Coping with geographic mosaics of stressors: unravelling the interplay of genetic and nongenetic trait variation in natural landscapes

A Data Management Plan created using DMPonline.be

Creators: Steven Van Belleghem, Robby Stoks, n.n. n.n., n.n. n.n.

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: n.n. n.n., n.n. n.n., Robby Stoks, Steven Van Belleghem

Grant number / URL: C16/23/003

ID: 202359

Start date: 01-10-2023

End date: 30-09-2029

Project abstract:

To improve our ability to predict responses of populations to global change, we aim to increase our knowledge of how populations cope with multiple stressors in a natural landscape through genetic and/or non-genetic (plastic, including microbiome-mediated) trait changes, embedded in an evolving metacommunity framework. Therefore, we will (1) study how populations of the water flea *Daphnia magna* and microbial communities respond to key environmental gradients (fish predation and pesticide exposure) in a heterogeneous landscape; (2) identify the molecular basis of plasticity and genetic adaptation in response to multiple stressors; (3) run experimental evolution trials with manipulated levels of genetic variation and apply resurrection ecology; and (4) further develop the concept and tools of the evolving metacommunity framework to integrate the data across the different endpoints, and synthesize a multistressor landscape perspective in global change research.

Last modified: 30-11-2023

Coping with geographic mosaics of stressors: unravelling the interplay of genetic and nongenetic trait variation in natural landscapes

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	
WP1.1A	Characterization 20 study ponds	N	D	N	.csv	<1GB	NA
WP1.1B	Dataset fish eDNA	N	D	N	.csv	<1GB	NA
WP1.1C	Free-living bacterioplankton samples of the study ponds for metagenomics	N	P	biological samples	NA	NA	20 filters (0.22 µm)
WP1.1D	*Daphnia* samples of the study ponds for landscape genomics	N	P	biological samples	NA	NA	20 eppendorf tubes
W1.1E	*Daphnia* samples of the study ponds for gut microbiome metagenomics	N	P	biological samples	NA	NA	20 eppendorf tubes
WP1.1F	*Daphnia* clones isolated from the study ponds (20 per pond)	N	P	biological samples	NA	NA	400 glass vials 500 ml
WP1.1G	zooplankton samples of the study ponds	N	P	biological samples	NA	NA	20 bottles 500 ml
WP1.1H	Density, community and size distribution zooplankton of the study ponds	N	D	N	.csv	<1GB	NA
WP1.2A	Re-sequenced genomes of 400 *Daphnia* clones	N	D	DNA sequences	.fastq	<5TB	NA
WP1.2B	Genetic population structure, demography and genome divergence 20 *Daphnia* study populations	N	D	N	.txt	<1GB	NA
WP1.2C	Phenotypic association analyses	N	D	N	.txt	<1GB	NA
WP1.3A	Landscape metagenetics (16S rRNA sequencing) of bacterioplankton of the study ponds (20 samples)	N	D	Sequencing data	.fastq	<100 GB	NA
WP1.3B	Landscape metagenetics (16S rRNA sequencing) of gut microbiome of the 200 *Daphnia* clones in lab culture + 20 pooled *Daphnia* field samples	N	D	Sequencing data	.fastq	<100 GB	NA
WP1.3C	Landscape metagenomics of free-living bacterioplankton of the 20 study ponds	N	D	Sequencing data	.fastq	<1TB	NA
WP1.3D	Landscape metagenomics of pooled gut microbiome of *Daphnia* of the 20 study ponds	N	D	Sequencing data	.fastq	<1TB	NA
WP1.4	Multistressor common garden experiment with *Daphnia*: phenotyping	N	D	N	.csv	<1GB	NA
WP1.5A	Multistressor gut microbiome transplant experiment with *Daphnia*: phenotyping	N	D	N	.csv	<1GB	NA
WP1.5B	Multistressor gut microbiome transplant experiment with *Daphnia*: metagenetic characterization gut microbiome (16S rRNA sequencing)	N	D	Sequencing data	.fastq	<100 GB	NA
WP1.6	Integrated analysis data collected in WP1	N	D	N, Model	.csv R code	<1GB	NA
WP2.1A	Construction 72 RNA libraries of *Daphnia*	N	D	Sequencing data	.fastq	<1TB	NA
WP2.1B	Construction of 168 ATAC-seq libraries of *Daphnia*	N	D	Sequencing data	.fastq	<1TB	NA
WP2.1C	Testing candidate genes (CRISPR/Cas9 and RNAi)	N	D	Genes and primers	.fastq	<1Gb	NA
WP2.2	Comparison plastic gene regulatory changes to genetic changes	N	D	N	.txt and .bed	<1Gb	NA
WP2.3A	Targeted gut microbiome transplant experiment: RNA-sequencing	N	D	sequencing data	.fastq	<1TB	NA
WP2.3B	Targeted gut microbiome transplant experiment: ATAC-sequencing	N	D	sequencing data	.fastq	<5TB	NA
WP3.1A	Production dormant eggs in experiment "Standing genetic variation"	N	P	biological samples	NA	NA	90 sets of dormant eggs stored in eppendorf tubes
WP3.1B	Full-genome resequencing of pooled samples *Daphnia* dormant eggs	N	D	Sequencing data	.fastq	<5TB	NA
WP3.2A	Dataset on *Daphnia* population densities in experiment on "Evolutionary potential"	N	D	N	.csv	<1GB	NA

WP3.2B	Full-genome resequencing of pooled samples of 180 experimental *Daphnia* populations	N	D	Sequencing data	.fastq	<5TB	NA
WP4.1A	Taking sediment cores of the study ponds	N	P	Sediment samples	NA	NA	20 cores (2x 5cm m cylinder, 5cm diameter)
WP4.1B	Characterization of the sediment cores (dating, fish presence, zooplankton composition)	N	D	N	.csv	<1GB	NA
WP4.1C	Isolation *Daphnia* from subset of the sediment cores	N	P	biological samples	NA	NA	400 vials 500 ml
WP4.1D	Multistressor common garden experiment with isolated *Daphnia* from the sediment core: phenotyping	N	D	N	.csv	<1GB	NA
WP4.2	Full-genome resequencing of 40 pooled samples of *Daphnia* isolated from the sediment core	N	D	Sequencing data	.fastq	<5TB	NA
WP5.1A	Feedback on community and ecosystem features: development of more refined analytical approaches	N	D	M	.csv model language code	<1GB	NA
WP5.1B	Feedback on community and ecosystem features: Pattern analysis	N	D	N	.csv .txt	<1GB	NA
WP5.2A	Conceptual synthesis and generalization: further development of existing models	N	D	M	.csv model language code	<1GB	NA
WP5.2B	Conceptual synthesis and generalization: comparison to existing data	N	D	N	.csv .txt	<1GB	NA

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:

To select the study sites for current project we will make use of existing databases:
Manscape, https://www.gbif.org/dataset/312bb844-4980-4b1a-a64b-e79f9ac083a4
Pondscape, https://www.gbif.org/es/dataset/a621b3ba-8415-41f2-a4af-7ec9511ae868
Midden-Limburg, https://www.gbif.org/dataset/3236cdc4-2f4d-4bd6-b53f-54fd8d8a0aa8
Tommelen, https://data.freshwaterbiodiversity.eu/ipt/resource?r=tommelen

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

For the invertebrate *Daphnia magna* no approval for experimental work is needed. For the experiments where we expose *Daphnia* to fish kairomones, we keep stickleback in separate aquaria and collect medium of this water to be used to expose the *Daphnia*. A responsible of ECD let us know by email that in this case (as no experiments on the fish itself is done) also no approval by ECD is required.

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).

For biological samples, these will be labelled and a separate text file will be added that identifies each sample (source population, experimental group, clone, etc).
For phenotypic data stored in .csv format columns will be added that identifies each sample (source population, experimental group, clone, etc). Units of measurement will be included in the .csv file. A separate sheet will be added with information how derived variables were calculated and which variables have been used in a manuscript.
For metagenetic data (composition free-living bacterioplankton and gut microbiome) stored in .fastq format, a separate text file will contain the metadata related to sample identification (source, experimental treatment, clone, etc).
For RNAseq data, the sample code is included in the .fastq file. A separate text file will contain the metadata related to sample identification (source, experimental treatment, clone, etc).
Metagenomic data (composition free-living bacterioplankton and gut microbiome) will be stored in .fastq format and will be deposited in NCBI together with the metadata related to sample identification (source, experimental treatment, clone, etc). MG-RAST (an open-source web application server that suggests automatic phylogenetic and functional analysis of metagenomes) will be used to store and share processed data results.
For the modelling part, similarly as for phenotypic data, the .csv-file containing the main modelling results will comprise a column that uniquely identifies each model run, along with the actual model output. A second .csv-file features a table with the evaluated parameter choices used for each model run. A third .txt-file includes descriptions of all model parameters and output columns in plain text. The high quality, concise & well documented code will be written in R, Python or Julia.

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

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- Personal network drive (I-drive)
- Other (specify below)
- Large Volume Storage

Each *Daphnia* clone (biological sample) will be duplicated and stored in separate temperature-controlled rooms.
Data and metadata on pond characteristics and phenotypic traits, and grazing rate will also be stored on external harddrives, and in the Dryad Digital Repository when the associated manuscripts are accepted.
Data and metadata on composition of bacteria (including gut microbiome) communities (metagenetic data) will be stored on personal computers and on OneDrive. When the associated manuscripts are accepted the data will be deposited in the Sequence Read Archive (SRA).
Data and metadata on gene expression (RNA-seq) will be stored on external hard disks, and on the High Performance Computing (HPC) cluster of the KU Leuven as staging storage. When the associated manuscripts are accepted the data will be deposited in the Gene Expression Omnibus (GEO) data repository managed by the National Center for Biotechnology Information (NCBI).
Metagenomic data will be stored on the KU Leuven VSC Tier-1 Data Storage
Modelling code will be stored on GitHub.

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

External harddrives
KU Leuven VSC Tier-1 Data Storage
Public submission of sequence read data to NVBI (with embargo until publication)

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.

- No (explain solution below)

All *Daphnia* clones will be stored in dedicated temperature-controlled rooms. Sediment samples will be stored in a cold room.
The large amount of (meta)genomic data will be shared and stored using KU Leuven's access to Flemish Tier-1 supercomputer infrastructure (VSC irods irods.hpc.kuleuven.be). Genomic data and corresponding metadata will be deposited on NCBI/SRA and made publicly available when published.
External harddrives will be purchased (~150 euro for 5Tb)

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

Biological samples are stored in the lab space of our teams, access to the labs outside office hours needs an activated personal badge.
External hardddrives will be stored safely in the office of the responsible PI.
All other data are stored in folders with limited access. The multifactor authentication with the KU Leuven Authenticator app will need to be used for access to the folders.
The KU Leuven VSC Tier-1 Data Storage has a strict permissions system (read/write/modify), which will be limited to the primary collectors of the data.

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

NCBI: Free
KU Leuven VSC Tier-1 Data Storage: ~35 euro per Tb. 30Tb is supported by the FWO Data component of the Flemish Tier-1 supercomputing platform until 2027.

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

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

- Other (specify below)

Biological samples will be stored in temperature-controlled rooms (biological samples) and in a cold room (sediment samples).
Data and metadata on pond characteristics and phenotypic traits will be archived in the Dryad Digital Repository when the associated manuscripts are accepted.
Data and metadata on bacteria and gut microbiota communities (metagenetic data) will be archived in the Sequence Read Archive (SRA)
Data and metadata on gene expression (RNA-seq) will be stored in the Gene Expression Omnibus (GEO) data repository.
Data and metadata on metagenomics will be stored at NCBI (National Center for Biotechnology Information)
Modelling code will be stored on GitHub.

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

We expect 35 euro costs per Tb, which will be covered by the C1 project.

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

Biological samples will be made available as restricted data.
Data on pond characteristics and phenotypic traits will be archived as open data in the Dryad Digital Repository when the associated manuscripts are accepted.
Data on bacteria and gut microbiota communities (metagenetic data) will be archived as open data on the Sequence Read Archive (SRA)
Data on gene expression (RNA-seq) will be archived as open data in the Gene Expression Omnibus (GEO) data repository.
Data on metagenomics will be archived as open data at NCBI (National Center for Biotechnology Information)
Modelling code will be publicly stored on GitHub.

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

Access to biological samples and under what conditions will be decided case-by-case based on specific requests.

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.

- Other data repository (specify below)
- KU Leuven RDR (Research Data Repository)

Data on pond characteristics and phenotypic traits will be made available in the Dryad Digital Repository.
Data on composition of bacteria communities (metabarcoding data) will be made available in the Sequence Read Archive (SRA).
Data on gene expression (RNA-seq) will be made available in the Gene Expression Omnibus (GEO) data repository managed by the National Center for Biotechnology Information (NCBI).
Genomic data will be made available at NCBI.
Modelling code will be stored on GitHub.

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

Data on pond characteristics and phenotypes stored in Dryad get a DOI.
16S RNA and RNAseq datasets stored on SRA and Gene Omnibus get an accession number.
Genomic (DNA, RNA and ATAC) data stored at NCBI get an accession number.

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

Dryad and NCBI: free
SRA and Gene omnibus: free
GitHub: free.

Responsibilities

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

PhD students and postdocs linked to the project

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

PhD students and postdocs linked to the project

Who will manage data preservation and sharing?

The responsible PI of a given (sub)work package.

Who will update and implement this DMP?

The responsible PI of a given (sub)work package.