DMP title

Project Name Discovery of novel antibiotics by highly sensitive genotype-based high-throughput selection - DMP title

Project Identifier 92261

Grant Title 1S20922N

Principal Investigator / Researcher Sandrien De Smedt

Project Data Contact sandrien.desmedt@kuleuven.be

Description The goal of my project is to develop a new PCR-assay for the detection of new gene clusters for antibiotic synthesis. For this purpose, soils will be collected and their DNA extracted, after which the DNA will be used to construct metagenomic libraries in an E.coli host. These libraries will then be screened, using PCR, with our own designed primer pairs, for the detection of new antibiotic biosynthetic gene clusters. In the case a positive PCR signal is detected, the corresponding clones are isolated and the vectors are recovered. Using sequencing, potentially interesting gene clusters might be discovered, in which case the cluster will be heterologously expressed in Streptomyces and screened for antibiotic activity against pathogens.

Institution KU Leuven

1. General Information

Name applicant

Sandrien De Smedt

FWO Project Number & Title

Number: 1S20922N

Title: Discovery of novel antibiotics by highly sensitive genotype-based high-throughput selection

Affiliation

- KU Leuven
- Other

VIB and IMEC

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of data	Format	Volume	How created
WP1: Assay optimization for detection and isolation of antibiotic BGCs			
Primers	Digitally preserved sequences	20 KB	Primer pairs described in literature or designed based on sequences from MiBIG database.
Primers	Vials stored at -20°C		Primers described in literature of designed based on sequences from MiBIG database and ordered at IDT.
Streptomyces, Pseudomonas putida and *Paenibacillus polymyxa* strains including lab strains and clones containing (parts of) the targeted antibiotic-producing gene cluster	Glycerol stocks frozen at -80°C	2 mL cryotubes and 96-well plates	*Streptomyces* strains obtained from Lieve Vanmellaert (Rega Institute), *Pseudomonas putida* from the lab of R. De Mot and *Paenibacillus polymyxa* ordered from the Belgian Coordinated Collections of Microorganisms (BCCM).
E. coli with pUC vectors containing amplicons	Glycerol stocks frozen at -80°C	2 mL cryotubes and 96-well plates	Cloning of amplicons in pUC vector and chemical transformation into *E.coli* lab strain..

PCR data	.xslx	5 GB	qPCR optimization on a dilution series of model DNA templates containing biosynthetic gene clusters of interest.
Droplet microfluidics data	.csv, .txt	2 TB	FACS for fluorescent read-out and sorting of ddPCR optimization and screening.
Plate counts	.xlsx	1 MB	Plate counts from recovery of sorted vectors after FACS-sorting, determination of transformation efficiency.
Optimized chips for the microfluidics system	/	/	/
Microscopy images	.nd	1 TB	Fluorescence images of chips used for optimization of ddPCR.
Optimized chips for the microfluidics system	-	-	Chips are provided by Imec and used for the optimization of ddPCR.
DNA sequences	.fasta	500 MB	Sanger sequencing to confirm that primers are suitable to amplify the desired sequences.
WP2: Droplet-PCR based and FACS-mediated isolation of novel antibiotic BGCs			
(Meta)genomic libraries	Glycerol stocks frozen at -80°C	2 mL cryotubes and 96-well plates	(Meta)genomic library construction starting from eDNA or gDNA from *Streptomyces*, *Pseudomonas putida* and *Paenibacillus polymyxa.*
Pool of recovered vectors	Cryopreserved plasmid stocks and transformed libraries in Eppendorf tubes		Extraction of the vectors out of the droplets after the sorting step. Transformation of these extracted vectors into an *E.coli* lab strain.
DNA sequences	.fasta	2 TB	Long read sequencing of the recovered vectors.
List of already known biosynthetic gene clusters and their sequences	.fasta	300 KB	MiBIG database.
WP3: Production of novel antibiotics in heterologous expression hosts			
Saccharomyces cerevisiae, *Streptomyces*, ESKAPE pathogens, including lab strains, natural/clinical isolates. Strains containing the biosynthetic gene cluster of interest.	Glycerol stocks frozen at -80°C	2 mL cryotubes and 96-well plates	*S.cerevisiae* from Natalay Kouprina, *Streptomyces* from Lieve Vanmellaert, ESKPAPE pathogens through commercial sources such as ATCC or BCCM or collaborations with other labs

DNA sequences	.fasta	500 MB	Sequencing to confirm that entire gene cluster is cloned into the heterologous expression host.
Production efficiencies	.xslx, .D	1 GB	Detection of secondary metabolites using HPLC, NMR and MS/MS.
Images of overlay assays, MIC, MBC tests	.tif, .jpeg	10 GB	Overlay assays, MIC- and MBC-tests to test the activity of a potential new antibiotic against ESKAPE pathogens.
WP4: Initial determination of the mode of action of the isolated antibiotics			
Microscopy images and timelapses	.nd	5 TB	Phase contrast and fluorescence images/time lapse microscopy of antibiotic sensitive cells.
ESKAPE pathogens, including lab strains, natural/clinical isolates, knock-out libraries, evolved strains	Glycerol stocks frozen at -80°C	2 mL cryotubes and 96-well plates	Strains and libraries will be commercially ordered or obtained through collaboration with other labs obtained by experimental evolution
DNA sequences	.fasta	5 GB	Whole genome sequences of antibiotic sensitive strains

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:
Short description of the kind of personal data that will be used:

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Yes

Potential tech transfer will be discussed with the research and development offices of KU Leuven, IMEC and VIB. Ownership of the generated data has been stipulated in a Cooperation Agreement. Concept note:

- IMEC: chip fabrication and design, setup development
- IMEC-KU Leuven Michiels: droplet assay development within this project

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

BIOLOGICAL DATA

Cryotubes and plates will be labelled with a reference number that links to an entry in our Microsoft Access Database which is hosted on a central server and accessible to all people involved in the project. All relevant information on the specific strains and vectors is included in this database. This includes strain identifier, a clear description of how the strains were constructed and a link to whole genome sequence if applicable.

EXPERIMENTAL RESULTS

Data files will be named using “yearmonthday_titleordescription”. The data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in the same folder also containing the results. Furthermore, a digital notebook in OneNote is used to register day-by-day activities in the lab or at the computer.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

Since there is no formally acknowledged metadata standard specific to our discipline, Dublin Core Metadata will be used. Moreover, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards that are more specific to our data.

5. Data storage and backup during the FWO project

Where will the data be stored?

BIOLOGICAL MATERIAL

Cryotubes and 96-well plates will be stored in a -80°C freezer present in the Michiels lab.

EXPERIMENTAL RESULTS

Data will be generally stored in two or more locations. For small-volume datasets, data will be stored on the device where the data is generated and a Onedrive/Sharepoint location.

A network drive will be used for large-scale data. A copy of these datasets will be made to desktop PCs with large computational power (or to the computational cluster of KU Leuven) whenever data analyses will be performed.

How is backup of the data provided?

BIOLOGICAL MATERIAL

A backup of selected strains will be stored in a different location**.**

EXPERIMENTAL RESULTS

The use of Onedrive/Sharepoint is backed up via a local copy and a (back-up) copy in the cloud/on the KU Leuven network and offers version control to allow easy recovery in case of mistakes. Similarly, network drives of KU Leuven follow daily back-up procedures and offer version control via the built-in features of Windows.

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

BIOLOGICAL MATERIAL

Sufficient storage is available.

EXPERIMENTAL RESULTS

OneDrive at KU Leuven alone already offers 5 TB of data per user. Network storage is purchased on a group level an increased whenever needed.

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

BIOLOGICAL MATERIAL

-80°C freezers are present in the Michiels lab (costs are covered by general lab expenses).

EXPERIMENTAL RESULTS

The costs for large volume storage will be covered by general lab financing (€569.20 per 5 TB per year).

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

BIOLOGICAL MATERIAL

Unauthorized people do not have access to the strain collection.

EXPERIMENTAL RESULTS

Access to data will be secured since data on the network drives are stored in the university’s secure environment. OneDrive storage is linked to my personal secured KU Leuven account. Both are secured by a two factor authorization and frequently changed passwords.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

BIOLOGICAL MATERIAL

All strains will be stored for at least 5 more years (and longer, complying with the 10 year data preservation rule of KU Leuven) after the end of the project.

EXPERIMENTAL RESULTS

All data will be stored for at least 5 more years (and longer, complying with the 10 year data preservation rule of KU Leuven) after the end of the project.

Where will the data be archived (= stored for the longer term)?

EXPERIMENTAL RESULTS

Data will be stored on the university’s central servers for 10 years. Published results will be deposited conform the journal’s policy.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

BIOLOGICAL MATERIAL

-80°C freezers are present (included in general lab costs). Deposit of biological material in public repositories is generally without a fee.

EXPERIMENTAL RESULTS

For the storage of essential data for 5-10 year, I will make use of publicly and free repositories (only for published data) and of the KU Leuven long-term, large-volume central storage servers (estimated cost of 100 euro per year).

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

For potential IP purposes, data will not be shared before publication or following a five-year embargo period after termination of the project.

Which data will be made available after the end of the project?

BIOLOGICAL MATERIAL

Microbial strains generated in this project will be made available upon simple request.

EXPERIMENTAL RESULTS

Published data are likely to include droplet microfluidics data, sequencing data, *S. cerevisiae* strain descriptions, platform optimization procedures, …

Where/how will the data be made available for reuse?

- Other (specify):

BIOLOGICAL MATERIAL

We aim at communicating our results in top journals that require full disclosure of all included data. Biological material will be shared upon simple request following publication.

EXPERIMENTAL RESULTS

We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in the supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply. Proper links to these data sets will be provided in the corresponding publications.

When will the data be made available?

Upon publication of the research results

After an embargo period of 5 years for unpublished data to allow the research group to publish research findings.

Who will be able to access the data and under what conditions?

BIOLOGICAL DATA

Biological data will be distributed to other parties if requested.

EXPERIMENTAL RESULTS

Depending on the journal: open access or standard subscription-based publication.

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

BIOLOGICAL MATERIAL

Generally, shipment is paid by requesting parties

EXPERIMENTAL RESULTS

Online repository is free of charge. Network storage at KU Leuven comes with a cost of 100 euro per TB per year, covered by the host lab after termination of the project.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Sandrien De Smedt

Who will be responsible for data storage & back up during the project?

Sandrien De Smedt

Who will be responsible for ensuring data preservation and reuse ?

Jan Michiels (Centre of Microbial and Plant Genetics, KU Leuven)

Who bears the end responsibility for updating & implementing this DMP?

Jan Michiels (Centre of Microbial and Plant Genetics, KU Leuven)