# Plan Overview

*A Data Management Plan created using DMPTool-Stage*

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

**Creator:** He, Bin - **ORCID:** [0000-0002-3072-6238](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](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:** 02-28-2023

**End date:** 02-27-2028

**Last modified:** 07-15-2022

**Copyright information:**

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

**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.

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

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.

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.

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).

**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.