DMP title

Project Name DMP_G0E8122N - DMP title
Grant Title G0E8122N
Principal Investigator / Researcher Dirk Springael
Project Data Contact Dirk Springael; Tel. 016 32 16 04; dirk.springael@kuleuven.be
Description Horizontal gene transfer plays an essential role in the adaptation of microbial communities to anthropogenic environmental changes. Typical examples are adaptations that confer antibiotic resistance, metal resistance and catabolism of xenobiotic organic pollutants. Among the vehicles that govern HGT are conjugative elements and insertion sequence (IS) elements. Conjugative elements such as plasmids transfer the adaptive gene functions between different organisms while IS elements are thought to play an essential role in the acquisition of the adaptive traits by the conjugative elements as such making the traits prone to lateral transfer. While much information exists on the molecular mechanisms of IS transposition and IS mediated gene acquisition, less information exist, and this in contrast to plasmids, about the contribution of IS elements to HGT and microbial adaptation in a wider community context. This project aims to acquire a better understanding of the role of IS elements in genetic adaptation at the community level and the anthropogenic environmental drivers underlying IS mediated gene recruitment and HGT in complex communities. Using innovative tools and experimental design, it will examine the abundances and genetic cargo of IS elements as well as IS mediated gene recruitment and HGT in soil and freshwater environments and will explore their relationships with environmental parameters and anthropogenic selective pressures.
Institution KU Leuven

1. General Information
Name applicant

Dirk Springael

FWO Project Number & Title

G0E8122N; Deciphering the role of insertion elements in community adaptation to anthropogenic environmental changes.

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).
Data files (.xls) on origin and characteristics of soil/water samples, around 30 in total (max I MB) created by gathering information on site and physico-chemical analysis of soil characteristics.
Sequencing data files (fastaq format) of long read PCR amplicons, around 40, total around 60 TB, generated by short read/long read sequencing.
Sequencing data files (fastq format) of 16S amplicon, around 40, ongeveer 10 TB, generated by short read sequencing.
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?

The project is a collaboration with Prof. Su Jian-Qiang of the Institute of Urban Environment in China and agreements have to be made when common results derived from collaborations need to be disseminated and published. We do not see a problem in there since the project has a high fundamenbtal character. This relate to all data sets mentioned.

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:

Metadata files on origin and characteristics of samples, will be provided with a clear description of the methods (like soil analytics) that were used to collect the data. The metadata will include information on the samples (where collected (GPS coordinates), environmental conditions at site when isolated, history of the location, soil physico-chemical characteristics). In addition it will provide where the related sequence data can be found.

Sequencing data files deposited in sequence data bases like EMBL 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 the documentation about origin of the sample, sequencing method, 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

Metadata files on origin and characteristics of samples, will be provided with a clear description of the methods (like soil analytics) that were used to collect the data. The metadata will include information on the samples (where collected (GPS coordinates), environmental conditions at site when isolated, history of the location, soil physico-chemical characteristics). In addition it will provide where the related sequence and mineralization data can be found.

Sequencing data files deposited in sequence data bases like EMBL 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?

The data will be stored in the university's secure environment for private data. Nucleic acid sequence data (16 S and long read PCR ampicons) will be submitted and stored in official nucleic

acid data bases like EMBL upon publication or at the end of the project. In addition, we will set up a shared repositort through a protected collaboration tool (like SharePoint or One Drive) (to be decided) for sharing data between the Belgian and Chinese partner.

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.

- Yes

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?
We expect this to be limited. If any, costs will be covered by the FWO project itself.

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 or in case of shared with the Chinese partner in the secured environment of SharePoint or One Drive

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. Nucleic acid sequence data (16S and long read PCR amplicons) will be stored at official public data bases like EMBL.

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 any costs, they will be covered by related research 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 nucleic acid sequence data (16S and long read PCR amplicons) 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
- Other (specify):

Through KU Leuven RDR. Sequencing data through EMBL.

When will the data be made available?

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 after agreement with the Chinese partner in case data from collaborative actions. 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 costs will be limited. In case, costs will be covered by FWO project itself.

8. Responsibilities
Who will be responsible for data documentation & metadata?

The scientist (to be hired) hired on the project to do the main work.

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

The scientist (to be hired) hired on the project to do the main work.

Who will be responsible for ensuring data preservation and reuse ?

The PI Dirk Springael

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

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