Synthetic microbial consortia engineered from the gut microbiome of wood-eating beetle larvae to unlock the convergence of lignin to value-added biochemicals.
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Synthetic microbial consortia engineered from the gut microbiome of wood-eating beetle larvae to unlock the convergence of lignin to value-added biochemicals.
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

When available, metagenomic, metaproteomic, and metabolomic data from WP1 will be annotated with metadata (beetle species, stage, substrate type and batch, gut segment, references to parallel measurements). This data will be reused from another project, but new data will also be generated during the current research. A curated enzyme database to explore lignin metabolism can also be created in this WP, while new data on culturing and metabolic traits will be collected in WP2. For the isolated strains, annotated genome data and reconstructed metabolic pathways will be produced in WP3 (new). Bioreactor monitoring data (pH, O2, consumption, and production rates) will be registered in WP4 (new).

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

ManGO will be used from the start for active research data management. Raw data will be stored on a dedicated large-volume storage drive with restricted access for internal use, as well as on the NCBI database (under embargo until publication). OneDrive (KU Leuven), with access permissions for the PhD researcher(s) and the promotor (Benjamin Horemans), will be used to keep active research data. At the end of the project, raw and processed annotated data (including unpublished) will be uploaded to the Research Data Repository at KU Leuven. At the moment of publication, data will be made open-access.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

I don't want to deviate from the principle of preservation of data.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

There are no such issues

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

There are no other issues

Synthetic microbial consortia engineered from the gut microbiome of wood-eating beetle larvae to unlock the convergence of lignin to value-added biochemicals.
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Synthetic microbial consortia engineered from the gut microbiome of wood-eating beetle larvae to unlock the convergence of lignin to value-added biochemicals.
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	
Beetle microbiome	Sequencing (amplicon + metagenomics+sanger) +metaproteomic	New and reused data	Digital	Experimental	.fasta/q, .txt, .xml, .csv	<5TB	
In vitro metabolites	Chemical analysis data on metabolites produced	New and reused data	Digital	Experimental	.txt, .xml, .csv	<1GB	
Metadata beetle physicochemical conditions	Measuring data on physicochemical properties in beetle gut	New and reused data	Digital	Experimental	.txt, .xml, .csv	<1GB	
Metadata bioreactor conditions	Measuring + monitoring data on bioreactor properties	New and reused data	Digital	Experimental	.txt, .xml, .csv	<1GB	
Metadata bacterial cultures	Description of bacterial isolates + growth conditions + metabolic traits	New data	Digital and physical	Experimental	.txt, .xml, .csv	<1GB	<3 L
Metadata metabolic reconstruction	Data on whole genome sequencing + annotated genomes + reconstructed metabolic pathways	New data	Digital	Experimental	.fasta/q, .txt, .xml, .csv	<5TB	

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:

For the beetle microbiome, *in vitro* metabolites, metadata on beetle physicochemical conditions and metadata on bioreactor conditions, I will reuse some of the data generated by Lilian Prinsen during the course of her PhD project with grant number 1SH8P24N.

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.

- No

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.

- No

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

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).

On the shared J-drive, we have a lab manual with standardized protocols that are being done in the lab. For new experiments, new protocols will be set-up and shared. Performed lab work will be written down in an Electronic Lab Notebook (Benchling), and shared with the supervisor, lab technician and other colleagues involved in the project. When the results will be published, materials and methods will be described in detail so that the experiment can be repeated by an experienced researcher.
All generated data will be stored on the KU Leuven server, which contains protected project directories to which only researchers involved have access, and also in ManGO. In addition, every researcher has a personal directory on the KU Leuven server for safe data storage and a OneDrive directory. Data is never stored on personal or work devices to prevent data loss upon technical failure.

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

A metadata template will be available for storage of raw sequencing data on the shared K-drive and in ManGO.

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

Where will the data be stored?

Raw sequencing data will be stored on a dedicated large volume storage drive with restricted access for internal use (K-drive and ManGO). Active research data will be kept on OneDrive and Benchling (electronic lab journal) with access permissions for PhD researcher(s) and promotor (Benjamin Horemans).

How will the data be backed up?

Standard back-up provided by KU Leuven ICTS and ManGO.

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

The research group has enough storage in the K-drive and ManGO. If the storage capacity is reached (unlikely), additional storage capacity can be requested/purchased.

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

The data will be stored in the University's secure environment with restricted/personalized access for different users.

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

Costs for data storage are covered by the research group.

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 data will be preserved for 10 years according to KU Leuven RDM policy.

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

At the end of the project, data (raw and processed, including unpublished) annotated with metadata will be uploaded to the Research Data Repository (RDR) at KU Leuven. Data will be made open-access on the moment of publication.

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

Costs for data storage are covered by the research group.

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

Sequencing data from beetle gut microbiome and whole genome sequencing/annotations from the microbial isolates.

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

NA

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.

Data will be available at the KU Leuven Research Data Repository, on ManGO, and on the publications that will come out of this project.

When will the data be made available?

Upon publication of the research results.

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

Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators.

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

All publications and associated data will receive a DOI.

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

Sharing data via NCBI is free of charge. Data present in manuscripts will cost the open access publication costs.

6. Responsibilities

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

Ángela González Álvarez (PhD), Benjamin Horemans (promotor), Dries Grauwels (lab technician)

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

Ángela González Álvarez (PhD) and ICT support

Who will manage data preservation and sharing?

Ángela González Álvarez (PhD) and Benjamin Horemans (promotor)

Who will update and implement this DMP?

Ángela González Álvarez (PhD)