



Horizon 2020 Work programme

Food Security, Sustainable Agriculture and Forestry, Marine, Maritime and Inland Water Research and the Bioeconomy

Call

H2020-FNR-2020: Food and Natural Resources

Topic name

FNR-16-2020: ENZYMES FOR MORE ENVIRONMENT-FRIENDLY CONSUMER PRODUCTS

FuturEnzyme:

Technologies of the Future for Low-Cost Enzymes for Environment-Friendly Products

Final ID: 101000327



23/12/2022

BIO-RESOURCES PREPARED AND EXCHANGED

D3.1

PETER N. GOLYSHIN
BANGOR
GWYNEDD, LL57 2DG, UNITED KINGDOM

Document information sheet

Work package:	WP3, Activity-based bio-prospecting for enzymes.
Authors:	BANGOR (Peter Golyshin, Alexander Yakunin, Olga Golyshina, Tatyana Chernikova), UHAM (Wolfgang R. Streit, Jennifer Chow, Pablo Pérez García), CSIC (Manuel Ferrer, Patricia Molina), UDUS (Karl-Erich Jaeger, Stephan Thies, Alexander Bollinger, Rebecka Molitor), CNR (Michael Yakimov), IST-ID (Carla de Carvalho, Pedro Fernandes)
Document version:	2
Date:	23/12/2022
Starting date:	01/06/2021
Duration:	36 months
Lead beneficiary:	BANGOR
Participant(s):	BANGOR, UHAM, CSIC, UDUS, IST-ID, CNR
Dissemination Level:	Confidential, only for members of the consortium (including the Commission Services)
Type	Other
Due date (months)	2
Contact details:	Peter N. Golyshin, p.golyshin@bangor.ac.uk

Summary

1. Scope of Deliverable 4

2. Reasons for the update 4

3. Bio-resources compiled and exchanged..... 4

Annex..... 7

Bio-resources prepared and exchanged

1. Scope of Deliverable

This deliverable enlists a set of bio-resources (enzymes, isolates, enrichment cultures, clone metagenomic libraries, genomes and shotgun metagenome sequences) available within the consortium. They represent bio-resources generated in the framework of previous European and national-funded projects, and that were compiled and exchanged within the consortium, at the beginning of the project, for screening those relevant to FuturEnzyme. It is worth noting that these bio-resources were not screened at the start date of the project, for the purposes of FuturEnzyme's objectives. The nature and information of these bio-resources, barcoded and accessible through a QR code readable with a mobile phone, have been compiled wherever possible or relevant, and included the origin, identity and culture conditions of the isolates, as well as the cloning system to produce the clone, the host, the vector, the inductor, the antibiotic resistance, the origin of the sample (GPS coordinates, microbe, environmental sample, etc.) and further processing and isolation and production of isolates and clone, production date, etc. The list of bio-resources and the QR code is available in the internal FuturEnzyme repository, and also as a report that will accompany this deliverable (the present document).

2. Reasons for the update

The first version of the Deliverable D3.1 was submitted in July 2021. This update is due to the fact that in these approx 18 months (from July 2021-original submission- to December 2022 -update submission), the partners have been able to collect a larger number of bio-resources that were available but not catalogued, which now is accomplished. In the present document, total numbers (with reference to the previous version) are detailed, and the documents linked to QR codes are also updated with the new entries. In November 2022, the Coordinator (Manuel Ferrer) contacted the Project Officer (Colombe Warin) to explain these circumstances and ask her to re-open the submission of this deliverable (amongst others), at which she agreed.

3. Bio-resources compiled and exchanged

This deliverable gathers all the bio-resources, which were generated in the framework of previous European (FP7, H2020, EraNet) and national-funded projects, and that were made available by and to partners for screening in search for candidates for FuturEnzyme's objectives. The available bio-resources comprised different types of products not screened at the start date of the project, for FuturEnzyme's purposes: 1) enzymes; 2) isolates; 3) enrichment cultures; 4) clone metagenomic libraries; 5) shotgun metagenome sequences. These bio-resources, collected and available at the laboratories of partners CSIC, UHAM, UDUS, BANGOR, CNR and IST-ID, originate from geo-graphically and bio-geochemically diverse extreme and non-extreme environments around the world, whose number and nature is detailed in **Table 1** (access available in the Annex).

Below is a full description of each type of bio-resources. For every type, a PDF file was produced which was attached to a QR code by using the free web www.qr-me.com. The data provided in these documents and the corresponding QR codes are described below.

- **Enzymes:** 1428 (previously 1353, by July 2021).
They represent highly diverse enzymes relevant to FuturEnzyme, available in expression systems, from diverse genomes and metagenomes; the enzymes have been isolated and characterized for purposes others than those in FuturEnzyme, and were now screened with project-relevant substrates and conditions.
The available enzymes belong to the following classes: oxidoreductases, 122; ester-hydrolases, 812; glycosyl-hydrolases, 291; proteases, 26; and others, 177.



The QR code leads to a table detailing: ID enzyme; amino acid sequence; phylum or domain; genera, order or family; screen system; GPS coordinates; latitude; longitude; depth (m); source or source organism (short ID); type of habitat; cloning or expression system; activity.

- **Isolates:** 1492, the genome of 223 of which were accessible (previously 1387, by July 2021). They represent psychrophilic, mesophilic, thermophilic, hyper-thermophilic, thermo-acidophilic, alcaliphilic, extreme halophilic, obligate anaerobic and facultative (micro)aerobic sulphur-respiring microorganisms. The collection includes strains growing at temperatures from 0° to 92°C, pH from 1.5 to 10.0, salinity up to 490 g/L, and pressure up to 50 Mpa. The available isolates belong to the following domains: Archaea, 166; Bacteria, 953; Eukarya, 17. They all together covered at least 22 different phyla, and 113 families.



The QR code leads to a table detailing: Sample ID/Sequence accession; strain name (original); status genome; corrected strain name; taxonomy ID; domain; phylum; class; order; family; coordinates; isolation site; source.

Among them, the genomes of 223 isolates (previously 197, by July 2021), representing lineages of (non)-extremophiles growing from 0 to 92°C, pH from 1.5 to 9.0, salinity up to 492 g/L, pressure up to 50 MPa, are available.



The QR code leads to a table detailing: Sample ID/Sequence accession; sequence and length of the 16S RNA; taxonomic information if available (gene; domain; phylum; class; order; family).

Among them, 55 isolates, did show confirmed activities among common commercial substrates.



The QR code leads to a table detailing: activities tested; if published, the DOI code.

- **Enrichment cultures:** 52 (previously 41, by July 2021). They derived from samples originated from multiple locations and representing enriched microorganisms of at least 16 different genera. They include enrichments from geothermal areas, beach fumaroles, blood drain biofilm, seawater and sediments from multiple locations, street sofa, plant *Sorghum bicolor* rhizosphere, to cite few.



The QR code leads to a table detailing: Starting material; sampling date; GPS coordinates; accession number; published in (journal); enrichment (substrate and medium); further comments including: sample ID; preliminary and validated strain (enriched) name; taxonomy ID; domain; phylum; class; order; family.

- **Clone metagenomic libraries:** 35 (previously 28, by July 2021)

They represent DNA material from communities inhabiting extreme environments (low pH from 1.1 to 4.4; high pH of 9.3-9.6; high salinity from 200 to 490 g/L; pressure up to 300 MPa; temperature up to 98°C) and non-extreme environments, including contaminated sites (close to neutral pH, low to moderate salinity (up to 50 g/L), temperatures from 4 to 30°C, up to 10.1 MPa). They include clone libraries from microbial communities or microbes inhabiting animal guts, deep sea, basins, seawater, marine sediments, single isolates, river water, wastewater treatment plant, bone (cow and turkey femur) biofilms, etc.



The QR code leads to a table detailing: Libraries site/material; sampling date; GPS coordinates of the sampling site; number of clones and vector used to construct the library; accession number (if published); published in (journal); enrichment (substrate) or native community; further comments.

- **Shotgun metagenome sequences:** 64 (previously 61, by July 2021)

They correspond to at least 16 different types of extreme and non-extreme environments. They include shotgun sequences from acidic environments, agricultural Biogas Plant Fermenter, microbiomes associated to animals and microalga, bone-degrading marine microbiome, brine and sediments of crystallizer ponds, contaminated and non-contaminated seawater, marine sediments, subglacial Lakes and soils, biofilm from Brackish Water Aquaculture Plant, surface seawaters, deep sea basins, to mention few.



The QR code leads to a table detailing: Sample description; BioSample/BioProject; GPS coordinates.

The bio-resources described in this deliverable and the QR codes and information detailed through them has been made available to, and exchanged between, the consortium members. The QR codes are confidential and available within the FuturEnzyme consortium. In order to increase the security, they have been blocked with a password (FuturEnzyme€01/06/2021). They are also included in the private area of the FuturEnzyme website (www.futureenzyme.eu), in the section *Shared data* under the designation *D3.1_Bioresources_Month 2_30.11.2022*. This private area that serves as a repository for the project is accessible to the members of the consortium through user and password.

The bio-resources have been massively subjected to bioinformatics and computational (WP2) and functional (WP3) screens, and the results of these screens are detailed in the following deliverables:

- D2.2 “Set of 250,000 sequences pre-selected”
- D2.3 “Set of 1,000 enzymes selected using motif screens”
- D3.3 “Set of 100 best clones, 10 isolates, and 10 enzymes shortlisted for sequencing or transfer to WP2”
- D3.4 “Sequence, activity, and stability datasets from best positive bio-resources”

Annex

Table 1. Detailed information of bio-resources available at month 2 (including the updates by month 18). The different types of bio-resources are: Enzymes, metagenomic clone libraries, enrichment cultures, shotgun metagenomes, isolates (including those with genomes sequenced). The table is available under the designation *Table D3.1_Bioresources_Month 2_30.11.2022* at the FuturEnzyme web intranet through the following QR code (password needed), in the section *Shared data, Datasets*:

