A high-throughput experimental evolution platform for the flexible optimization of antimicrobial probiotic and biocontrol strains

FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Microbial cell counts via flow cytomtery	Read-out for all the bacterial competition and fitness assays (both in vitro and in vivo)	Generate new data	Digital	Experimental	.xls	<1GB	
Microbial cell counts via CFU	Read-out for all the bacterial competition and fitness assays (both in vitro and in vivo)	Generate new data	Digital	Experimental	.fcs	<100GB	
Bacterial biomass production	Measurements of biofilm biomass via crystal violet assay	Generate new data	Digital	Experimental	.xls	<1GB	
Optical density measurements	Measurements of bacterial population densities based on optical density	Generate new data	Digital	Experimental	.xls	<100MB	
Microscopy images	Microscopy images of bacterial biofilms and wound tissues	Generate new data	Digital	Experimental	.czi .tiff	<100GB	

Whole genome sequencing	WGS data of the ancestral and selected evolved strains	Generate new data	Digital	Experimental	.FASTQ .BAM	<5TB	
RNA sequencing	RNA-sequencing of the ancestral and selected evolved strains	Generate new data	Digital	Experimental	.FASTQ .BAM	<1TB	
Evolved bacteria	Evolved bacterial populations and isolated clones from the evolved populations	Generate new data	Physical				~5000-10000 samples, stored in 96-well plate format
Porcine tissue samples	Wound biopsies for histopathological analysis	Generate new data	Physical				~50 samples, embedded in paraffin, stored at room temperature
RT-qPCR	qPCR measurements of the immune response in the porcine trials and the confirmation of of *Agrobacterium* infection in the green house trials	Generate new data	Digital	Experimental	.SDS	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

We will make use of a published reference genomes of *L. rhamnosus* GG (NCBI RefSeq assembly GCF_000011045.1) to map our whole genome and RNA sequencing data.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, animal data

In *vivo* validation of the evolved probiotic strains are planned in project year 3. An application to the animal ethical committee of the KU Leuven will be submitted once the required preliminary data is collected.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

Both the improved probiotic strains and the experimental evolution platform will be valorized. IP will be protected by coyright and patent applications.
The project Exploitation Committee will assess the opportunity of patenting, monitor developments in the targeted markets, re-align the project's R&D objectives if need be according to the evolving needs of the industrial end-users, seek out potential application areas and business opportunities for the project's developed technology, and generate and update the dissemination and exploitation plan of the R&D results.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

We have FTO for the probiotic and biocontrol strains selected for the different case studies. The IP on the Paenibacillus strain is owned by KU Leuven, whereas the patents covering the use of Bacillus and Lactabcillus in the envision application are expired or aborted.

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

1. The data will be accompanied by a detailed metadata txt file denoting important characteristics (e.g. strain, material, time point,..) necessary for interpretation of the results. The key characteristics will also be denoted in the filename of the data files.
2. A detailed experimental protocol will be added to the directory of the corresponding experimental results. This step-wise description will facilitate potential future reproduction of the experiments.
3. For every deliverable in the project, a general outline txt file will be created. This file provides an overview of all the available data, the design of the experiment and the structure of the data saving.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

Microscopy data: OME-XML metadata standard
Sequencing data: MIxS metadata standard
Flow cytometry data: MIFlowCyt metadata standard
qPCR data: MIQE metadata standard
A metadata template will be constructed in the frame of the project for data types where no general metadata standard is available.

3. Data storage & back-up during the research project

Where will the data be stored?

1. Data generated by the different partners will be saved on a collective sharepoint
for the data during the project. In view of the size of the raw microscopy and sequencing data,these data will be exempt from the sharepoint.
2. Additionally each partner will store their generated data on the central storage facility of their respective research units with the same structure as the sharepoints.
3. Personal copies can be made and kept on personal devices.

How will the data be backed up?

1. The collective sharepoint is backed up three times per day.
2. The internal back-up of the specific partners is managed according the procedures of their respective research institutions.

**Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.**

- Yes

Sufficient storage space is provided by the respective research institutions.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

1. Viewing and modification rights in sharepoint are granted based on the involvement in the specific work packages.
2. Critical documents, e.g. reports, presentations,... can
be (temporarily) locked by the author(s).
3. Sharepoint provides a changelog for detecting and reverting possible unauthorized changes.
4. The internal storage of the partners provides a back-up for the sharepoint and vice versa.
5. Physical data is stored in a secured -80°C freezer at the facility with limited-access.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

1. The total cost of storage are estimated on ~€3000-€3500 per year, mainly for the raw sequencing data (~4TB).
2.The costs will be covered by the respective partners using the allocated project budget (working budget: consumables).

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All of the generated data will be stored for minimum 5 years after the end of the
project.

Where will these data be archived (stored and curated for the long-term)?

1. Every partner will store their generated data for at least 5 years according to the storage and back-up procedures present at their research institute. 2. The project coordinator will store the total generated data in the research project using the archive network (K:) drives provided by the KU Leuven.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

In light of the size of the sequencing data (4 TB), the current storage volume of the archive drive at KU Leuven needs to be increased. The costs for an expansion of 4 TB for 5 years are estimated on € 4,000, and will be covered by the allocated project budget.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

All digital data will be made available upon publication. The evolved bacterial strains will not be made available in order to protect the valorisation potential of these strains.

If access is restricted, please specify who will be able to access the data and under what conditions.

n/a

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The sequencing data will be deposited International Nucleotide Sequence Database Collaboration (INSDC).

When will the data be made available?

The data will be made available upon publication of the corresponding research results.

Which data usage licenses are you going to provide? If none, please explain why.

We will make use of open licenses such as CC-BY or their equivalent as our data is not considered sensitive and open licenses are often required for high impact publications.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

No costs are expected.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

The WP leaders carry the end responsibility for the correct documentation of the data generated in their respective WP

Who will manage data storage and backup during the research project?

Each partner is responsible for correct data storage, i.e. sharepoint and internal, for the duration of the project. The project coordinator is responsible for the coordination of the data storage

Who will manage data preservation and sharing?

The coordinator of the project will collect all the data after the project and ensure correct preservation of the data

Who will update and implement this DMP?

The coordinator, together with the project partners, bears the end responsibility of updating & implementing this DMP.