THE ROLE OF smORF ENCODED POLYPEPTIDES IN LAGER BEER FERMENTATION

A Data Management Plan created using DMPonline.be

Creator: Florian Weiland

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Florian Weiland

Grant number / URL: G018923N

ID: 199858

Start date: 03-07-2023

End date: 30-06-2027

Project abstract:

This project uses ribosomal sequencing and proteomics to identify small open reading frames (smORF) encoded polypeptides in S. pastorianus. This will lay the foundation for understanding the regulatory role of SEPs in lager type-beer fermentation under conditions relevant to the brewing industry. Further, the application potential of SEPs as screening markers for beneficial brewing traits in yeasts will be assessed. New data will be generated in the form of ribosomal sequencing data, proteomics data (imaging data from gels, western-blots, LC-MS/MS data) and yeast phenotyping data.

Last modified: 15-06-2023

THE ROLE OF smORF ENCODED POLYPEPTIDES IN LAGER BEER FERMENTATION 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
01_WP1+2+4_Fermentations	Data regarding growth curves, pH values, extract and alcohol concentrations during fermentations	Generate new data	Digital	Observational	.csv	< 100MB	
02_WP1_Fermentations_Analysis_Scripts	Data analysis script files to analyse dataset 01	Generate new data	Digital	Software	.R	< 1MB	
03_WP1_Imaging_Data	Images of SDS-PAGE gels and Western-Blots WP1	Generate new data	Digital	Experimental	.tiff	< 50GB	
04_WP1_Fractionaton_Data	UV-Spectra data for RNA fractionation	Generate new data	Digital	Experimental	.csv	< 100MB	
05_WP1_RNA_Sequencing_Data	TIS-Riboseq data	Generate new data	Digital	Experimental	FASTQ and FASTA	< 100GB	
06_WP2+3_Imaging_Data	Images of SDS-PAGE gels and Western-Blots WP2	Generate new data	Digital	Experimental	.tiff	< 100GB	
07_WP2+3_Fractionation_Data	UV-Spectra of protein and peptide fractionation prior to proteomics	Generate new data	Digital	Experimental	.csv	< 100MB	
08_WP2+3_Mass_spectrometry_data	Quantitative (phospho-)proteomics dataset	Generate new data	Digital	Experimental	.raw and .mzXML	< 1TB	
09_WP2+3_Mass_spectrometry_data_analysis_scripts	Software scripts to analyse 08. Can be based on dataset 20	Generate new data	Digital	Software	.R	< 1MB	
10_WP4_CRISPR-Cas9_vector_data	Vector sequences of the used sgRNAs from gene knock-out	Generate new data	Digital	Experimental	FASTA	< 10MB	
11_WP4_Imaging_Data_Knock-Out_validation	Images of PCR and RT-PCR agarose gels	Generate new data	Digital	Experimental	.tiff	< 50GB	
12_WP4_Sequencing_Data_Knock-Out_Validation	DNA sequencing data of potential knock-out clones	Generate new data	Digital	Experimental	.ab1	< 10GB	
13_WP4_Knock-out_Clones	Yeast knock-out clones (both failed or validated clones) as per dataset 10, 11 and 12	Generate new data	Physical				450 cryovials
14_WP4_PCR_Screening	PCR screening of SEPs as beneficial trait	Generate new data	Digital	Experimental	.ab1	< 10GB	
15_WP4_Beer_Analysis_GC-MS	Chemical analysis of beer fermented in WP4 using SEP knock-out clones	Generate new data	Digital	Experimental	.raw and mzXML	< 100 GB	
16_WP4_Beer_Analysis_LC	Chemical analysis of beer fermented in WP4 using SEP knock-out clones	Generate new data	Digital	Experimental	.raw and .csv	< 10GB	
17_WP4_Beer_Analysis_General	Yeast stress assays	Generate new data	Digital	Experimental	.csv	< 10MB	
18_WP4_Phenotype_screening	Influence of SEP Knock-out on phenotype	Generate new data	Digital	Experimental	.csv	< 100MB	
19_Existing_data_FASTA	Yeast FASTAs	Reused	Digital	Experimental	.FASTA	< 100MB	
20_Existing_data_Software	R scripts	Reused	Digital	Software	.R	< 1MB	

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:

Dataset 19: FASTA files for Saccharomyces pastorianus are downloadable from https://www.ncbi.nlm.nih.gov
Dataset 20: R scripts for data analysis of proteomics datasets are available via https://doi.org/10.5281/zenodo.4311474

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

Dataset 14 has the potential to discover new quality traits for novel beer fermentation yeast strains and species

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).

The principles of FAIR (Findable, Accessible, Interoperable and Reusable) will be applied.
All experimental procedures will be written down in lab books. Each step will wither follow the methodology as described in the applicable standard operating procedure (SOP)while deviations from the SOP will be explicitly marked.
The project will be placed into a main folder, with subfolders for each work package (WP folder). Within each WP folder will be subfolders according to the task structure (Task folder). Any folder below the Task folder level will be named according to following convention: The date followed by a task and an experiment identifier and a short description of the experiment, all divided by underscores: YYYYMMDD_Task1-1_FW001_SDS-PAGE. Words in the short description field are separated by "-"
Each file will be named accordingly: The date followed by a task and an experiment identifier, a running number, a short description of the experiment, all divided by underscores: YYYYMMDD_Task1-1_FW001-01_SDS-PAGE-Gel01.
Files will be put into a structure where each datatype (e.g raw data) will be placed into its own subfolder below the Task folder level (experiment folder):
Folder: YYYYMMDD_Task1-1_FW001_SDS-PAGE
Subfolder
1) Raw_data
Experimental data. In case of machine derived data these will include metadata, e.g. mass spectrometry data derived from proteome analysis.
2) Analysis
Analysis scripts will be placed in the "Analysis folder", while the output will be placed in a "Output_R" subfolder: ./Analysis/Output_R/
Any software scripts will be placed on GitHub for version control.
3) Documentation
Here metadata about the experiments themselves will go in. Examples include pipetting schemes for SDS-PAGE or Protein quantification assays. this includes explanations of non-standard abbreviations.

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

For Proteomics data the MIAPE standard will be used, also in case of publication, all files will be made available in a format as per HUPO-PSI.
The RNA-seq data will use the MINSEQE standard.

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

Where will the data be stored?

Data will be stored at the facilities provided by KU Leuven. Access will be provided via Microsoft Teams.

How will the data be backed up?

Backup data stored in the cloud will be managed by KU Leuven ICTS. In case of non-digital data, e.g. yeast knock-out clones, these will be stored in temperature monitored freezers connected to the emergency power supply. In case the temperature is too high, the system contacts persons responsible for the freezer (e.g. Florian Weiland).

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

The total data volume is expected to be around < 5TB. This volume is supported at KU Leuven via Microsoft Teams (max 5TB).

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

Microsoft Teams supports multifactor authentication as measure against unauthorized access and modification.
For non-digital data, e.g. storage of yeast knock-out clones in freezers: Access to the location of these freezers is restricted via a card system.

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

Micorsoft Teams is free for KU Leuven staff.

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...).

Data will be preserved according to KU Leuven guidelines for 10 years, which exceeds the current FWO guidelines of 5 years.

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

Data preserved in dedicated data repositories is listed in Question 5-1.
All other data will be stored at KU Leuven Large Volume Storage.

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

PRIDE, MetaboLights, EMBL Expression Atlas and Zenodo are free of charge.
KU Leuven Large Volume Storage cost approx. 100 Euro/year per TB

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

All datasets which are not part of publication supplementary datasets and/or deposited at publicly available storage services will be made available on request.

Datasets will be made available at:
01: Part of Supplementary data of publications and Zenodo
02: Zenodo
03: Part of Supplementary data of publications and Zenodo
04: Part of Supplementary data of publications and Zenodo
05: EMBL Expression Atlas
06: Part of Supplementary data of publications and Zenodo
07: Part of Supplementary data of publications and Zenodo
08: PRIDE
09: Zenodo
10: Part of Supplementary data of publications and Zenodo
11: Part of Supplementary data of publications and Zenodo
12: Part of Supplementary data of publications and Zenodo
13: Data will be made available on request
14: Part of publication and Zenodo
15: MetaboLights
16: MetaboLights
17: Part of publication and Zenodo
18: Zenodo

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

We aim to pre-publish all manuscripts and access will be restricted until publication on BioRxiv.

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.

Datasets:
02: Zenodo
05: EMBL Expression Atlas
08: PRIDE
09: Zenodo
15: MetaboLights
16: MetaboLights
18: Zenodo
All other datasets will be uploaded to Zenodo (if not part of supplementary data in publication)

When will the data be made available?

Upon publication of results

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

Software: GNU Library or "Lesser" General Public License 3.0 (LGPL-3.0)
Data: Public Domain Mark (PD) or (if applicable) CC0

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?

No costs expected.

6. Responsibilities

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

The postdoctoral researcher

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

KU Leuven

Who will manage data preservation and sharing?

KU Leuven

Who will update and implement this DMP?

Florian Weiland (Supervisor)

THE ROLE OF smORF ENCODED POLYPEPTIDES IN LAGER BEER FERMENTATION 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)

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)

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)

Question not answered.

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)

Question not answered.

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

Question not answered.

THE ROLE OF smORF ENCODED POLYPEPTIDES IN LAGER BEER FERMENTATION DPIA

DPIA

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

- Not applicable

THE ROLE OF smORF ENCODED POLYPEPTIDES IN LAGER BEER FERMENTATION GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable