
FWO Odysseus - Characterizing developmental mechanisms of plasticity and their importance in adaptation to rapid environmental change

A Data Management Plan created using DMPonline.be

Creator: Steven Van Belleghem

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Steven Van Belleghem, Robby Stoks

Data Manager: Steven Van Belleghem

Project Administrator: Steven Van Belleghem

Grant number / URL: G0DCM23N

ID: 202402

Start date: 01-10-2023

End date: 30-09-2028

Project abstract:

Understanding the molecular details of how organisms adapt to changing environments remains a prime challenge in biology. Developmental plasticity, which enables organisms to adjust their phenotype in response to environmental cues within a single generation, is particularly important in this process. Although the adaptive significance of such a response is generally recognized, the underlying molecular mechanisms and how they evolve remain highly elusive. Fortunately, novel genome methods enable the study of the mechanisms that orchestrate the development of plastic phenotypes. My aim is to integrate these novel methods into evolutionary studies that test the importance of plasticity in the process of long-term adaptation to changing environments and the resulting genetic adaptation. Specifically, I will test a mechanism by which genetic diversity is masked in populations that plastically express alternative phenotypes and how this plasticity may precede genetic adaptation - a contentious but undemonstrated theory called 'plasticity-led evolution'. I will perform these experiments on the water flea *Daphnia*, an established model system for plastic development. Understanding genes involved in plasticity and how they interact with the process of genetic adaptation has the potential to offer insights into the very essence of life and to provide various applications in biotechnology, medicine, and mitigation of species loss under extreme environmental change.

Last modified: 20-10-2023

FWO Odysseus - Characterizing developmental mechanisms of plasticity and their importance in adaptation to rapid environmental change

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
		<i>Please choose from the following options:</i> <ul style="list-style-type: none">• Generate new data• Reuse existing data	<i>Please choose from the following options:</i> <ul style="list-style-type: none">• Digital• Physical	<i>Please choose from the following options:</i> <ul style="list-style-type: none">• Observational• Experimental• Compiled/aggregated data• Simulation data• Software• Other• NA	<i>Please choose from the following options:</i> <ul style="list-style-type: none">• .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ...• NA	<i>Please choose from the following options:</i> <ul style="list-style-type: none">• <100MB• <1GB• <100GB• <1TB• <5TB• <10TB• <50TB• >50TB• NA	
DNA-seq short	Short-read DNA sequences of Daphnia clones	Generate new data	Digital	Observational	.fastq	<5TB	
DNA-seq long	Long-read DNA sequences of Daphnia clones	Generate new data	Digital	Observational	.fastq	<1TB	
ATAC-seq	Chromatine accessibility profiles of Daphnia clones, developmental changes and exposures	Generate new data	Digital	Experimental	.fastq	<5TB	
RNA-seq	Gene expression profiles of Daphnia clones, developmental changes and exposures	Generate new data	Digital	Experimental	.fastq	<1TB	
Hi-C	Chromatin contacts in Daphnia clones and exposure	Generate new data	Digital	Experimental	.fastq	<1TB	
Single cell RNA and ATAC-seq	Multiomics of Daphnia clone	Generate new data	Digital	Experimental	.fastq	<1TB	
DNA-seq short	Short-read DNA sequences of Daphnia clones	Reuse existing data	Digital	Observational	.fastq	<1TB	
Reference genome	Reference genome of Daphnia magna	Reuse existing data	Digital	Observational	.fastq	<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:

DNA-seq short reads: Published NCBI BioProject under accession code PRJNA344883 and PRJNA624267
Reference genome: Published NCBI BioProject under accession code PRJNA727483 and PRJNA777104

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.

- No

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.

- 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

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).

For genomic sequence data (including DNA, RNA, ATAC, Hi-C and single cell), NCBI metadata standards for data submission will be used. This includes:

- Sample ID (NCBI accession and lab ID)
- Sample origin
- Date of sampling
- Location of sampling
- Tissue
- Life stage
- Sex
- Sequence type and sequence platform

Data processing pipelines stored on GitHub will be accompanied with readme.txt files.

Phenotypic measurements of Daphnia clones will be stored in excel sheets including clone ID, clone origin, phenotypic value, experimenter and date of experiment.

An electronic lab book in which all metadata is shared and curated to common standards will be kept through a lab group on Microsoft teams.

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.

- Yes

1. Raw sequencing reads for: NCBI metadata standards for data submission
2. Reference genome assemblies: NCBI metadata standards for data submission
3. Processed sequenced data: external harddrives and KU Leuven VSC Tier-1 Data Storage
4. Bioinformatic analyses pipeline: readme.txt files with descriptions of the pipelines
5. Phenotypic measurements of Daphnia clones: excel sheets including clone ID and phenotypic value

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

Where will the data be stored?

1. Raw sequencing reads for:
 - External harddrives
 - NCBI (National Center for Biotechnology Information)
 - KU Leuven VSC Tier-1 Data Storage
2. Reference genome assemblies:
 - External harddrives
 - NCBI (National Center for Biotechnology Information)
 - KU Leuven VSC Tier-1 Data Storage
3. Processed sequenced data:
 - External harddrives
 - KU Leuven VSC Tier-1 Data Storage
4. Bioinformatic analyses pipeline:
 - Personal computer
 - External harddrives
 - GitHub
5. Phenotypic measurements of Daphnia clones:
 - Personal computer
 - External harddrives
 - KU Leuven VSC Tier-1 Data Storage

How will the data be backed up?

Personal computer
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 yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

30Tb is supported by the FWO Data component of the Flemish Tier-1 supercomputing platform until 2027.

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?

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.
External harddrives will be stored safely in the office of Steven Van Belleghem.
Personal computers will be protected password protected.

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

Costs for data storage are included in the funded project and allow purchasing external harddrives and KU Leuven VSC Tier-1 Data Storage until the end of the project.

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...).

All sequencing and experimental data will be publicly stored >25 years.

1. Raw sequencing reads for:
 - Short-read DNA sequences (.fastQ)
 - Long-read DNA sequences (.fastQ)
 - ATAC-seq (.fastQ)
 - RNA-seq (.fastQ)
 - Hi-C data (.fastQ)
 - Single cell RNA and ATAC-seq (.fastQ)
2. Reference genome assemblies (.fasta)
3. Processed sequenced data:
 - read alignments to genomes (.bam)
 - gene expression counts (.counts)
 - open chromatin regions (.bw .bed)
 - genome alignments (.maf)
4. Bioinformatic analyses pipeline
5. Phenotypic measurements of Daphnia clones

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

1. Raw sequencing reads for: NCBI (National Center for Biotechnology Information)
2. Reference genome assemblies: NCBI (National Center for Biotechnology Information)
3. Processed sequenced data: external harddrives and KU Leuven VSC Tier-1 Data Storage
4. Bioinformatic analyses pipeline: GitHub
5. Phenotypic measurements of Daphnia clones: Github and publication supplements

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

NCBI: Free

GitHub: Free (up to 2gb, which is sufficient for storing pipelines and scripts)

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.

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, ...)

1. Raw sequencing reads for:
 - Short-read DNA sequences (.fastQ)
 - Long-read DNA sequences (.fastQ)
 - ATAC-seq (.fastQ)
 - RNA-seq (.fastQ)
 - Hi-C data (.fastQ)
 - Single cell RNA and ATAC-seq (.fastQ)
2. Reference genome assemblies (.fasta)
3. Processed sequenced data:
 - read alignments to genomes (.bam)
 - gene expression counts (.counts)
 - open chromatin regions (.bw .bed)
 - genome alignments (.maf)
4. Bioinformatic analyses pipeline
5. Phenotypic measurements of Daphnia clones

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

Processed sequence data will be stored in at least two location using external harddrives and KU Leuven VSC Tier-1 Data Storage. These data can be reproduced from raw reads and the available processing pipelines (shared on GitHub). These processed data types are not typically uploaded to NVBI for storage and sharing, but after publication, these data will also be freely available upon request and shared through web transfers.

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.

- No

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

1. Raw sequencing reads for: NCBI (National Center for Biotechnology Information)
2. Reference genome assemblies: NCBI (National Center for Biotechnology Information)
3. Processed sequenced data: external harddrives and KU Leuven VSC Tier-1 Data Storage
4. Bioinformatic analyses pipeline: GitHub
5. Phenotypic measurements of Daphnia clones: Github and publication supplements

When will the data be made available?

Upon publication.

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

Creative Commons Attribution (CC BY) License

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

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

NCBI: Free

GitHub: Free (up to 2gb, which is sufficient for storing pipelines and scripts)

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.

6. Responsibilities

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

Steven Van Belleghem, Zorimar Vilella-Pacheco, Alexandros Kourtidis

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

Steven Van Belleghem, Zorimar Vilella-Pacheco, Alexandros Kourtidis

Who will manage data preservation and sharing?

Steven Van Belleghem

Who will update and implement this DMP?

Steven Van Belleghem

..