-101	96.02%
CA-6	Forward (5F)	< 100 nt							
	Reverse (5R)	Labrys neptuniae	100%	0	99.44%				
CA-7	Forward (6F)	< 100 nt							
	Reverse (6R)	Luteibacter jiangsuensis	100%	0	99.57%				
CA-10	Forward (7F)	Rhodococcus erythropolis Rhodococcus erythropolis Nocardia coeliaca	100%	0	99.56%	For strand 7R	100%	0	98.71%
	Reverse (7R)	Rhodococcus qingshengii Rhodococcus qingshengii Rhodococcus erythropolis Nocardia coeliaca Rhodococcus qingshengii	100%	0	98.71%	For strand 7F For strand 7F	100% 100%	0 0	98.97% 98.97%
CA-13	Forward (8F)	Rhodococcus electrodiphilus	100%	0	100.00%				
	Reverse (8R)	Rhodococcus electrodiphilus	99%	0	97.84%				
CA-14	Forward (9F)	Inquilinus limosus	100%	0	98.58%	For strand 9R	100%	0	96.47%
	Reverse (9R)	Inquilinus ginsengisoli	100%	0	97.48%	For strand 9F	100%	0	98.10%
CA-15	Forward (10F)	Gordonia amicalis	100%	0	98.77%				
	Reverse (10R)	Gordonia amicalis	99%	0	99.07%				
CA-16	Forward (11F)	Achromobacter pulmonis	100%	0	100.00%	For strand 11R	100%	0	99.34%
	Reverse (11R)	Achromobacter veterisilvae	100%	2.00E-154	99.34%	For strand 11F	100%	2.00E-154	99.33%
CA-19	Forward (12F)	< 100 nt							
	Reverse (12R)	Labrys neptuniae	100%	0	98.64%				
CA-22	Forward (13F)	Achromobacter aegrifaciens	100%	0	99.35%	For strand 13R	100%	0	99.26%
	Reverse (13R)	Achromobacter dolens Achromobacter anxifer	100%	0	99.26%	For strand 13F	100%	0	98.06%
CA-23	Forward (14F)	< 100 nt							
	Reverse (14R)	Labrys neptuniae	100%	0	98.88%				

Among all the identified strains, 7 were selected based on their biosafety level (level 1) and their ability to biodegrade hydrocarbons as described scientific publications

- *Pseudarthrobacter phenanthrenivorans* is a Gram-positive aerobic bacterium able to degrade phenanthrene, a polycyclic aromatic hydrocarbon (PAH) consisting of three fused benzene rings, with a higher rate compared to other bacteria belonging to the *Pseudarthrobacter* genus (Tsagogiannis et al., 2021; Asimakoula et al., 2023).

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

- *Nocardioides luteus* is a Gram-positive bacterium whose strain BAFB has been shown to be highly efficient in the aerobic degradation of C8 to C11 normal and mono-branched alkane compounds (Brown et al., 2017). Other strains belonging to the genus *Nocardioides* have been shown to degrade phenanthrene (Saito et al., 2000), vinyl chloride (Coleman et al., 2002), hexachlorobenzene and chlorophenol, including 2,4-dichlorophenol, 2,4,5-trichlorophenol and pentachlorophenol (Ito et al., 2019), and benzo[a]pyrene (Wang et al., 2021).
- *Achromobacter pulmonis* strain HDK3 has been shown to degrade C14-C19 long hydrocarbons (Khalifa and Aldayel, 2018).
- *Rhodococcus erythropolis* showed the presence of *alkB* genes encoding the enzyme alkane 1-monooxygenase (Van der Geize and Dijkhuizen, 2004), being a potential biodegrader for n-alkanes. *R. erythropolis* has also been shown to degrade 2-chlorophenol, 2,4-dichlorophenol, catechol, p-chlorophenol, p-cresol and other phenolic compounds (Gorlatov and Golovleva, 1992; Goswami et al., 2005; Strnad et al., 2014).
- *Gordonia amicalis* has been reported to degrade the aliphatic hydrocarbon n-hexadecane and diesel oil (Sowani et al., 2020).
- *Achromobacter aegrifaciens* has been shown to degrade chrysene in oil contaminated seawater (Lazzem et al., 2023).
- *Luteibacter jiangsuensis* is Gram-negative, non-motile, rod-shaped bacterial strain capable of degrading methamidophos that *was isolated from* a methamidophos-manufacturing factory in China (Wang et al., 2011). Other study (Cul et al. 2020) demonstrated the capacity of petroleum hydrocarbons degradation of a bacterial community where the dominant bacterial genera are *Burkholderia-Paraburkholderia, Luteibacter, and Acinetobacter.*

3 Aim and Scope – Shotgun metagenomics analysis

Shotgun metagenomics was performed on pooled replicates of hydrocarbon-contaminated soils from the Azkoitia site (Spain) and the Pola Osobowickie site (Poland) for a deeper characterization of their microbial component, with the aim of identifying potential degrading microorganisms.

3.1 Metagenomics analysis: taxonomic characterization

From the Azkoitia site (Spain), 20 brownfield soil samples were collected in a former bronze foundry closed around 40 years ago in an industrial park, in February and May 2023, with 5 biological replicates per sampling. From the Pola Osobowickie site (Poland), 15 brownfield soil samples were collected from the area of former natural sewage treatment plant closed around 10 years ago in the Wroclaw city in winter, spring, and summer 2023, with 5 biological replicates per sampling. Total microbial DNA was extracted separately from each biological replicate from approximately 7-10 g of soil using the DNeasy PowerSoil Max kit (Qiagen, Hilden, Germany) for Azkoitia samples and from approximately 20-30 mg of soil using the DNeasy PowerSoil Pro kit (Qiagen, Hilden, Germany) for Pola Osobowickie samples. Extraction was performed according to the manufacturer's instructions with only minor adjustments. In the DNeasy PowerSoil Pro Kit, for the homogenization step all samples were homogenized using a FastPrep instrument (MP Biomedicals, Irvine, CA, U SA) and for the elution step samples were incubated for 5 min at 4°C before centrifugation; in the DNeasy PowerMax Soil Kit, in the steps of cell lysis, samples were incubated in a warm bath at 65°C for 40 minutes, shaking samples in falcon tubes every 10 minutes. For

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

both protocols the elution step was repeated twice in 70 µL and 5 ml, respectively. DNA quantification was performed using a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, United States). Biological replicates were then pooled to the same concentration to be processed for deep metagenomics shotgun sequencing (see **Table 4** for sample quantification, pooling and assigned IDs).

DNA libraries were prepared using the QIAseq FX DNA Library Kit (Qiagen) according to the manufacturer's instructions. Shortly, 100 ng of each DNA sample was fragmented to a 450-bp size, end-repaired, and A-tailed using the FX enzyme mix with the following thermal cycle: 4°C for 1 min, 32°C for 8 min, and 65°C for 30 min. Adapter ligation was performed by incubating DNA samples at 20°C for 15 min in the presence of DNA ligase and Illumina adapter barcodes. A first purification step with Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) was performed, followed by library amplification with a 10-cycle PCR and a further purification step. Samples were pooled at an equimolar concentration of 4 nM to obtain the final library. Sequencing was performed on an Illumina NextSeq platform using a 2 × 150-bp paired-end protocol, following the manufacturer's instructions (Illumina, San Diego, CA, USA). Raw sequences were quality filtered using fastp v0.23.4 (Chen et al., 2018) and assembly was performed with metaSPAdes v3.15.3 (Prjibelski et al., 2020). Raw and filtered sequencing outputs are shown in **Figure 1** and **Figure 2**, respectively, including million reads and mean quality score.

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table 4 – Summary table of extracted samples. DNA quantification through Nanodrop following microbial DNA extraction is reported together with samples IDs and pools for shotgun metagenomics analysis.

Site	Sample ID	Nanodrop conc. (ng/uL)	Pool ID	Sampling area and period
Azkoitia, Spain	7039-T3.1 SI-5	46.49	Sy5	Soil inside the foundry Feb. 2023
	7039-T3.1 SI-6	43.74		
	7039-T3.1 SI-7	44.74		
	7039-T3.1 SI-8	27.77		
	7039-T3.1 SI-9	11.61		
	7039-T3.1 SE-10	35.33	Sy6	Soil outside the foundry Feb. 2023
	7039-T3.1 SE-11	34.9		
	7039-T3.1 SE-12	42.45		
	7039-T3.1 SE-13	21.49		
	7039-T3.1 SE-14	36.57		
	7039-T3.1 SI-15	77.38	Sy7	Soil inside the foundry May 2023
	7039-T3.1 SI-16	75.87		
	7039-T3.1 SI-17	57.07		
	7039-T3.1 SI-18	62.49		
	7039-T3.1 SI-19	27.24		
	7039-T3.1 SE-20	49.98	Sy8	Soil outside the foundry May 2023
	7039-T3.1 SE-21	53.17		
	7039-T3.1 SE-22	37.99		
	7039-T3.1-SE-23	35.31		
	7039-T3.1 SE-24	64.42		
Pola Osobowickie, Poland	PO1	151.31	Sy1	Pola Osobowice Winter 2023
	PO2	132.75		
	PO3	143.38		
	PO4	161.45		
	PO5	148.63		
	PO1/2	160.42	Sy2	Pola Osobowice Spring 2023
	PO2/2	133.74		
	PO3/2	137.6		
	PO4/2	217.3		
	PO5/2	167.15		
	PO1/3	146.23	Sy3	Pola Osobowice Summer 2023
	PO2/3	182.08		
	PO4/3	201.37		
	PO3/3	136.25		
	PO5/3	214.86		

Sample	Direction	% dups	% GC	M reads	G bp
Sy1	Forward	1	58	33	4.95
	Reverse	1.2	57	33	4.95
Sy2	Forward	0.7	63	27.8	4.17
	Reverse	0.9	62	27.8	4.17
Sy3	Forward	0.7	63	27.9	4.19
	Reverse	0.8	63	27.9	4.19
Sy5	Forward	7	67	71.5	10.7
	Reverse	6.4	67	71.5	10.7
Sy6	Forward	3.4	65	74	11.1
	Reverse	4	66	74	11.1
Sy7	Forward	4	66	60.5	9.08
	Reverse	3.9	65	60.5	9.08
Sy8	Forward	4.7	65	69	10.4
	Reverse	4.4	65	69	10.4

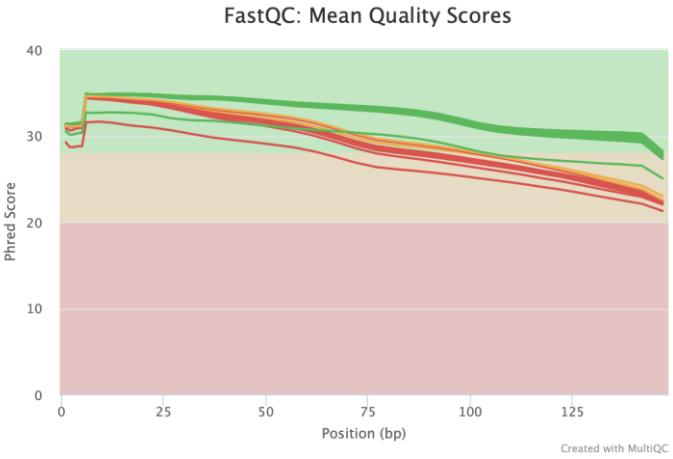

Figure 1 – Summary of the raw sequencing data. In the table on the left, we reported for each sample the percentage of duplicates (% dups), the percentage of GC content (% GC), the output in million reads (M reads) and Giga base pairs (G bp). In the figure on the right, we reported the mean quality score of the samples.

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Sample	Direction	% dups	% GC	M reads	G bp
Sy1	Forward	0.8	57	28.3	4.2
	Reverse	0.8	57	28.3	4.2
	Unpaired_1	0.4	63	4.0	0.6
	Unpaired_2	0.4	64	0.2	0.03
Sy2	Forward	0.5	62	23.4	3.5
	Reverse	0.5	62	23.4	3.5
	Unpaired_1	0.3	66	3.7	0.56
	Unpaired_2	0.4	67	0.3	0.05
Sy3	Forward	0.5	62	22.7	3.4
	Reverse	0.5	63	22.7	3.4
	Unpaired_1	0.4	67	4.5	0.68
	Unpaired_2	0.3	67	0.3	0.05
Sy5	Forward	5.6	66	55.9	8.4
	Reverse	5.7	67	55.9	8.4
	Unpaired_1	3	70	12.7	1.9
	Unpaired_2	1.6	71	0.8	0.12
Sy6	Forward	2.6	64	60.4	9.1
	Reverse	2.7	65	60.4	9.1
	Unpaired_1	1.6	68	10.8	1.6
	Unpaired_2	1	69	0.8	0.12
Sy7	Forward	3.1	65	49.7	7.5
	Reverse	3.2	65	49.7	7.5
	Unpaired_1	1.4	69	8.5	1.3
	Unpaired_2	1.1	70	0.7	0.11
Sy8	Forward	3.6	65	56.4	8.5
	Reverse	3.7	65	56.4	8.5
	Unpaired_1	2.4	68	10.0	1.5
	Unpaired_2	1.7	70	0.8	0.12

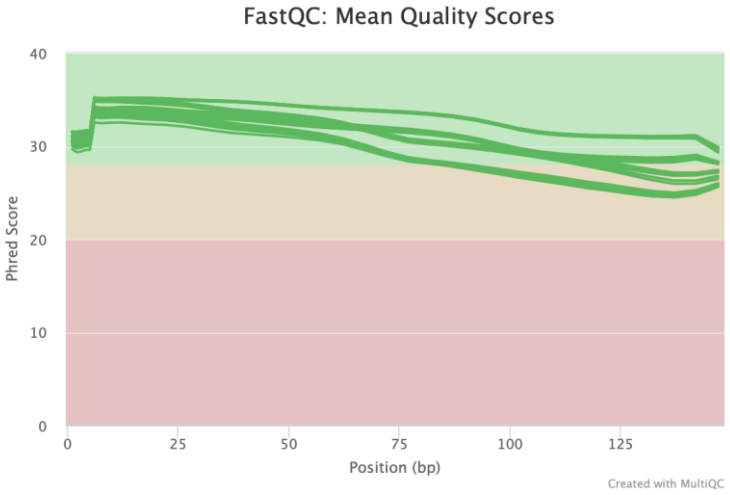

Figure 2 – Summary of the quality-filtered sequencing data. In the table on the left, we reported for each sample the percentage of duplicates (% dups), the percentage of GC content (% GC), the output in million reads (M reads) and Giga base pairs (G bp). In the figure on the right, we reported the mean quality score of the samples.

Assembled contigs outputs were processed through QUAST (QUality ASsessment Tool, v5.1.0rc1) (Gurevich et al., 2013) for basic statistics (**Table 5**). Contigs were then processed for metagenome-assembled genomes (MAGs) production using MaxBin 2.0 (Wu et al., 2016), MetaBAT 2 (Kang et al., 2019) and CONCOCT v1.1.0 (Alneberg et al., 2014) for binning and MetaWRAP v1.3.2 (Uritskiy et al., 2018) for bin refinement. Overall, 64 MAGs were retrieved, specifically 5 bins for Sy1 sample, 3 bins for Sy2, 2 bins for Sy3, 13 bins for Sy5, 15 bins for Sy6, 14 bins for Sy7 and 12 bins for Sy8. Only bins with >50% completeness and <5% contamination were selected for MAGs analysis, namely 4 bins for Sy1, 3 bins for Sy2, 1 bin for Sy3, 13 bins for Sy5, 15 bins for Sy6, 14 bins for Sy7 and 12 bins for Sy8. Taxonomic assignment of the retrieved high-quality MAGs was performed using GTDB-Tk v2.3.2 (Chaumeil et al., 2022) (**Table 6**).

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table 5 – Assembled contigs statistics based on QUAST. N50/N90 are the sequence length of the shortest contig at 50% and 90% of the total assembly length. L50/L90 are the count of the smallest number of contigs whose length sum makes up 50 or 90% of genome size.

Parameter	Sy1	Sy2	Sy3	Sy5	Sy6	Sy7	Sy8
# contigs (>=0 bp)	671480	517329	494924	845830	1279860	823139	1196694
# contigs (>= 1000 bp)	52587	27913	22985	97538	108017	77838	105082
# contigs (>= 5000 bp)	2241	785	1021	5727	4524	4251	4822
# contigs (>= 10000 bp)	828	257	181	1399	1447	1155	1476
# contigs (>= 25000 bp)	211	32	5	162	267	155	248
# contigs (>= 50000 bp)	34	3	0	17	33	20	48
Total length (>= 0 bp)	395467672	270484489	254309205	572009483	767874353	517363304	732872194
Total length (>= 1000 bp)	107764093	49847981	42902570	210542183	215029815	165271598	215616622
Total length (>= 5000 bp)	27126313	7993212	8045940	52328871	48037884	40936437	50491750
Total length (>= 10000 bp)	17640802	4410126	2438341	23365995	27182090	19887770	27883869
Total length (>= 25000 bp)	8186547	1070498	152738	5770652	10077943	5629228	10144437
Total length (>= 50000 bp)	2363357	175511	0	1000565	2434118	1243918	3510650
# contigs	238987	157283	143858	359630	483347	318316	459880
Largest contig	157721	63241	43501	84426	155187	84121	168190
Total length	231368898	134182080	121355070	385982149	463902383	325246171	451114754
GC (%)	59.85	63.7	64.14	66.54	64.99	65.68	64.79
N50	936	805	784	1106	933	1016	957
N90	550	536	534	564	549	556	551
L50	60780	47290	43290	80831	125553	75211	115249
L90	194768	131322	120338	286787	394695	256512	373814

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table 6 – Taxonomic assignment of retained bins with GTDBTK. Only bins with >50% completeness and <5% contamination were selected for MAGs analysis and taxonomically assigned.

MAGs	Classification
Sy1_bin.2	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__Bradyrhizobium;s__
Sy1_bin.3	d__Bacteria;p__Actinomycetota;c__Thermoleophilia;o__Solirubrobacterales;f__70-9;g__VAYN01;s__
Sy1_bin.4	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JAAYBP01;g__JAFDWI01;s__
Sy1_bin.5	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__REEB76;f__REEB76;g__;s__
Sy2_bin.1	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Mycobacteriales;f__Mycobacteriaceae;g__Mycobacterium;s__
Sy2_bin.2	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Micropepsales;f__Micropepsaceae;g__Rhizomicrobium;s__
Sy2_bin.3	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Puia;s__
Sy3_bin.2	d__Bacteria;p__Chloroflexota;c__Ktedonobacteria;o__Ktedonobacterales;f__JADMIN01;g__;s__
Sy5_bin.1	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__JADFDG01;g__JADFDG01;s__
Sy5_bin.10	d__Bacteria;p__Deinococcota;c__Deinococci;o__Deinococcales;f__;g__;s__
Sy5_bin.11	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Mycobacteriales;f__Mycobacteriaceae;g__Williamsia;s__
Sy5_bin.12	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Cellulomonadaceae;g__Cellulomonas;s
Sy5_bin.13	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Micrococcaceae;g__Paeniglutamicibacter;s
Sy5_bin.14	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Rhodocyclaceae;g__Rugosibacter;s__Rugosibacter sp002422995
Sy5_bin.2	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JACDCH01;g__;s__
Sy5_bin.3	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__Sva-07;s__
Sy5_bin.4	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JACDCH01;g__UBA9382;s__
Sy5_bin.6	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__;s__
Sy5_bin.7	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__;s__
Sy5_bin.8	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Mycobacteriales;f__Pseudonocardiaceae;g__Saccharopolyspora;s__Saccharopolyspora dendranthemae
Sy5_bin.9	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JACDCH01;g__VFJN01;s__
Sy6_bin.1	d__Bacteria;p__Patescibacteria;c__Saccharimonadia;o__Saccharimonadales;f__JAGQNK01;g__JAGQNK01;s__
Sy6_bin.10	d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__UBA2999;g__;s__
Sy6_bin.11	d__Bacteria;p__Myxococcota_A;c__UBA9160;o__SZUA-336;f__SZUA-336;g__VGRW01;s__
Sy6_bin.12	d__Bacteria;p__Zixibacteria;c__MSB-5A5;o__GN15;f__FEB-12;g__FEB-12;s__
Sy6_bin.13	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Acidiferrobacterales;f__Sulfurifustaceae;g__JACQFZ01;s__
Sy6_bin.15	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__Porticoccaceae;g__JAGPUQ01;s__
Sy6_bin.16	d__Bacteria;p__Desulfobacterota_B;c__Binatia;o__Bin18;f__Bin18;g__JABFSC01;s__
Sy6_bin.2	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__JADFDG01;g__JADFDG01;s__
Sy6_bin.3	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__CADCSY01;g__JACOTK01;s__
Sy6_bin.4	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__UBA5794;f__ZC4RG35;g__;s__
Sy6_bin.5	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__JADFDG01;g__JADFDG01;s__
Sy6_bin.6	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Gemmatimonadales;f__GWC2-71-9;g__;s__
Sy6_bin.7	d__Bacteria;p__Patescibacteria;c__Saccharimonadia;o__Saccharimonadales;f__JAGQNK01;g__JAGQNK01;s__
Sy6_bin.8	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA5335;f__UBA5335;g__Macondimonas;s__
Sy6_bin.9	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JACDCH01;g__VFJN01;s__
Sy7_bin.1	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JACDCH01;g__VFJN01;s__
Sy7_bin.10	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__;s__
Sy7_bin.11	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Rhodoglobus;s__
Sy7_bin.12	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__;s__
Sy7_bin.13	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__LMDS01;s__
Sy7_bin.14	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JACDCH01;g__UBA9382;s__
Sy7_bin.2	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Microbacterium;s__
Sy7_bin.3	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Agromyces;s__
Sy7_bin.4	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__Sva-07;s__
Sy7_bin.5	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__JADFDG01;g__JADFDG01;s__
Sy7_bin.6	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__JACDCH01;g__;s__
Sy7_bin.7	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__UBA5794;f__ZC4RG35;g__;s__
Sy7_bin.8	d__Bacteria;p__Deinococcota;c__Deinococci;o__Deinococcales;f__;g__;s__
Sy7_bin.9	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Parvibaculales;f__Parvibaculaceae;g__Parvibaculum;s__
Sy8_bin.1	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__JADFDG01;g__JADFDG01;s__
Sy8_bin.10	d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__UBA2999;g__;s__
Sy8_bin.11	d__Bacteria;p__Myxococcota_A;c__UBA9160;o__SZUA-336;f__SZUA-336;g__VGRW01;s__
Sy8_bin.12	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__CADEEN01;s__
Sy8_bin.2	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__CADCSY01;g__JACOTK01;s__
Sy8_bin.3	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__JADFDG01;g__JADFDG01;s__
Sy8_bin.4	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__Porticoccaceae;g__JAGPUQ01;s__
Sy8_bin.5	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__UBA5794;f__ZC4RG35;g__;s__
Sy8_bin.6	d__Bacteria;p__Patescibacteria;c__Saccharimonadia;o__Saccharimonadales;f__JAGQNK01;g__JAGQNK01;s__
Sy8_bin.7	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Acidiferrobacterales;f__Sulfurifustaceae;g__JACQFZ01;s__
Sy8_bin.8	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA5335;f__UBA5335;g__Macondimonas;s__
Sy8_bin.9	d__Bacteria;p__Zixibacteria;c__MSB-5A5;o__GN15;f__FEB-12;g__FEB-12;s__

3.2 Metagenomics analysis: functional characterization

High-quality retrieved MAGs from Azkoitia and Pola Osobowickie sites were processed through MicrobeAnnotator v2.0.5 (Ruiz-Perez et al., 2021) for prediction of metabolic functions. **Figure 3** shows the metabolic module category from the KEGG pathway database (Kanehisa and Goto, 2000) for each identified MAG for modules with >50% completeness in at least one genome.

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

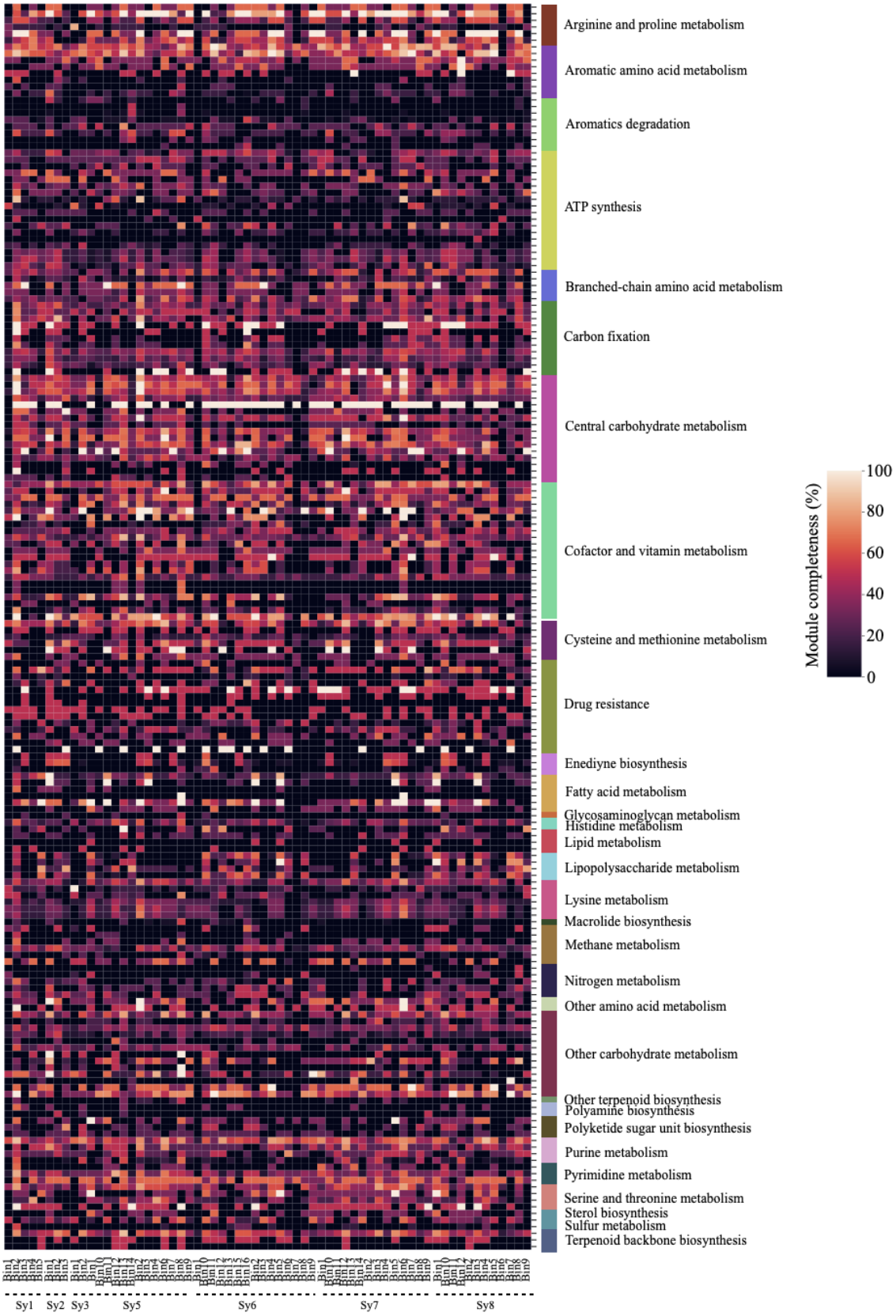

Figure 3 – Metabolic module category from the KEGG pathway database for each identified MAG for modules with >50% completeness. Bins and corresponding samples are indicated at the bottom of the heatmap, KEGG pathways are indicated by a color-coded bar on the right. The legend for module completeness is shown on the right.

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

In particular, there are 19 different KEGG modules associated with hydrocarbons degradation. Three pathways are involved in PAH degradation, namely M00624 (terephthalate degradation), and M00623 and M00636 (phthalate degradation), whereas 16 are involved in aromatic hydrocarbons degradation, namely M00538 (toluene degradation), M00537 (xylene degradation), M00419 (cymene degradation), M00547 (benzene/toluene degradation), M00548 (benzene degradation), M00551 and M00540 (benzoate degradation), M00568 (catechol ortho-cleavage), M00569 (catechol meta-cleavage), M00539 (cumate degradation), M00543 (biphenyl degradation), M00544 (carbazole degradation), M00418 (toluene degradation, anaerobic), M00541 (benzoyl-CoA degradation), M00534 (naphthalene degradation), M00545 (trans-cinnamate degradation). These modules all belong to the xenobiotic degradation pathways, specifically aromatics degradation (in green on the heatmap of **Figure 3**). **Table 7** lists all KEGG modules associated with aromatics degradation. Amongst these, 8 modules are >50% complete in at least one genome (highlighted in green), with 6 of them belonging to the pathways of interest. Despite having a completeness <50% in at least one genome, the 3 modules associated with PAH degradation and 9 of the 10 remaining modules associated with aromatics degradation are represented in the majority of the MAGs.

Table 7 – KEGG modules associated with aromatics degradation. Eight modules have >50% completeness (green columns). The 3 pathways associated with PAH degradation and 9 of the 10 remaining modules for aromatics degradation (in bold) have <50% completeness**.**

Pathway group		Aromatics degradation																				
Name		Toluene degradation, anaerobic, toluene => benzoyl-CoA	Cymene degradation, p-cymene => p-cumate	Naphthalene degradation, naphthalene => salicylate	Xylene degradation, xylene => methylbenzoate	Toluene degradation, toluene => benzoate	Cumate degradation, p-cumate => 2-oxopent-4-enoate + 2-methylpropanoate	Benzoate degradation, cyclohexanecarboxylic acid =>pimeloyl-CoA	Benzoyl-CoA degradation, benzoyl-CoA => 3-hydroxypimeloyl-CoA	Biphenyl degradation, biphenyl => 2-oxopent-4-enoate + benzoate	Carbazole degradation, carbazole => 2-oxopent-4-enoate + anthranilate	Benzene--toluene degradation, benzene => catechol -- toluene => 3-methylcatechol	Benzene degradation, benzene => catechol	Benzoate degradation benzoate => catechol -- methylbenzoate => methylcatechol	Catechol ortho-cleavage, catechol => 3-oxoadipate	Catechol meta-cleavage, catechol => acetyl-CoA + 4-methylcatechol => propanoyl-CoA	**Phthalate degradation 1, phthalate => protocatechuate**	**Terephthalate degradation, terephthalate => 3,4-dihydroxybenzoate**	**Phthalate degradation 2, phthalate => protocatechuate**	Anthranilate degradation, anthranilate => catechol	Salicylate degradation, salicylate => gentisate	Phenylacetate degradation, phenylacetate => acetyl-CoA--succinyl-CoA
Module		**M00418**	**M00419**	**M00534**	**M00537**	**M00538**	**M00539**	**M00540**	**M00541**	**M00543**	**M00544**	**M00547**	**M00548**	**M00551**	**M00568**	**M00569**	**M00623**	**M00624**	**M00636**	M00637	M00638	M00878
Sy1	Bin1	**16.67**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin2	**8.33**	**0**	**20.83**	**0**	**0**	**0**	**20**	**0**	**25**	**0**	**0**	**0**	**0**	**25**	**60**	**33.33**	**0**	**0**	**0**	**0**	45.71
	Bin3	**8.33**	**0**	**16.67**	**0**	**0**	**0**	**40**	**25**	**25**	**0**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin4	**0**	**0**	**0**	**0**	**0**	**0**	**40**	**25**	**0**	**0**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin5	**0**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**6.25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	0
Sy2	Bin1	**8.33**	**50**	**4.17**	**0**	**5.56**	**20**	**20**	**0**	**25**	**0**	**0**	**0**	**0**	**25**	**60**	**0**	**16.67**	**8.33**	**0**	**0**	28.57
	Bin2	**8.33**	**16.67**	**16.67**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**16.67**	**0**	**0**	**0**	17.14
	Bin3	**8.33**	**0**	**0**	**0**	**0**	**5**	**20**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	22.86
Sy3	Bin1	**8.33**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**11.11**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin2	**0**	**16.67**	**0**	**0**	**0**	**0**	**40**	**0**	**31.25**	**0**	**0**	**0**	**0**	**50**	**20**	**0**	**0**	**0**	**0**	**0**	28.57
Sy5	Bin1	**8.33**	**33.33**	**16.67**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	0
	Bin10	**0**	**33.33**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**50**	**0**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin11	**0**	**16.67**	**0**	**0**	**5.56**	**25**	**20**	**0**	**6.25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**16.67**	**0**	**0**	**0**	42.86
	Bin12	**0**	**0**	**0**	**0**	**0**	**5**	**0**	**0**	**6.25**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0
	Bin13	**0**	**0**	**4.17**	**0**	**5.56**	**0**	**20**	**25**	**6.25**	**11.11**	**0**	**0**	**0**	**75**	**40**	**0**	**0**	**0**	**33.33**	**0**	42.86
	Bin14	**16.67**	**16.67**	**20.83**	**50**	**50**	**20**	**20**	**0**	**31.25**	**11.11**	**12.5**	**0**	**16.67**	**0**	**60**	**0**	**16.67**	**8.33**	**50**	**0**	28.57
	Bin2	**8.33**	**0**	**0**	**0**	**5.56**	**5**	**20**	**25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	42.86
	Bin3	**0**	**0**	**0**	**0**	**0**	**5**	**20**	**25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin4	**8.33**	**0**	**0**	**0**	**0**	**0**	**20**	**25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	0
	Bin6	**16.67**	**0**	**4.17**	**0**	**0**	**0**	**40**	**0**	**6.25**	**0**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**25**	0
	Bin7	**8.33**	**0**	**0**	**0**	**0**	**20**	**20**	**25**	**6.25**	**0**	**0**	**0**	**0**	**25**	**60**	**0**	**0**	**0**	**33.33**	**0**	0
	Bin8	**8.33**	**0**	**0**	**33.33**	**38.89**	**0**	**20**	**18.75**	**6.25**	**0**	**0**	**0**	**0**	**25**	**60**	**0**	**0**	**0**	**0**	**0**	45.71
	Bin9	**0**	**0**	**0**	**0**	**0**	**10**	**20**	**25**	**0**	**11.11**	**0**	**0**	**0**	**50**	**0**	**0**	**0**	**0**	**33.33**	**0**	28.57
Sy6	Bin1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0
	Bin10	**8.33**	**0**	**0**	**0**	**5.56**	**0**	**20**	**37.5**	**0**	**11.11**	**0**	**0**	**0**	**50**	**40**	**0**	**0**	**33.33**	**33.33**	**0**	14.29
	Bin11	**25**	**0**	**20.83**	**0**	**0**	**0**	**20**	**0**	**31.25**	**0**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**25**	**0**	31.43
	Bin12	**8.33**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**6.25**	**0**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin13	**8.33**	**16.67**	**16.67**	**0**	**0**	**5**	**20**	**25**	**6.25**	**0**	**0**	**0**	**0**	**25**	**20**	**33.33**	**0**	**0**	**0**	**0**	28.57
	Bin15	**16.67**	**0**	**8.33**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	0
	Bin16	**41.67**	**0**	**8.33**	**0**	**0**	**5**	**20**	**0**	**56.25**	**0**	**12.5**	**0**	**0**	**0**	**0**	**0**	**0**	**8.33**	**25**	**50**	17.14
	Bin2	**8.33**	**16.67**	**16.67**	**0**	**0**	**0**	**20**	**0**	**31.25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	0
	Bin3	**8.33**	**0**	**0**	**0**	**0**	**5**	**20**	**0**	**6.25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin4	**8.33**	**0**	**0**	**0**	**0**	**0**	**20**	**25**	**6.25**	**0**	**0**	**0**	**0**	**50**	**20**	**0**	**0**	**0**	**0**	**0**	0
	Bin5	**0**	**33.33**	**20.83**	**0**	**0**	**0**	**0**	**25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**25**	0
	Bin6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin7	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	0
	Bin8	**0**	**16.67**	**16.67**	**16.67**	**5.56**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	0
	Bin9	**8.33**	**0**	**0**	**0**	**0**	**5**	**20**	**25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	0
Sy7	Bin1	**0**	**0**	**0**	**0**	**0**	**0**	**40**	**25**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin10	**8.33**	**0**	**0**	**0**	**0**	**20**	**40**	**25**	**6.25**	**11.11**	**0**	**0**	**0**	**25**	**60**	**0**	**16.67**	**0**	**33.33**	**0**	0
	Bin11	**0**	**0**	**0**	**0**	**0**	**5**	**0**	**0**	**0**	**11.11**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	0
	Bin12	**8.33**	**0**	**4.17**	**0**	**5.56**	**0**	**20**	**0**	**6.25**	**0**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**25**	0
	Bin13	**16.67**	**16.67**	**16.67**	**0**	**0**	**0**	**20**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**25**	**0**	0
	Bin14	**8.33**	**0**	**4.17**	**0**	**0**	**20**	**20**	**0**	**31.25**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	28.57
	Bin2	**16.67**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**6.25**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	0
	Bin3	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**50**	**40**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin4	**0**	**0**	**0**	**0**	**0**	**5**	**20**	**25**	**0**	**0**	**0**	**0**	**0**	**25**	**40**	**0**	**0**	**0**	**0**	**0**	28.57
	Bin5	**0**	**33.33**	**16.67**	**0**	**0**	**0**	**0**	**0**	**6.25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**25**	**25**	0
	Bin6	**8.33**	**0**	**0**	**0**	**5.56**	**5**	**20**	**25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	28.57
	Bin7	**8.33**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**6.25**	**0**	**0**	**0**	**16.67**	**50**	**20**	**0**	**0**	**0**	**0**	**0**	0
	Bin8	**8.33**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	**0**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin9	**8.33**	**0**	**16.67**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**40**	**0**	**0**	**0**	**0**	**0**	28.57
Sy8	Bin1	**0**	**33.33**	**16.67**	**0**	**0**	**0**	**0**	**25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	0
	Bin10	**8.33**	**0**	**0**	**0**	**5.56**	**0**	**0**	**31.25**	**0**	**11.11**	**0**	**0**	**0**	**25**	**40**	**0**	**0**	**33.33**	**33.33**	**0**	42.86
	Bin11	**8.33**	**0**	**20.83**	**0**	**0**	**5**	**20**	**0**	**31.25**	**0**	**0**	**0**	**0**	**0**	**40**	**0**	**0**	**0**	**25**	**0**	31.43
	Bin12	**8.33**	**0**	**16.67**	**0**	**0**	**0**	**40**	**25**	**0**	**0**	**0**	**0**	**0**	**25**	**40**	**0**	**0**	**33.33**	**0**	**0**	14.29
	Bin2	**8.33**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**6.25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	14.29
	Bin3	**8.33**	**0**	**20.83**	**0**	**0**	**0**	**0**	**0**	**31.25**	**0**	**0**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**25**	0
	Bin4	**33.33**	**0**	**8.33**	**0**	**0**	**0**	**20**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	0
	Bin5	**16.67**	**0**	**0**	**0**	**0**	**0**	**20**	**25**	**6.25**	**0**	**0**	**0**	**0**	**50**	**20**	**0**	**0**	**0**	**0**	**0**	0
	Bin6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0
	Bin7	**8.33**	**0**	**0**	**0**	**0**	**10**	**20**	**25**	**6.25**	**0**	**0**	**0**	**16.67**	**50**	**20**	**33.33**	**0**	**0**	**0**	**0**	14.29
	Bin8	**8.33**	**16.67**	**4.17**	**16.67**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**25**	0
	Bin9	**8.33**	**0**	**0**	**0**	**5.56**	**0**	**0**	**0**	**6.25**	**0**	**0**	**0**	**0**	**25**	**20**	**0**	**0**	**0**	**0**	**0**	14.29

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

4 CONCLUSIONS

Among all bacterial isolates, we identified 6 strains potentially capable of degrading hydrocarbons according to previous literature results, namely *Pseudarthrobacter phenanthrenivorans*, *Nocardioides luteus*, *Achromobacter pulmonis*, *Rhodococcus erythropolis*, *Gordonia amicalis* and *Achromobacter aegrifaciens* from the Azkoitia site.

As for new potential hydrocarbons degraders identified by shotgun metagenomics analysis, we found that all identified genomes carried at least one potential degrading function, regardless of the level of completeness of the module, except for Sy6 bin1 and Sy8 bin6, with promising implications in the potential of the brownfield soil associated microbial communities for hydrocarbons degradation in both selected sites.

5 DATA AVAILABILITY

All sequences belonging to the 14 isolated bacterial strains that were processed for identification through Sanger sequencing of the 16S rRNA gene using the universal bacterial primers 27F and 1492R were deposited in the **Annex 1** attached to this deliverable D3.1. Sequences produced via shotgun metagenomics of hydrocarbon-contaminated soils from the Azkoitia site (Spain) and the Pola Osobowickie site (Poland) were deposited in the public repository ENA (European Nucleotide Archive) under the project number **PRJEB75648**.

6 REFERENCES

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., ... & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nature methods, 11(11), 1144-1146.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Asimakoula S, Marinakos O, Tsagogiannis E, Koukkou AI. Phenol Degradation by *Pseudarthrobacter phenanthrenivorans* Sphe3. Microorganisms. 2023;11(2):524.

Brown LM, Gunasekera TS, Ruiz ON. Draft Genome Sequence of *Nocardioides luteus* Strain BAFB, an Alkane-Degrading Bacterium Isolated from JP-7-Polluted Soil. Genome Announc. 2017;5(4):e01529-16.

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics, 38(23), 5315-5316.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), i884-i890.

Coleman NV, Mattes TE, Gossett JM, Spain JC. Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. Appl Environ Microbiol. 2002;68(12):6162-71.

Cul J, Chen H, Sun M& Wen J (2020). Comparison of bacterial community structure and function under different petroleum hydrocarbon degradation conditions. Bioprocess and Biosystems Engineering 43:303–313.

Dos Santos, H. R. M., Argolo, C. S., Argôlo-Filho, R. C., & Loguercio, L. L. (2019). A 16S rDNA PCR-based theoretical to actual delta approach on culturable mock communities revealed severe losses of diversity information. BMC microbiology, 19(1), 1-14.

Gorlatov SN, Golovleva LA. Effect of cosubstrates on the dechlorination of selected chlorophenolic compounds by *Rhodococcus erythropolis* 1CP. J Basic Microbiol. 1992;32(3):177-84.

Goswami M, Shivaraman N, Singh RP. Microbial metabolism of 2-chlorophenol, phenol and rho-cresol by *Rhodococcus erythropolis* M1 in co-culture with *Pseudomonas fluorescens* P1. Microbiol Res. 2005;160(2):101-9.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075.

Ito K, Takagi K, Kataoka R, Kiyota H, Iwasaki A. Dissipation, dehalogenation, and denitration of chloroaromatic compounds by *Nocardioides* sp. strain PD653: Characterization of the substrate specificity. J Pestic Sci. 2019;44(3):171-176.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ, 7, e7359.

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Khalifa A, Aldayel MF. Metabolic diversity of the diesel-oil-degrading bacterium, *Achromobacter pulmonis* HDK3, obtained from Eastern region of Saudi Arabia. Asian journal of microbiology Biotechnology and environmental Sciences. 2018;20(3):778-785.

Lazzem A, Lekired A, Ouzari HI, Landoulsi A, Chatti A, El May A. Isolation and characterization of a newly chrysene-degrading *Achromobacter aegrifaciens*. Int Microbiol. 2023.

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes de novo assembler. Current protocols in bioinformatics, 70(1), e102.

Ruiz-Perez, C. A., Conrad, R. E., & Konstantinidis, K. T. (2021). MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. BMC bioinformatics, 22, 1-16.

Saito A, Iwabuchi T, Harayama S. A novel phenanthrene dioxygenase from *Nocardioides* sp. Strain KP7: expression in *Escherichia coli*. J Bacteriol. 2000;182(8):2134-41.

Sowani H, Kulkarni M, Zinjarde S. Uptake and detoxification of diesel oil by a tropical soil Actinomycete *Gordonia amicalis* HS-11: Cellular responses and degradation perspectives. Environ Pollut. 2020;263(Pt B):114538.

Strnad H, Patek M, Fousek J, Szokol J, Ulbrich P, Nesvera J, Paces V, Vlcek C. Genome Sequence of *Rhodococcus erythropolis* Strain CCM2595, a Phenol Derivative-Degrading Bacterium. Genome Announc. 2014;2(2):e00208-14.

Tsagogiannis E, Vandera E, Primikyri A, Asimakoula S, Tzakos AG, Gerothanassis IP, Koukkou AI. Characterization of Protocatechuate 4,5-Dioxygenase from *Pseudarthrobacter phenanthrenivorans* Sphe3 and In Situ Reaction Monitoring in the NMR Tube. Int J Mol Sci. 2021;22(17):9647.

Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome, 6(1), 1-13.

Van der Geize R, Dijkhuizen L. Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications, Current Opinion in Microbiology. 2004;7(3):255-261.

Wang B, Teng Y, Yao H, Christie P. Detection of functional microorganisms in benzene [a] pyrene-contaminated soils using DNA-SIP technology. J Hazard Mater. 2021;407:124788.

Wang L, Wang G, Li S &Jiang J (2011). *Luteibacter jiangsuensis* sp. nov.: A Methamidophos-Degrading Bacterium Isolated from a Methamidophos-Manufacturing Factory. Curr Microbiol 62:289–295.

Wu, Y. W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics, 32(4), 605-607.

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

ANNEX 1 – HIGH-QUALITY FULL SEQUENCES OF ISOLATED STRAINS.

- *CA-1 reverse:*

ACAAGGGGTTAGGCCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAG
CGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAATCCGAACTGAGACCGGCTTTTTGGGATTAGC
TCCACCTCACAGTATCGCAACCCTTTGTACCGGCCATTGTAGCATGCGTGAAGCCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCC
CACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCCCCACCATCACGTGCTGGCAACATAGAACGAGGGTTGCGCTCGTTGCGGG
ACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTAAACCGACCGCAAGCGGGGCACCTG

- *CA-2 forward:*

TGCAGTCGAGCGGAAGGCCCTTCGGGGTACTCGAGCGGCGAACGGGTGAGTAACACGTGAGTAATCTGCCCCAGGCTCTGGGATAGCCA
CCGGAAACGGTGATTAATACCGGATACGACAACCGATTGCATGATCTGGTTGTGGAAAGTTTTTCGGCCTGGGATGTGCTCGCGGCCTATC
AGCTTGTTGGTGAGGTAATGGCTCACCAAGGCTTCGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCC
CAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGGAAGCCTGATCCAGCAACGCCGCGTGAGGGATGACGGCCTTCG
GGTTGTAAACCTCTTTCAGCACAGACGAAGCGCAAGTGACGGTATGTGCAGAAGAAGGACCGGCCAACTACGTGCCAGCAGCCGCGGTAA
TACGTAGGGTCCGAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTCTGTCGCGTCGGGAGTGAAAACCAGGTGCTTAAC
ACCTGGCCTGCTTTCGATACGGCCAGACTAGAGGTACTCAGGGGAGAATGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAG
GAACACCGGTGGCGAA

- *CA-2 reverse:*

GGGCCACTGGCTTCGGGTGTTGCCGACTTTCGTGACGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGAT
CTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAATCCGAACTGAGACCGGCTTTTTGGGATTCGCTCACCCTCAC
GGGATCGCAGCCCTTTGTACCGGCCATTGTAGCATGCGTGAAGCCCTGGACATAAGGGGCATGATGACTTGACGTCATCCCCCACCTTCCTC
CGAGTTGACCCCGGCAGTCTCCTATGAGTCCCCAACCGAATTGCTGGCAACATAGGACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAA
CATCTCACGACACGAGCTGACGACAGCCATGCACCACCTGTACACCGACTAAAAGGGGCCGTATCTCTACGGCTTTCCGGTGTATGTC

- *CA-3 forward:*

TGCAGTCGACGGCAGCACAGAGGAGCTTGCTCCTTGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGAATCTACTTTTTCGTGGGG
GATAACGTAGGGAAACTTACGCTAATACCGCATACGACCTACGGGTGAAAGCAGGGGACCTTCGGGCCTTGCGCGATTGAATGAGCCGAT
GTCGGATTAGCTAGTTGGCGGGGTAAAGGCCCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGAACTGAGA
CACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATACCGCGTGGGTGAAGAAG
GCCTTCGGGTTGTAAAGCCCTTTTGTTGGGAAAGAAATCCAGCCGGCTAATACCTGGTTGGGATGACGGTACCCAAAGAATAAGCACCGGC
TAACTTCGTGCCACCAGCCGCGGTAATACGAAGGGTGC

- *CA-3 reverse:*

AGCGCCTCCGAGTTAGCTACCTGCTTCTGGTGCAACAAACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCG
CAGCAATGCTGATCTGCGATTACTAGCGATTCCGACTTCATGGAGTCGAGTTGCAGACTCCAATCCGGACTGAGATAGGGTTTCTGGGATT
GGCTTACCGTCGCCGGCTTGCAGCCCTCTGTCCCTACCATTGTAGTACGTGTGTAGCCCTGGCCGTAAGGGCCATGATGACTTGACGTCAT
CCCCACCTTCCTCCGGTTTGTCACCGGCGGTCTCCTTAGAGTTCCCACCATTACGTGCTGGCAACTAAGGACAAGGGTTGCGCTCGTTGCG
GGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTGTTCGAGTTCCCGAAGGCACCAATCCATCTCTGGAAA
GTTCTCGACATGTCAAGGTCTAGTTACGGTTCTTCGCGTTGCATCGAATTAAACCACATACTCCACCGCTTGTGCGGGGCCCCGTCAATTCC
TCTCCACCTCCCGCCTTGCGACGTACTCCCCCAGCCGGGGAACTTAACGCGTTAGCTTCCATACTGGTTGCCAAATTGCACCCAACATCCA
GTCGCATCGGTTAGGGCGGGGAATACCAG

- *CA-5 forward:*

TGCAGTCGACGGCAGCACGGACTTCGGTCTGGTGGCGAGTGGCGAACGGGTGAGTAATGTATCGGAACGTGCCCAGTAGCGGGGGATAA
CTACGCGAAAGCGTAGCTAATACCGCATACGCCCTACGGGGGAAAGCAGGGGATCGCAAGACCTTGCACTATTGGAGCGGCCGATATCGG
ATTAGCTAGTTGGTGGGGTAACGGCTCACCAAGGCGACGATCCGTA

- *CA-5 reverse:*

ATCGCCTCCTTGCGGTTAGGCTAACTACTTCTGGTAAACCCACTCCCATGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCAC
CGCGACATGCTGATCCGCGATTACTAGCGATTCCGACTTCACGCAGTCGAGTTGCAGACTGCGATCCGGACTACGATCGGGTTTCTGGGAT
TGG

- *CA-6 reverse:*

CTGCCTCCTTGCGGTTGGCGCAGCGCCTTCGGGTAACCAACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACC
GCAGCATGCTGATCTGCGATTACTAGCGATTCCACCTTCATGCACTCGAGTTGCAGAGTGCAATCTGAACTGAGACGGCTTTTGGAGATTTG
CTAGGGGTCACCCCTTCGCTTCCCACTGTCACCGCCATTGTAGCACGTGTGTAGCCCAGCCCGTAAGGGCCATGAGGACTTGACGTCATC
CCCACCTTCCTCGCGGCTTATCACCGGCAGTCCCCCTAGAGTGCCCAACTTAATGCTGGCAACTAGGGGCGAGGGTTGCGCTCGTTGCGG
GACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTCTCCGGTCCAGCCGAACTGAAGGATCTCATCTCTGTGA
TCCGCGACCGGGATGTCAAGGGCTGTAAGGTTCTGCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGG

- *CA-7 reverse:*

TTGCGGTTAGACTAACGGCTTCTGGAGCAGCTCACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGCA
TAGCTGATCTGCGATTACTAGCGATTCCGACTTCATGGAGTCGAGTTGCAGACTCCAATCCGGACTGGGATCGGCTTTCTGGGATTAGCTC
CACCTCGCGGTCTTGCAACCCTCTGTACCGACCATTGTAGTACGTGTGTAGCCCTGGCCGTAAGGGCCATGATGACTTGACGTCATCCCCA
CCTTCCTCCGGTTTGTCACCGGCAGTCTCCTTAGAGTTCCCGACTTTACTCGCTGGCAACTAAGGACAAGGGTTGCGCTCGTTGCGGGACT

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

TAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTGTTCCGATTCCCGAAGGCACTCCCGCATCTCTGCAGGATTCC
GGACATG

- *CA-10 forward:*

TGCAGTCGAGCGGTAAGGCCTTTCGGGGTACACGAGCGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCTGCACTTCGGGATAAGC
CTGGGAAACTGGGTCTAATACCGGATATGACCTCCTGTTGCATGACTTGGGGTGGAAAGATTTATCGGTGCAGGATGGGCCCGCGGCCTAT
CAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGACCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGC
CCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTC
GGGTTGTAAACCTCTTTCAGCAGGGACGAAGCGCAAGTGACGGTACCTGCAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTA
ATACGTAGGGTGCAAGCGTTGTCCGGAATTACTGGGCGTAAAGAGTTCGTAGGCGGTTTGTCGCGTCGTTTGTGAAAACCAGCAGCTCAAC
TGCTGGCTTGCAGGCGATACGGGCAGACTTGAGTACTGCAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAG
GAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTGACGCTGAGGAA

- *CA-10 reverse:*

GGCTCCTCCACAGGGTTAAGCCACCGGCTTCGGGTGTTACCGACTTTCATGACGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTC
ACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCACGGGGTCGAGTTGCAGACCCCGATCCGAACTGAGACCAGCTTTAAG
GGATTCGCTCCACCTCACGGTCTCGCAGCCCTCTGTACTGGCCATTGTAGCATGTGTGAAGCCCTGGACATAAGGGGCATGATGACTTGAC
GTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCTTACGAGTCCCCACCATAACGTGCTGGCAACATAAGATAGGGGTTGCGCTCG
TTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCACCACCTGTATACCGACCACAAGGGGGGCCACATCTCTGC
AGCTTTCCGGTATATGTCATATCCAGGTAAGGTTCTTCGCGTTGCATCGAATTACTCCACATGCTCCGCCGCTTGTGCGGGCCCCCG

- *CA-13 forward:*

TGCAGTCGACGATGAAGCCCAGCTTGCTGGGTGGATTAGTGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCTGCACTTCGGGATAA
GCCTGGGAAACTGGGTCTAATACCGGATAGGACCTCGGGATGCATGTTCCGGGGTGGAAAGGTTTTCCGGTGCAGGATGGGCCCGCGGC
CTATCAGCTTGTTGGTGGGGTAACGGCCCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACA
CGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGG
CCTTCGGGTTGTAAACCTCTTTCAGTACCGACGAAGCGCAAGTGACGGTAGGTACAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCGTCTGTGAAAACCCGCAG
CTCAACTGCGGGCTTGCAGGCGATACGGGCAGACTTGAGTACTGCAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATAT

- *CA-13 reverse:*

TCCTCCACGAGGGGTTAGGCCACCGGCTTCGGGTGTTACCGACTTTCATGACGTGACGGGCGCGTGTGTACAGGCCCGGGACGTATTCAC
CGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCACGGGGTCGAGTTGCAGACCCCGATCCGAACTGAGACCGGCTTTAAGGG
ATTCGCTCCACCTCGCGGTATCGCAGCCCTCTGTACCGGCCATTGTAGCATGTGTGAAGCCCTGGACATAGGGGGCATGATGACTTGACCT
CGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTGCGAGTCCCCACCATTACGTGCCTGCCAACACAGGACAAGGGTTGCGCTCG
TTGCGGGA

- *CA-14 forward:*

TGCAGTCGAACGCCCCGCAAGGGGAGTGGCGCACGGGTGAGTAACACGTGGGAACCTACCTTCTGGTACGGAACAACCAAGGGAAACTTT
GGCTAATACCGTATACGACCTCCGGGTGAAAGATTTATCGCCGGAAGAGGGGCCCGCGTCCGATTAGGTAGTTGGTGGGGTAACGGCCTA
CCAAGCCGACGATCGGTAGCTGGTCTGAGAGGATGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTG
GGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGTAAAGCTCTTTCACCCACGAC
GATGATGACGGTAGTGGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCAAGCGTTGTTCGGAATGA
CTGGGCGTAAAGGGCGCGTAGGCGGTTCGTTGCGTCAGATGTGAAAGCCCCGGGCTCAACCTGGGAACTGCATTTGATACGGGCGGGCT
TGAATCCAAGAGACGGTGCTGGAATTCCCAGTGTAGAGGTGAAATTCGTAGATATTGGGAAGACCACCAGTGGCGAACGCGCCCACCTGG
CTTG

- *CA-14 reverse:*

CCGGCGCCCCATTGCTGGTCAGCGCACCGTCTTCGGGTAACCAACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATT
CACCGCGGCGTGCTGATCCGCGATTACTAGCGATTCCAACTTCATGCACCCGAGTTGCAGAGTGCAATCCGAACTGAGACGGCTTTTGGG
GATTAGCTCCAGGTCGCCCCTTCGCTGCCCACTGTCACCGCCATTGTAGCACGTGTGTAGCCCAACCCGTAAGGGCCATGAGGACTTGAC
GTCATCCCCGCCTTCCTCCGGCTTGTCACCGGCAGTATCTCCAGAGTGCCCAACTCAATGATGGCAACTGAAGAGGAGGGTTGCGCTCGTT
GCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTGTGGAAGCCAGCCGAACTGAAGGACCAGATCTC
TCCGGACCATACTTCCCATG

- *CA-15 forward:*

TGCAGTCGACGGAAGGCCCGCTTGCGGGTACTCGAGTGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCTGGACTCTGGGATAAGC
CTGGGAAACTGGGTCTAATACCGGATATGACCTTACATCGCATGGTGTTTGGTGGAAAGCTTTTGCGGTTCAGGATGGGCCCGCGGCCTAT
CAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGC
CCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTC
GGGTTGTAAACCTCTTTCACCAGGGACGAAGCGCAAGTGACGGTACCTGGAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGT
AATACGTAGGGTGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCTCGTAGCGGTTTGTCGCGTCGTCTGTGAATTCTGCAACTCATTG
TAGGCGTGCAGGCGATACGGGCAGACTTGAGTACTACAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAAGAGGA
ACCCCGGTGGCGAAG

- *CA-15 reverse:*

CACTTCAGCTCCTCCACAAGGGGTTAGGCCACCGGCTTCGGGTGTTACCGACTTTCATGACGTGACGGGCGGTGTGTACAAGGCCCGGGA
ACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAATCCGAACTGAGACTGG

CTTTAAGGGATTCGCTCCACCTCACGGTATCGCAGCCCTCTGTACCAGCCATTGTAGCATGTGTGAAGCCCTGGACATAAGGGGCATGATG
ACTTGACGTCATCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTGCAAGTCCCCGGCATAACCCGCTGGCAATACAGGACAAGGGTT
GCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCACCACCTGTACACCAACCACAAGGGGGGCTATA
TCTCTATAGCTTTCTGGTGTATGTCATATCCAGGTAAAGGTTCTTCGCGTTGCATCGAATTAATCCACATGCTCCGCCGCTTGTGCGG

- CA-16 forward:

ACTTCGGTCTGGTGGCGAGTGGCGAACGGGTGAGTAATGTATCGGAACGTGCCCAGTAGCGGGGGATAACTACGCGAAAGCGTAGCTAAT
ACCGCATACGCCCTACGGGGGAAAGCAGGGGATCGCAAGACCTTGCACTATTGGAGCGGCCGATATCGGATTAGCTAGTTGGTGGGGTAA
CGGCTCACCAAGGCGACGATCCGTAGCTGGTTTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
AGCAGTGGGGAATTTTGGACAATGGGGGAAACCCTGATCCAGCCATCCCGCGTGTGCGATGAAGGCCTTCGGGTTGTAAAGCACTTTTGG
CAGGAAAGAAACGTCGCGGGTTAATACCCCGCGGAACTGACGGTACCTGCAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGT

- CA-16 reverse:

AGGCTACTACTTCTGGTAAACCCACTCCCATGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGCGACATGCTGATCCG
CGATTACTAGCGATTCCGACTTCACGCAGTCGAGTTGCAGACTGCGATCCGGACTACGATCGGGTTTCTGGGATTGGCTCCCCCTCGCGG
GTTGGCGACCCTCTGTCCCGACCATTGTATGACGTGTGAAGCCCTACCCATAAGGGCCATGAGGACTTGACGTCATCCCCACCTTCCTCCG
GTTTGTCACCGGCAGTCTCATTAGAGTG

- CA-19 reverse:

CGCAGCGCCTTCGGGTAACCAACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGCATGCTGATCTGCG
ATTACTAGCGATTCCACCTTCATGCACTCGAGTTGCAGAGTGCAATCTGAACTGAGACGGCTTTTGGAGATTTGCTAGGGGTCACCCCTTCG
CTTCCCACTGTCACCGCCATTGTAGCACGTGTGTAGCCCAGCCCGTAAGGGCCATGAGGACTTGACGTCATCCCCACCTTCCTCGCGGCTT
ATCACCGGCAGTCCCCCTAGAGTGCCCAACTTAATGCTGGCAACTAGGGGCGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCA
CGACACGAGCTGACGACAGCCATGCAGCACCTGTCTCCGGTCCAGCCGAACTGAAGGATCTCATCTCTGTGATCCGCGACCGGGATGTCT
TTGGTCTGTTAAGGTTCTGCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGT

- CA-22 forward:

TGCAGTCGACGGCAGCACGGACTTCGGTCTGGTGGCGAGTGGCGAACGGGTGAGTAATGTATCGGAACGTGCCCAGTAGCGGGGGATAA
CTACGCGAAAGCGTAGCTAATACCGCATACGCCCTACGGGGGAAAGCAGGGGATCGTAAGACCTTGCACTATTGGAGCGGCCGATATCGG
ATTAGCTAGTTGGTGGGGTAACGGCTCACCAAGGCGACGATCCGTAGCTGGTTTGAGAGGACGACCAGCCACACTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGGGAAACCCTGATCCAGCCATCCCGCGTGTGCGATGAAGGCCTTC
GGGTTGTAAAGCACTTTTGGCAGGAAAGAAACGTCATGGGTTAATACCCCGTGAAACTGACGGTACCTGCAGAATAAGCACCGGCTAACTA
CGTGCCAGCAGC

- CA-22 reverse:

CCCCCTTGCGGTTAGGCTACTACTTCTGGTAAACCCACTCCCATGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGCG
ACATGCTGATCCGCGATTACTAGCGATTCCGACTTCACGCAGTCGAGTTGCAGACTGCGATCCGGACTACGATCGGGTTTCTGGGATTGGC
TCCCCCTCGCGGGTTGGCGACCCTCTGTCCCGACCATTGTATGACGTGTGAAGCCCTACCCATAAGGGCCATGAGGACTTGACGTCATCC
CCACCTTCCTCCGGTTTGTCACCGGCAGTCTCATTAGAGTGCCCTTTCGTAGCAACTAATGACAAGGGTTGCGCTCGTTGCGGGACTTAAC
CCAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTGTTCCGGTTCTCTTGCGAGCACTCCTAAATCTCTTCAGGATTCCAG
ACATGTCAAGGGTAGGTATTGTTTTTCGCGTTGCATCGAATTAATCCACATCATCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGA

- CA-23 reverse:

CTGCTCCTTGCGGTTGGCGCAGCGCCTTCGGGTAAACCAACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACC
GCAGCATGCTGATCTGCGATTACTAGCGATTCCACCTTCATGCACTCGAGTTGCAGAGTGCAATCTGAACTGAGACGGCTTTTGGAGATTTG
CTAGGGGTCACCCCTTCGCTTCCCACTGTCACCGCCATTGTAGCACGTGTGTAGCCCAGCCCGTAAGGGCCATGAGGACTTGACGTCATC
CCCACCTTCCTCGCGGCTTATCACCGGCAGTCCCCCTAGAGTGCCCAACTTAATGCTGGCAACTAGGGGCGAGGGTTGCGCTCGTTGCGG
GACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTCTCCGGTCCAGCCGAACTGAAGGATCTCATCTCTGTGA
TCCGCGACCGGGATGTCAATTGTCTGTTAAGGTTCTGCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGG

Funded by the European Union

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 the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.