

Research on bacteriophage based biocontrol of bacterial canker in cherry

DPIA

DPIA

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

- No

Research on bacteriophage based biocontrol of bacterial canker in cherry

GDPR

GDPR

Have you registered personal data processing activities for this project?

- No

Research on bacteriophage based biocontrol of bacterial canker in cherry

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)

The project will reuse existing data, bacterial strains and phages which are in the current collection of KU Leuven and/or ILVO and their corresponding sequence and meta data currently available as FastaQ and TSV files on servers of KU Leuven and/or ILVO.

New data:

- New bacterial strains and phages stored at -80°C freezer, corresponding sequence data will be stored as GB and FastaQ (400 MB) and meta data as TSV files (200 KB).
- Images will be stored as JPG (1250 MB).
- Cloning vectors will be stored in a bacterial strain at -80°C at their sequence as GB files (50 MB).
- Numerical data will be stored as TSV file (200 KB)
- Statistic analysis data will be stored as Rmarkdown files (100 KB).
- Bioinformatic pipelines will be made using GitHub.
- A digital lab book will used and made accessible by both KU Leuven and ILVO.

Both KU Leuven and ILVO (in agreement) hold the rights to all data. A MTA has been set up with the University of Warwick for the use of their phages.

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)

1. Designation of responsible person (KU Leuven: Jeroen Wagemans, ILVO: Jolien Venneman)
2. Storage capacity/repository
 - during the research
 - KU Leuven: Bacterial and phage strains: -80°C freezer, digital data: Reserach group server and back up server.
 - ILVO: Bacterial and phage strains: -80°C freezer, digital data: Reserach group server and back up server.
 - Sequences will be made public available and deposited at NCBI
 - Bioinformatic pipelines will be deposited on Github.
 - Interesting bacterial strains will be deposited to BCCM
 - after the research
 - KU Leuven: Bacterial and phage strains: -80°C freezer, digital data: Long term storage server
 - ILVO: Bacterial and phage strains: -80°C freezer, digital data: Long term storage server

Both KU Leuven and ILVO will have access to all data and material, this extends after the project. All data will be kept up to 5 years. After 5 years, the bacteria and phages with the corresponding metadata and sequences will be kept. Other data (images, numerical data and statistic analysis) will not be kept unless considered necessary for/be potential publications.

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)

Not applicable

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)

Not applicable

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

Not applicable

Research on bacteriophage based biocontrol of bacterial canker in cherry

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
Phage isolates	Phages isolated from cherry/pear orchards	New	Physical				1 Eppendorf Tube per phage
Numerical data	Results from twig/leaf/fruit/tree infection assays and microbial assays	New	Digital	Experimental	.csv	<100MB	
Pictures	Pictures from infection assays	New	Digital	Experimental	.jpg	<1GB	
Genomic data	Results from sequencing, annotating, microbiome ...	New	Digital	Experimental	.fasta .fastaqz .tbl .gbk .gff .TSV	<1GB	
Mutants	Genetically modified phages and bacteria	New	Physical				1 Eppendorf Tube per phage
Bacterial collection ILVO	Bacterial collection of ILVO mainly isolated in the CTRL-PSEUDO project	Reuse	Physical				1 Eppendorf Tube per strain
Cloning vectors	Plasmide used to clone	New	Physical				1 Eppendorf Tube per vector
Statistic analysis	statistical analysis of the experiments	New	Digital	Other	.Rproj	<100MB	
Bioinformatic pipelines	Pipelines made and used to analyse data	New	Digital	Software	.txt	<100MB	
Lab books	summary of experimental research	New	Digital	Experimental	.docx	<100MB	

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:

Bacterial collection ILVO, obtained during CTRL-PSEUDO VLAIO-project: HBC.2021.1064

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.

- Yes

The phage and bacterial isolates and engineered strains can be used for commercial valorization. It can be licensed out to industrial partners.

Novel techniques developed may be patented.

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.

- Yes

Phages from the University of Warwick, MTA has been set up.

Potentially bacterial strains from the university of Birmingham, a MTA will be put in place.

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).

A detailed lab book is maintained containing information on the methods used to generate and analyze data:

- protocols to describe how the data is obtained
- code used to analyze the data
- graphs/tables to interpret the data

All data will be timestamped, enabling easy lookup in the lab books how it was acquired.

Excel files will have sufficient annotations and an explanatory sheet to make the experimental build-ups and the gathered data

understandable.

If possible, the data will be linked to an R Project file, where annotations will be made in an R markdown file to provide guidance for the correct use of the data.

Snakemake pipelines will be accompanied by a README file and deposited in github.

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.

- No

Bacterial strains will be included in the BioNumerics database and registered in Excel collection sheets.

Phages and mutants will have a TSV file for metadata.

Numeric data will have a tab containing metadata.

All Excel files will have the same template to enhance interpretation.

For the genome sequence data generated, samples are registered and results and metadata recorded in standard bioinformatics file formats according to established guidelines agreed upon at ILVO and KUL level.

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

Where will the data be stored?

During the PhD:

- Phage strains will be stored at 4°C as PEG stock in KUL and ILVO.
- Bacterial strains will be stored at -80°C as glycerol stock in KUL and ILVO.
- Digital data will be stored in the shared network drive (J:) from KU Leuven and the shared ILVO folder 'ILVO\Fytobacteriologie - Documenten\General\Projecten\Phage therapy'
- At ILVO, sequencing results will be stored on the Genomics server, in the 'Fytopathologie (W:)\Sequencing' folder and in the BioNumerics database.

After the PhD:

- Phage strains will be stored at 4°C as PEG stock in KU Leuven and ILVO
- Bacterial strains will be stored at -80°C as glycerol stock in KU Leuven and ILVO.
- Digital data will be stored in the large volume storage from KU Leuven (K:) and ILVO.
- Published data will be deposited in public databases/repositories (such as GenBank for sequencing data)

How will the data be backed up?

Data will be backed-up according to the standard back-up procedures provided by KU Leuven ICTS. This involves automatic version management of the files. Version management is done using "snapshot" technology, where the previous versions of the changed files are kept online in a snapshot on the same storage system.

- by default, 1 snapshot is taken daily and is kept for 14 days. So you can go back to previous versions of the file up to 14 days.
- end users can restore older files themselves from within their Windows PC via the "previous versions | previous versions" functionality.

At ILVO, the backup process includes:

- A (working) daily backup where versions are kept up to 3 months back.
- A (working) daily file copy where only the latest version is available.
- A monthly backup to offsite tape.
- The ILVO sharepoint services in the cloud rely on the MS backup policy. Deleted documents can be recovered up to 3 months back.

**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

Physical data: Sufficient space at 4°C and -80°C has been allocated for this project.

Digital data: the large storage volume (K:) at KUL has a storage capacity of 6TB. If more space is needed, extra storage space can be bought per TB. Also the storage capacity of ILVO's drives and Genomics server is more than sufficient for the expected data.

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

Physical data can only be accessed by authorized personnel (utilizing batch systems in both host institutions).

Digital data access is controlled through standard IT practices, restricting access to relevant personnel only. Data is always transferred over encrypted connections.

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

The costs for storing and backup of large datasets equals:

- €104,42/TB each year at KU Leuven.
- ILVO server: €55/TB per year
- ILVO online storage: 132€/TB per year

Upon publication, relevant data will be made available in either appropriate repositories or e.g. using generic services such as zenodo.org, dryad, etc. These services will ensure long-term storage.

Costs will be covered by the research groups.

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...).

Computational data (e.g., sequencing data, lab books,...) and biological data

- stored for at least 10 years according to the KU Leuven RDM policy
- stored for at least 5 years at ILVO

For both datatypes a selection will be made for data that is either hard/costly to replicate or has potential future applications (such as engineered strains)

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

Computational data

- large volume storage drive (K:-drive) of KU Leuven
- ILVO:
 - Sequencing data: 'Fytopathologie (W:)\Sequencing', Genomics server, BioNumerics databases
 - Finalized projects from the folder 'ILVO\Fytobacteriologie -Documenten\General\Projecten\Phage therapy': archived on the ILVO storage drive

Biological samples

- At 4°C/-80°C freezers in KU Leuven and ILVO

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

Most of the data will be made available through online repositories, reducing the storage costs to below 1TB/year (<€200/TB/year).

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
- Yes, in a restricted access repository (after approval, institutional access only, ...)

If the data is published in a peer-reviewed journal, this publication will have a DOI.

Sequencing data will be made available through NCBI.

Scripts/bioinformatic workflows will be documented (for example in markdown language as README.md), and the documentation including the scripts will be uploaded on a GitLab repository. Optionally, the GitLab repository with the scripts can be deposited at Zenodo to generate a DOI for the scripts at the time of publication.

Other datasets accompanying the paper (but not deposited in public archives as stated above) will be submitted to Dryad to generate a DOI.

Upon request, other data can be shared with interested parties.

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

Unless otherwise stated data will be made available Open Access. For data shared directly by the PI, a material transfer agreement (and a nondisclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.

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.

- Yes, Intellectual Property Rights

The (isolated or engineered) phage and bacterial strains can be used for commercial valorization. It can be licensed out to industrial partners for biocontrol purposes.

Techniques with patent potential will be protected through the KU Leuven Research & Development Technology Transfer Office.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Data will be made available through public repositories:

- Genomic data will be deposited in NCBI
- Manuscripts will be published in Open Access journals or made Green Open Access using the KU Leuven repository
- Scripts, datasets, etc accompanying a manuscript will be uploaded to a GitLab repository, Zenodo, etc
- Other metadata for which no specific tools is available can be stored on the KU Leuven RDR platform.

When will the data be made available?

Upon publication of an associated research article.

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

Data submitted to a public repository will adhere to the licenses being put in place by the repositories themselves. Data that is being made public outside such public repositories will be released under the CC-BY-4.0 license (or the CC-BY-SA-4.0 for specific cases).

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

DOI number will be provided by the repositories where the data will be submitted.

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

The proposed repositories don't charge a fee for uploading data

6. Responsibilities

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

Viktor Devlieghere (PhD student), under supervision of Jeroen Wagemans (co-promotor), Jolien Venneman (co-promoter) and Rob Lavigne (promotor)

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

Viktor Devlieghere (PhD student), under supervision of Jeroen Wagemans (co-promotor), Jolien Venneman (co-promoter) and Rob Lavigne (promotor)

Who will manage data preservation and sharing?

Rob Lavigne (PI)

Who will update and implement this DMP?

Viktor Devlieghere (PhD student), under supervision of Jeroen Wagemans (co-promotor), Jolien Venneman (co-promoter) and Rob Lavigne (promotor)