

Plan Overview

A Data Management Plan created using DMPTool

Title: Evolution of the Gene Regulatory Networks for Oxidative Stress Response in Yeasts

Creator: He, Bin - ORCID: [0000-0002-3072-6238](https://orcid.org/0000-0002-3072-6238)

Affiliation: University of Iowa (uiowa.edu)

Data Manager: Raluca Gordân

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: nsf22586

Grant: <https://www.nsf.gov/pubs/2022/nsf22586/nsf22586.htm>

Template: NSF-BIO: Biological Sciences

Project abstract:

Dissect the evolution of oxidative stress response networks between related yeast species that exhibit strongly different resistance to reactive oxygen species such as hydrogen peroxide. The species being compared include a commensal and opportunistic human pathogen and a related low pathogenic potential yeast, namely the baker's yeast *S. cerevisiae*. The much stronger oxidative stress resistance in the former suggests potential relevance of the stress response evolution in host adaptation.

Start date: 03-01-2023

End date: 02-28-2028

Last modified: 07-15-2022

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Evolution of the Gene Regulatory Networks for Oxidative Stress Response in Yeasts

Data and Materials Produced

Describe the types of data, physical samples or collections, software, curriculum materials, and other materials to be produced in the course of the project. (For collaborative proposals, the DMP must cover all the various data types being collected by each collaborator.)

Data: functional genomics such as RNA-seq, ChIP-seq; quantitative PCR; oxidative stress enzymatic assays; flow cytometry; growth curves; Protein Binding Microarray measurements for TF binding affinities; protein crystallography; fluorescence microscopy.

Scripts and analyses: Python, Shell and R scripts, research notebooks, markdown documentations.

Samples: bacterial and yeast strains as glycerol stocks

Curriculum materials: Scientific Computing Workshop materials, including syllabus, workshop materials; Syllabus and workshop materials for an entry-level Bioinformatics for Beginners course.

Standards, Formats and Metadata

Describe the standards to be used for all the data types anticipated, including data or file format and metadata. [Note: Where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies.]

Type of data	Form of data
Genomics, e.g. RNA-seq, ChIP-seq	<u>fastq</u> files and accompanying metadata conforming to GEO standard
Flow cytometry	<u>FCS</u> files, with meta data included
Microscopy	<u>Leica endogenous formats</u> , meta data included
qPCR and enzymatic assays	<u>tabular text files</u> with header information as comments or separate
binding affinity measurements	<u>original files</u> by the Biolayer Interferometry instrument, <u>tabular text</u> files for data processing with meta information.
Protein crystallography	<u>original image</u> files from the instrument, standard <u>PDB</u> files
Python scripts	.py and .ipynb (jupyter notebook)
R scripts	.Rmarkdown (Rstudio)
Shell script	standard bash script

Roles and Responsibilities

Describe the roles and responsibilities of all parties with respect to the management of the data (including contingency plans for the departure of key personnel from the project).

The PI, Bin He, is responsible for the deposit, maintenance and management of the data. Lab members including graduate students, research technicians and undergraduate researchers will work with the PI to properly document, deposit and maintain the data they generate consistent with the NSF guideline on DMP. Collaborators, including the protein and crystallography core and Dr. Raluca Gordân's lab, will assist the PI in depositing and managing data.

Dissemination Methods

Describe the dissemination methods that will be used to make data and metadata available to others during the period of the award, and any modifications or additional technical information regarding data access after the grant ends.

Genomics data will be uploaded to the NCBI GEO database according to their guidelines. Flow cytometry data will be deposited on FlowRepository.org. Microscope imaging data will be uploaded onto [Image Data Resource](#). All custom scripts and analysis documentation, including Rmarkdown and plain markdown documents, will be hosted on the project-specific GitHub repository, effectively producing a "computational research notebook" that is version controlled throughout the project period ([example](#) of PI's lab's most recent computational research notebook repo). Additional data suitable for display in the browser, e.g. images, tabular text files, will be uploaded onto FigShare. Educational modules, including the syllabus, training materials such as Rmarkdown files, will be made available through a GitHub repository. All of the above will continue to be available through the same means for the foreseeable future even after the grant ends.

Policies for Data Sharing and Public Access

Describe the PI's policies for data sharing, public access and re-use, including re-distribution by others and the production of derivatives. Where appropriate, include provisions for protection of privacy, confidentiality, security, intellectual property rights and other rights.

All data and scripts generated in this project will be made publicly available and can be re-used and re-distributed for academic or commercial purposes with proper credit attribution (under MIT license).

Archiving, Storage and Preservation

Where relevant, describe plans for archiving data, samples, software, and other research products, and for on-going access to these products through their lifecycle of usefulness to research and education. Consider which data (or research products) will be deposited for long-term access and where. (What physical and/or cyber resources and facilities (including third party resources) will be used to store and preserve the data after the grant ends?)

Data: Raw data from instruments, such as the sequencer, microscope, flow cytometer and X-ray crystallography will be archived on the Research Data Storage Service and Large Storage Service provided by the PI institution's Information Technology Service. Each offers 5 TB space and are

backed up daily.

Sample: All bacterial and yeast strains will be archived as glycerol stocks and will be stored in our UltraLow Temperature freezers. Both have a backup copy in a separate ULT in a different building. All strains are available to any researcher upon request. Plasmids will be made available upon request, either via the lab or through an organization such as Addgene.

Script: As mentioned before, all scripts will be made available through a GitHub repository.

Educational modules: The workshop and tutorial materials for both the new course and the workshop derived from it are made available through a GitHub repository (e.g., <https://github.com/hezhaobin/2020-Data-Skills>). The materials are freely available to the community under the MIT license.
