Plan Overview

A Data Management Plan created using DMPonline.be

Title: Metabolic Profiling of bioactive Antimicrobial Metabolites from Some of the selected Tanzanian Medicinal Plants: Bio-guided fractionation and Untargeted Metabolomics approaches

Creator: Meshack Damian Lugoba

Principal Investigator: n.n.

Data Manager: n.n., Meshack Damian Lugoba

Project Administrator: n.n., Meshack Damian Lugoba

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: n.n. n.n.

Data Manager: n.n. n.n., Meshack Damian Lugoba https://orcid.org/0000-0001-8642-3343

Project abstract:

Infectious diseases have long been among the main causes of morbidity and mortality in the general population worldwide, and are more prevalent in developing countries. The presence of antimicrobial resistance strains to current available effective antibiotics threatens efforts to eradicate microbial infections. There is an urgent need for the discovery and development of new antimicrobial drugs to cope with the need and avoid running out of treatment options. Natural sources such as plants continue to be a better option as sources of newer or alternative antimicrobials, since they are cheaply available and offer a wide array of compounds that can serve as antimicrobials or at least precursor molecules. The Tanzanian nature is a rich source of different plant species, some of which are known to be a source of interesting compounds for healing and preventing diseases such as microbial infections.

ID: 210566

Start date: 01-04-2023

End date: 31-03-2027

Last modified: 24-10-2024

Metabolic Profiling of bioactive Antimicrobial Metabolites from Some of the selected Tanzanian Medicinal Plants: Bio-guided fractionation and Untargeted Metabolomics approaches

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Plant materials	Plant materials (root barks, stem barks, leaves, and fruits) are identified, collected, and prepared for extraction and endophytic identification	N	P & D	I & Other: Observational	.TIFF .jpg .png	<1GB	
Extracts (samples)	Extracts (samples) are obtained from the prepared plant materials (root barks, stem barks, leaves, and fruits). The plant samples are stored at 4°C	N	P & D	I & Other: Experimental/Observational	.TIFF .jpg .png	<1GB	Some vials of 100 mL
Endophytes	Endophytes are identified from fruits and stored at -80°C.	N	P & D	I & Other: Experimental/Observational	.TIFF .jpg .png .xlsx .csv	<1GB	Some boxes of 2 mL glycerol stocks
Antimicrobial Susceptibility Testing (AST) for antimicrobial activity screenings (qualitative assays)	To give a first indication on whether the plant samples have activity or not against standard bacterial strains. The size of the halos is directly related to the activity produced.	N	D	I & Other: Observational	.TIFF .jpg .png	<1GB	
MIC Testing (quantitative assays)	To look at the antibacterial spectrum (strength) and potency of the plant samples. The MICs are measured and stored in an Excel file.	N	D	I, N, T & Other: Experimental/Observational	.TIFF .jpg .png .xlsx	<1GB	

DNA agarose electrophoresis gels	To visualize DNA and verify whether a PCR reaction succeeded or not or whether DNA is present or not. A digital image of the gel is made to store.	N	D	I & Other: Experimental/Observational	.TIFF .jpg .png	<1GB	
Sequencing data	DNAs of the isolated endophytes are sequenced for endophytic identification. This is done by sending the DNA samples for Sanger sequencing to Eurofins Genomics. They send the data back as .abi file to process. The data is processed in SnapGene and stored as .dna file. An inventory and descriptions are stored in an Excel file.	N	D	Other: DNA sequences from an external company, Metadata: excel file with inventory and description of sequencing data	Received as: .abi and stored as: .dna, .xlsx .csv	<100GB <1GB	
LC-MS data	LC-MS analyses are performed to verify the presence, molecular weight and quantity of the compounds of interest.	N	D	Other: LC-MS chromatograms	.FID	<5TB	
HPLC data	Chromatograms from the HPLC purification of compounds	N	D	Other: HPLC chromatograms	.pdf	<1TB	
NMR data	NMR spectroscopic will be performed for structural analyses of the compounds	N	D	Other: NMR spectra	.FID	<100GB	
Lab book notes and reports	Description of all of the experiments performed in the lab and the results obtained	N	D	N, T & Other: Compiled/aggregated data	.docx, .pdf	<1GB	

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:

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

There are ethical issues concerning the transportation and use of plants from Tanzania. An ethical approval with reference number NIMR/HQ/R.8a/Vol.IX/4484 is obtained from Tanzania together with the Material Transfer Agreement (MTA)

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- 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.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or 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.

- No

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

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).

All data will be collected and documented on KU Leuven servers and personal One drive. An inventory of all data and a description will be collected in an Excel file. All work performed in the lab will be saved as an Electronic Lab Notebook (ELN). This notebook will contain all information on methodology, protocols, results, and conclusions. The ELN will be ordered chronologically, with a title for every date and a subtitle for every experiment.

All data will be accompanied by appropriate documentation styles and tabs outlining procedures and standards followed during data collection and processing. Standard operating procedures and guidelines to be followed during data collection, processing, and laboratory work will be well documented in the lab.

General protocols and standard operating procedures will be collected in a dedicated folder on a Shared network drive (J:) at KU Leuven.

Raw LC-MS, HPLC, and NMR data will be collected alongside a .txt file (README) describing the experimental conditions.
All the images with their explanation will be saved in separate folders in the ELN.
Background information will be stored in an Excel file and the Electronic Lab Notebook.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

The aforementioned methods of data documentation clearly ensure the exact documentation and finding of the data. Any specific data related to this project can be retrieved from the above documentation methods. In short, the metadata for most of the experiments are generated in a descriptive format that can be interpreted easily by other people in the future. The metadata for all NMR, LC-MS, and UV spectroscopic data will be available within the file format they will be stored in.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- OneDrive (KU Leuven)
- Large Volume Storage

Data will be stored on a network drive (J:) which is backed up by the ICTS service at KU Leuven. This includes monthly, time-stamped copies of each lab notebook, images, and spectroscopic data. Additional copies will be made and kept on personal devices. The network drive is only accessible to group members. Once a researcher leaves the lab, their data will be transferred to a large-volume network archive drive. All LC-MS data will be stored in duplicate on dedicated external hard drives. Bacterial strains and DNA plasmids will be kept in duplicate at -80 and -20°C freezers, respectively.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

The data will be stored on the university's central servers with automatic daily backup procedures. Copies of the LC-MS data will be stored on dedicated external hard drives which is a researcher-based personal data back-up.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

KU Leuven provides sufficient storage and backup capacity during and after the project. Dedicated external hard drives of 5 and 10 TB are available for storing the LCMS, HPLC, and NMR data.

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

The Shared network drive (J:) of the KU Leuven is only accessible to research group members. Their access is determined by their

KU Leuven personnel number. The drive has a high level of security, where the two-factor authentication of KU Leuven provides secure storage against other unauthorized
personnel access. Furthermore, the data on the external hard drives will be protected by a password.

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

The KU Leuven shared network drive costs € 503,66/ TB / year. Additionally, a large volume storage drive from the KU Leuven costs € 104,42 / TB / year. This drive will be used to store all the large files. These costs, along with the costs for the external hard drives (approx. € 200) will be covered by the project budget.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

The Principal Investigator (Prof. Dr. Joleen Masschelein) will take responsibility for data preservation after the end of the research project. All the data will be stored in an online lab management system with an integrated Electronic Lab Notebook. The data will be preserved for 10 years according to KU Leuven RDM policy.

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

- Large Volume Storage (longterm for large volumes)
- KU Leuven RDR
- Shared network drive (J-drive)

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

The KU Leuven large-volume network archive drive costs € 104,42 / TB / year. These costs and the costs of the external hard drives (approximately € 200) will be covered by the project budget.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as restricted data (upon approval, or institutional access only)

During the project, all data will be stored on the shared network drive of the KU Leuven. Lab members and people participating in the project can get access to the data stored on this drive based on their personnel numbers.

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

Under the control of the Principal Investigator (Prof. Dr. Joleen Masschelein), only researchers participating in the project and lab

members will be able to access the data before publishing. Upon publication, everyone will be able to access the data.

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 per dataset or data type where appropriate.

- Yes, intellectual property rights

If novel antimicrobial compounds are discovered, a patent application will be filed. This may temporarily restrict the sharing of data.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

Genome data will be made available in GenBank (NCBI).

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- Data Transfer Agreement (restricted data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

A unique identifier will be added to the published data. Additionally, genome data deposited in GenBank will also get a unique accession number.

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

No expected costs for data sharing.

Responsibilities

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

PhD researcher (Meshack Damian Lugoba), Principal Investigator (Prof. Dr. Joleen Masschelein), and co-Principal Investigator (Dr. Joseph Sempombe).

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

PhD researcher (Meshack Damian Lugoba), Principal Investigator (Prof. Dr. Joleen Masschelein), and co-Principal Investigator (Dr. Joseph Sempombe).

Who will manage data preservation and sharing?

PhD researcher (Meshack Damian Lugoba), Principal Investigator (Prof. Dr. Joleen Masschelein), and co-Principal Investigator (Dr. Joseph Sempombe).

Who will update and implement this DMP?

The researcher will remain responsible for updating and implementing the DMP (day-to-day management), while the Principal Investigator will be responsible for overall data management.