DMP title

Project Name DMP_G0D0322N - DMP title

Grant Title G0D0322N

Principal Investigator / Researcher Dirk Springael

Project Data Contact Dirk Springael, Tel. 016 32 16 04; e-mail: dirk.springael@kuleuven.bel

Description Biological treatment involving bioaugmentation has been proposed as a green alternative for physico-chemical treatment of intake groundwaters contaminated with organic micropollutants (OMPs) for drinking water production. A main bottle neck though is the inefficiency of bioaugmentation. Bioaugmentation merely involves an invasion process. In this project, ecological invasion theories are put to the test addressing different determinants of invasion, i.e., selection, dispersion and diversification, to improve bioaugmentation. This will be done for three different OMP catabolic bacterial strains in continuous biofilm ecosystems and the effects of these manipulations on the resident community and its functionality will be examined. Selection is manipulated by creating new niches by adding selective C-sources, by implementing environmental disturbances to remove competing organisms or opening space, and by co-inoculation with beneficially interacting bacteria. Dispersion will focus on propagule pressure while diversification will examine the evolution of the inoculum towards variants with improved invasion abilities. Invasion success will be scored based on both cell numbers and OMP biodegradation activity. The project will provide both novel applied and fundamental information for and regarding the microbial invasion of complex ecosystems.

Institution KU Leuven

1. General Information

Name applicant

Dirk Springael

FWO Project Number & Title

G0D0322N; Putting ecological invasion theories to the test: improving bioaugmentation of drinking water treatment systems for organic micropollutant removal

Affiliation

- KU Leuven

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).

Degradation data in biofilm settings under different conditions (.xls); 100 files, max. 200 GB.

Biofilm microscopy images (.tif); generated by confocal laser scanning microscopy); 10 GB per image; up to 1000 images.

16S rRNA gene amplicon high throughput Illumina sequencing data (FastQ format); 100 files; size unknown.

OTU table derived from the 16S amplicon sequences and containing the relative abundances of each Operational Taxonomic Unit (OTU) for each sample (in.csv format)

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?

- No

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?

Overall, (e-)labbooks will contain information on experimental design, protocols, sampling location, abbreviations used, structure of the data (including link with physical storage of data), and steps involved in data analysis and relevant analysis scripts (R scripts, MOTHUR/QIIME scripts). A clear coding for all data files related to the project will be used. In the concluding stage of the project, a master index file containing the combined information for all experiments will be compiled which will be archived and also stored on the personal harddrives/PC of the PI. Altogether, this should allow any secondary analyst to use the data accurately and effectively.

More specifically, the following information will be given on the items described in section 2:

Degradation data in biofilm settings under different conditions will be documented with a detailed description of the used protocol in the xls file and will contain information about the sample of origin (which biofilm, which experiment, aim of experiment etc.).

Biofilm microscopy images (.tif); generated by confocal laser scanning microscopy); The methodology and protocol will be described in detail in the lab book and a ReadMe file of the image collection will be written including the protocol.

16S rRNA gene amplicon high throughput Illumina sequencing data deposited at EMBL via BANKIT, will include the information/documentation required by the data base.

OTU table derived from 16S amplicon sequences and containing the relative abundances of each Operational Taxonomic Unit (OTU) for each sample will contain information about the sample of origin (which biofilm, which experiment, aim of experiment etc.)

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

16S rRNA gene amplicon high throughput Illumina sequencing data deposited at EMBL via BANKIT, will include the information/documentation required by the data base.

5. Data storage and backup during the FWO project

Where will the data be stored?

All data will be stored in the university's secure environment. Nucleic acid sequence data will be submitted and stored in official nucleic acid data bases like EMBL upon publication or at the end of the project.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures.

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.

The amount of data foreseen to not encompass foreseen capacities

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

Not expected to be high. If any, will be covered by the FWO project, if any.

Data security: 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.

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, ...).

All data will be preserved for at least 5 years after completion of the project.

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

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy. 16 S nucleic acid sequence data will be stored at official public data bases like EMBL via BANKIT.

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

Not expected to be high. If costs, they will be covered by on-going related projects.

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)?

- No

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

The full dataset (except the nucleic acid sequence data) will be deposited in a cvs format in KU Leuven RDR under a CC-BY license. The 16S nucleic acid sequence data will be avalailble through official nucleic acid databases like EMBL.

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

- In an Open Access repository

Through KU Leuven RDR.

When will the data be made available?

- Upon publication of the research results

The full dataset will be uploaded and made available in a cvs format in RDR immediately afer the end of the project in case published. Others will be added upon publication. If not published within 1,5 years of project complation, all datasets will be made available.

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

Open access data in RDR.

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

No direct idea but expected to be not high. In case, costs will be covered by related projects.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Afshin Abolhasani who will act as the PhD working on the project.

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

Afshin Abolhasani who will act as the PhD working on the project.

Who will be responsible for ensuring data preservation and reuse ?
PI Dirk Springael

Who bears the end responsibility for updating & implementing this DMP?
The PI bears the end responsibility of updating & implementing this DMP.