



*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

30/11/2022



# SEQUENCE, ACTIVITY, AND STABILITY DATASETS FROM BEST POSITIVE BIORESOURCES

D3.4

PETER GOLYSHIN

BANGOR

Bangor University, Bangor, Gwynedd, LL57 2DG, UK

## Document information sheet

<b>Work package:</b>	WP3, Activity-based bio-prospecting for enzymes
<b>Authors:</b>	Bangor (Peter Golyshin)
<b>Document version:</b>	1
<b>Date:</b>	30.11.2022
<b>Starting date:</b>	01.06.2021
<b>Duration:</b>	36 months
<b>Lead beneficiary:</b>	Bangor
<b>Participant(s):</b>	CSIC, Bangor, UHAM, UDUS, IST-ID, CNR
<b>Dissemination Level:</b>	Confidential, only for consortium's members (including the Commission Services)
<b>Type</b>	Other
<b>Due date (months)</b>	18
<b>Contact details:</b>	Peter Golyshin (p.golyshin@bangor.ac.uk), Manuel Ferrer (mferrer@icp.csic.es)

## Summary

1. Scope of deliverable .....	4
2. Introduction.....	4
3. Bio-resources available at month 2 .....	4
4. Bio-resources newly generated .....	5
5. Results of massive screens .....	5
6. Annex.....	6

# SEQUENCE, ACTIVITY, AND STABILITY DATASETS FROM BEST POSITIVE BIO-RESOURCES

## 1. Scope of deliverable

This deliverable consists in the datasets informing about the sequences, performances and stabilities of best preselected bio-resources (isolates and clones). This information, is detailed in a comprehensive manner, and is associated through a QR code linked to the bio-resources, available in the internal FuturEnzyme repository.

## 2. Introduction

One of the objectives of WP2 and WP3 was to use available bio-resources from previous projects and new ones to be generated during this project to search for enzymes or enzyme containing microorganisms, relevant to the project. Those bio-resources (available at month 2, and newly generated during the project) consisting of sequences are transferred to WP2 for performing screens by bioinformatics and computational techniques. In parallel, those consisting of other types of bio-resources (available at month 2, and newly generated during the project) are transferred to WP3 for performing screens by functional techniques. It is these latter bioresources that form the basis of this deliverable, in which functional screens have allowed the identification of a series of enzymes and microorganisms with activities of interest to the project. The functional screens, carried out by partners IST-ID, CSIC, Bangor, UDUS, UHAM, and CNR, were performed following the indications and methodologies described in the deliverable D3.2 “Standard assays, analytics and calculations for monitoring enzymatic performance” (July, 2021).

Below, the bio-resources available at month 2 (Section 3), and newly generated during the project (Section 4) are summarized.

## 3. Bio-resources available at month 2

At the beginning of the project, the deliverable 3.1 “Bio-resources prepared and exchanged” was prepared and submitted (July 2021). This deliverable gathers all the bio-resources, which were generated in the framework of previous European and national projects, and that were made available for functional screening in search for candidates for FuturEnzyme’s objectives. The number and nature of all bio-resources, which are going to be updated (submission re-opened on 28 November 2022, and to be re-submitted by December 2022 to the EU portal), are briefly summarized below:

- Enzymes: 1428 (previously 1353, by July 2021)
- Isolates: 1482, of which 223 have their genomes accessible (previously 1387, of which 197 have their genomes accessible, by July 2021)
- Clone metagenomic libraries: 35 (previously 28, by July 2021)
- Enrichment cultures: 52 (previously 41, by July 2021)
- Shotgun metagenome sequences: 64 (previously 61, by July 2021)

This update is due to the fact that since the beginning of the project we have been able to collect a larger number of bio-resources that were available but not catalogued, which we have now been able to do. It is to highlight that the bio-resources collected come from more than 420 geographically and environmental diverse extreme and non-extreme sites. The number and nature of those is compiled in the **Table 1** (access available in the Annex).

#### 4. Bio-resources newly generated

In this deliverable, the information corresponding to the new bio-resources generated, or to which we have had access, since the beginning of the project, is described. The number and nature of those is compiled in the **Table 2** (access available in the Annex).

- Raw environmental samples: 51 new samples were made available from different sites around the world and a variety of habitats to be explored in the seeking of candidates.
- Isolates: 208 new isolates have been delivered from 28 different locations, of which 22 were sequenced.
- Clone metagenomic libraries: 7 fosmid libraries were produced, altogether gathering libraries of at least 2 million clones.
- Enrichment cultures: 42 enrichment cultures have been newly delivered.
- Shotgun metagenome sequences: 54 shotgun metagenomes are newly available from different sampling sites and were transferred to WP2 “Machine learning enzyme bioprospecting integrated into an industrial context”, so new enzymes can further be obtained.

#### 5. Results of massive screens

The screening of a part of the bio-resources available in month 2 and of the new ones generated by applying different functional methods detailed in the deliverable 3.2 “Standard assays, analytics and calculations for monitoring enzymatic performance” has resulted in the following enzymes and isolates as having activities relevant to the project.

- A total of 257 new enzymes with interest, that have been disclosed at this point of the project:
  - 185 enzymes with esterase, lipase, or polyester degrading hydrolase activities
  - 70 enzymes with glycosyl hydrolase activity, 30 being hyaluronidases
  - 2 enzymes with proteolytic activity
- A total of 41 isolates with interesting activity for the project (from 18 different sites):
  - 19 being positive for esterase or lipase activity
  - 20 being positive for glycosyl hydrolase activity, 17 with hyaluronidase activity
  - 1 being positive for protease activity

Note that the identified enzymes, include: i) those enzymes available at month 2 that after screening were found relevant to the project, and ii) those include in positive metagenomic clones and whose activity was confirmed after sequencing, open reading frames selection, expression and characterization. In the case of isolates being positives, those with higher activity phenotype were subjected to genome sequencing to further proceed with annotation and selection of open reading frames encoding the enzyme of interest, which was further cloned or synthesized, expressed and characterized. For details, see the **Table 2** (access available in the Annex). We would like to highlight that for the screening, both commercial substrates relevant to the project and real-life substrates provided by industrial partners, were used to guarantee the selection of candidates best fitting to the project.

All the enzymes and isolates identified as positives after functional screens were transferred to WP4. They constitute an important part of the total enzymatic materials generated during the project until month 18, together with those identified by *in silico* tools that are detailed in the deliverable 2.2 “Set of 250,000 sequences pre-selected” (November 2021; to be updated by December 2022). The enzymatic performances (activities and stabilities) of all enzymes and isolates identified and successfully expressed, produced and/or characterized, including the ones reported in this deliverable, have been extensively described in the deliverables 4.6 “The metadata on expression yield activity and stability” (November, 2022), 4.7 “At least 180 enzymes (recombinant, native, biomimetic) with attractive properties” (November, 2022), and 5.1 “The shortlist of at least 18 enzymes nominated for engineering” (November, 2022). These deliverables gather the information regarding description and characterization (sequence, activity, stability, source, etc.) of the best positive bio-resources. For this reason, the performance phenotypes of all enzymes and isolates have not been detailed in this document, in order to avoid repetition.

## 6. Annex

**Table 1.** Detailed information of bio-resources available at month 2. 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*:



**Table 2.** Detailed information of bio-resources newly sampled and generated since the beginning of the project. The different types of bio-resources are: new samples, new metagenomic clone libraries, new enrichment cultures, new shotgun metagenomes, and new isolates (including those with genomes sequenced). The table also contains information of the series of enzymes and microorganisms with activities of interest to the project identified by performing functional tests of the bio-resources available at month 2 (Table 1), and newly generated during the project (Table 2). The table is available under the designation *Table D3.4\_Bioresources\_New for FuturEnzyme\_30.11.2022* at the FuturEnzyme web intranet through the following QR code (password needed), in the section *Shared data, Datasets*:

