DMP ZB/22/019 - Molecular and ecological features of *Ideonella sakaiensis* mediated biodegradation of Polyethylene terephthalate (PET) plastics.

Project Name: Molecular and ecological features of *Ideonella sakaiensis* mediated biodegradation of Polyethylene terephthalate (PET) plastics.

Grant Title: DMP ZB/22/019

Principal Investigator / Researcher: Dirk Springael/Hazrat Noor

Project Data Contact: Dirk Springael

Description: Petroleum-based plastics form serious environmental contamination, especially eminent in developing countries with poor waste management. Despite the inherent recalcitrance character of petroleum-based plastics, several recent studies report on bacteria that metabolize synthetic plastics using the plastic as sole source of carbon and energy. This indicates that, in analogy with the bacterial catabolism of xenobiotic chemicals like pesticides, microbes can metabolically adapt to use plastics at their advantage. This project has two aims. First, it aims to acquire a better understanding of the microbial adaptations to degrade the classic petroleum-based plastic polyethylene terephthalate (PET) by identifying auxiliary functions for PET biodegradation that go beyond the catabolic enzymes, in the PET assimilating *Ideonella sakaiensis* 201F-6. Second, that information is used to acquire an understanding of the distribution of bacterial PET degraders and the evolutionary tracks in PET infested areas like landfills, by analyzing DNA and PET degrading isolates from PET debris samples.

Institution: KU Leuven

1. GENERAL INFORMATION

Name applicant

Dirk Springael

Project Number & Title

ZB/22/019 - Molecular and ecological features of *Ideonella sakaiensis* mediated biodegradation of Polyethylene terephthalate (PET) plastics.

Affiliation

KU Leuven

2. DATA DESCRIPTION

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

Generate new data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume, ideally per objective or WP of the project. You might consider using the table in the guidance.

Metadata files (.xls) on origin and characteristics of PET samples, around 30 in total (max I MB) created by gathering information on site and physico-chemical analysis of the PET samples.
Sequencing data files (fastq format), around 30 genome sequences, approx. 1 TB, generated by short read/long read sequencing

16S amplicon sequencing data files (fastq format) from around 30 plastic samples generated by Illumina sequencing, approx. 100 GB.
ASV tables derived from the 16S amplicon sequences and containing the relative abundances of each ASV for each plastic sample (in.csv format)
Bacterial culture collection (approx. 30 strains) in glycerol medium, stored at -80°C. 2 replicates per strain.
PET degradation kinetic data (.xls), around 60 in total (max. 10 MB) generated by lab-experiments collecting CO2 produced from PET

3. LEGAL & 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 your host institution's privacy register - not relevant yet)
No

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 & 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 hard drives/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 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.

ASV tables derived from 16S amplicon sequences and containing the relative abundances of each ASV for each sample will be deposited in an official nucleic acid database like EMBL and contain the documentation requested by the depository.
The bacterial cultures isolated in the study will be preserved as frozen stocks in glycerol in our laboratory collection at -80°C, and a file with strain details (in Access) (identification/source of origin/main characteristics/storage medium/revival guide/location in the freezer) will be maintained. We have a dedicated on line own system for that.
Mineralization kinetic data will be provided including a detailed description of the method for collecting the data.

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 PET samples, will be provided with a clear description of the methods 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, physico-chemical characteristics). In addition it will provide where the related nucleic acid 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 & BACK UP DURING THE FWO PROJECT

Where will the data be stored?

How is back up of the data provided?

The data will be stored on the One Drive KU Leuven 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?

Costs are not expected to be high. In case required, costs will be covered by the lab budget whenever needed.

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

The access to the folder containing all project related data is restricted to the project researchers and access can only be granted by the PI.

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, ...).

Digital data: 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.All data will be preserved for at least 5 years after completion of the project.
Physical data: Freezer stocks of microbial strains will be maintained in the lab at -80°C and will be available upon request. Samples taken from the field or from experiments will be stored in freezers (-20°C) for up to 5 years after the end of the project.

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

After the project, hard copy data notebooks will remain at the host lab, and all the research data will be archived in KU Leuven's public Research Data Repository (RDR) for a minimum of 10 years, ensuring scientific integrity, conform to the KU Leuven RDM policy and complying with the FAIR principles.

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

No direct idea but not expected to be high. If costs, they will be covered by related research project funds.

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 deposited at official repositories for sequencing results) will be deposited in a cvs format in KU Leuven RDR under a CC-BY license. The nucleic acid sequence data will be available 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 after the end of the project in case published. Others will be added upon publication. If not published within 1,5 years of project completion, 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 high costs expected. In case, costs will be covered by research funds acquired by the promotor.

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?

Hazrat Noor who acts as the PhD student working on the project and PI Dirk Springael.

Who will be responsible for data storage & back up during the project?
Hazrat Noor who acts as the PhD student working on the project and PI Dirk Springael.

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

Who bears the end responsibility for updating & implementing this DMP?
PI Dirk Springael.