Plan Overview

A Data Management Plan created using DMPonline.be

Title: The role of the microbiome and epigenetic factors in the local adaption of medically important snails

Creator: Ruben Schols

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Project abstract:

Snail-borne diseases (SBDs) are caused by trematode parasites and affect more than 300 million people worldwide, while also affecting wildlife and livestock. The most important of these diseases is schistosomiasis, which ranks second only to malaria in terms of prevalence and morbidity. Despite the availability of adequate tools for diagnosis and the concerted control efforts to date, schistosomiasis continues to (re-)emerge with unexpected distributions and unprecedented intensities. The complexity of this disease involving human, animal and environmental health necessitates a *One Health* approach. Furthermore, global change causes shifts in vector-borne disease distribution and the first cases of schistosomiasis have already been reported in Europe. Interestingly, current research provides increasing evidence for a potential link between the microbiome of the snail intermediate host and its resistance to schistosome infection. Moreover, the microbiome has been implied in host epigenome modifications. Here we want to test such a relationship, as well as the impact on temperature tolerance, by analysing microbial diversity through 16S metabarcoding, microbiome functionality through shotgun sequencing, and the molecular basis of phenotypic plasticity through ATAC-seq based on field samples and transplant experiments. Thereby, revealing a potential interaction and **the relative contribution of the microbiome and epigenetic factors on the adaptive capacity of intermediate snail hosts to climate change and disease resistance**. Our results could alter disease control measures towards sustainable alternatives, limiting the use of detrimental molluscicides, and prepare Europe for a changing world.

ID: 211715

Start date: 03-12-2024

End date: 31-10-2025

Last modified: 16-12-2024

The role of the microbiome and epigenetic factors in the local adaption of medically important snails

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	
µCT-images	Images associated with the µCT paper	E	D	I	various	<1TB	
µCT-sequences	sequences associated with the µCT paper	E	D	Other - Sequences	ab1	<1GB	
µCT-DNA extracts	DNA extracts associated with the µCT paper	E	P	DNA			30 1.5ml eppendorf tubes
Columella paper	16S rRNA gene metabarcoding	E	D	Other - Sequences	.fastq	<100GB	
Columella paper	rscripts and metadata	E	D	Other - rscript, excel,...	.R & others	<1GB	
TRISS-images	Images associated with the TRISS paper	E	D	I	various	<100GB	
TRISS-sequences	sequences associated with the TRISS paper	E	D	Other - Sequences	ab1 and fastq	<100GB	
TRISS-DNA extracts	DNA extracts associated with the TRISS paper	E	P	DNA			400 1.5ml eppendorf tubes & 5 96-well plates
Temporal-sequences	sequences associated with the temporal paper	E	D	Other - Sequences	fastq	<100GB	

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:

Existing DNA extracts from my PhD project are reused to continue data analysis and complete missing data when required for publication. These will be part of the collection of the Royal Museum for Central Africa (RMCA) in Tervuren and are currently actively being used.

Existing sequencing output is stored on the Biocol/parasitology internal server of RMCA, my OneDrive account, my local PC, a local SSD, and on the shared J-drive:KOR-Biologie/People's folders/9_Ruben Schols. Similarly, images/pictures, metadata, preliminary versions of the manuscript are all stored on the same backup locations.

The metabarcoding data set generated and analyzed during the Temporal-sequences project is freely available on the European Nucleotide Service (ENA), under the bioproject PRJEB77058.

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.

- No

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.

- Yes

All DNA extracts will become part of the parmanent collection of RMCA, thereby falling under loan agreements for future use.

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 DNA extracts that will be curated in the permanent collection of RMCA require a standard template of metadata to ensure all available metadata is present.
All sequencing files will be organised by sequencing run (incl. date) and contain the used primers and sample name in the file name. Each folder is then associated with a metadata.xlx file to provide any available metadata for each sample. Experimental outline and other material and methods are listed in the associated version of the manuscript and are kept up to date while the experiment progresses.
Similarly, all images contain the sample name and are sorted per research project. All associated metadata is then linked in the associated .xlx file of that research project.
All information is recorded on the previously mentioned locations: the Biocol/parasitology internal server of RMCA, my OneDrive account, my local PC, a local SSD, and on the shared J-drive:KOR-Biologie/People's folders/9_Ruben Schols.

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.

- Yes

We use the metadata .xlx template from RMCA to list data associated with DNA extracts, samples, and images. Earlier metabarcoding datasets have been uploaded to the European Nucleotide Service (ENA). However, future sequencing data for publications will be uploaded to KU Leuven Research Data Repository, following DataCite as a metadata standard.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Other (specify below)

It is additionally stored on my personal laptop, SSD disk, OneDrive from RMCA, and the internal servers of RMCA (nbiocol).

How will the data be backed up?

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

backups will be made on a personal SSD disk, OneDrive from RMCA, the internal servers of RMCA (nbiocol), and the Shared network drive (J-drive).
RMCA OneDrive and internal servers are equally well protected as KUL with an authenticator app.

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

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

This is indeed a valid concern for the Shared network drive (J-drive) where other members of the group may access the data. Nevertheless, only restricted KUL personnel can access these folders. Moreover, RMCA-linked OneDrive is a valid backup with personalized access and nbiocol only providing writing access to a very limited set of employees. The versious backup methods mentioned above enablea consistent backup that can be used to validate any suspicious data.

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

All databack methods have been paid for in advance or are part of larger functioning authorities that ensure funding into the future (KUL & RMCA).

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

All DNA extracts will be dried and integrated into the permanent collection of RMCA.
Digital files linked to the project will be kept for a minimum of 10 years for future use.

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

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

All data linked to this project will be centralized on the Shared network drive (J-drive). Currently, my dataset of amplicon sequencing has been uploaded to ENA (European Nucleotide Archive) for a manuscript under review. All other final datasets for publication will also be stored on ENA while non-sequencing files will be stored on the KU Leuven RDR to obtain unique identifiers. All data is also stored on the nbiocol internal server of RMCA.

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

The long term datastorage platforms used in this project are part of larger enterprises to ensure its persistence (KUL & RMCA)

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)
- Yes, as open data

Sequencing data, analysis files, and metadata will all be made publically available upon publication of the study.
Images not included as part of a publication will be stored on the RMCA's internal server, and will be restrictedly available until it has been uploaded on the publically accessible DARWIN website.
DNA extracts will be preserved but will only be available after approval of RMCA curators.

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

DNA extracts are accessible at the RMCA upon valid research requests. Images are available upon any request.

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.

- No

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)
- Other data repository (specify below)

The sequencing data of the manuscript under review is already publically avialable on the European Nucleotide Archive (ENA) as bioproject PRJEB77058.
Future amplicon sequencing data will be uploaded and made publically available through ENA.
Associated files will be uploaded on the KU Leuven RDR.

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.

- CC-BY 4.0 (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

One dataset already has a PID: bioproject PRJEB77058 on ENA
Future metabarcoding datasets will also be uploaded here while assocaited smaller files will be stored on the KULeuven RDR

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

We ensure to use free datasharing platforms.

Responsibilities

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

I will

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

I will

Who will manage data preservation and sharing?

I will

Who will update and implement this DMP?

I will

Created using DMPonline.be. Last modified 16 December 2024

7 of 7