Comparative Genomics of Arsenic Detoxification in *S. cerevisiae* and *C. reinhardtii*

Garret Genco, Biology Department, Colby College, Waterville, ME 04901

Crysten Blaby, Biology Department, Brookhaven National Laboratory, Upton, NY 11973

Abstract

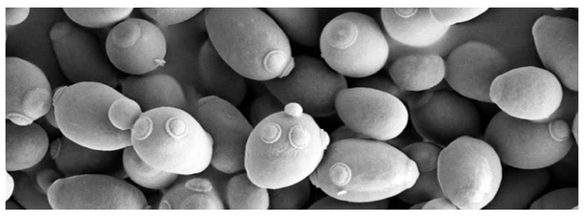
Arsenic is a pervasive, heavily toxic, naturally occurring metalloid that poses significant health risks through contamination of water sources, air-born particulate matter and soil-derived matter. In response to the ubiquitous presence of arsenic, plants, fungi and algae are well documented to have developed a myriad of arsenic detoxification mechanisms. To determine these gene responses, this study focused on two microbial model organisms – *Saccharomyces cerevisiae* and *Chlamydomonas reinhardtii*, a fungus and green alga respectively, which are the most well understood eukaryotic organism in the stress response field. 28 S. cerevisiae proteins were identified by the comprehensive literature review and comparative genomics analyses, of which 10 had SSNs generated and only seven of these SSNs contained isofunctional cluster links between the algae and *S. cerevisiae* proteins. The *S. cerevisiae* proteins Pho87, Pho88, Acr3 and Fps1 were generally involved in regulating the influx and efflux of arsenic. Those mediating the arsenate fluxes were typically inhibited to prevent further uptake, while those mediating arsenite fluxes were upregulated to increase the efflux of arsenite. To cope with the effects of intracellular arsenic concentrations, Xrn1 activates the Fet3 protein which significantly decreases mRNA levels in the presence of arsenic, repressing high affinity Fe uptake and eventual Fe deficiency, but increasing arsenic resistance. While Aco1 and Leu1 are overexpressed in the presence of arsenate, alleviating growth defects of arsenate that increase Fe toxicity. The role of alternative algae, plant and fungi proteins could not be described, since some proteins in the phylogenetic trees may have been possibly clustered together into a single node, thus hindering the ability to identify potential paralogs and orthologs. Future research should rectify this mistake made when constructing the phylogenetic trees, then gene knockout experiments should be done to investigate the functions of those proteins.

Introduction

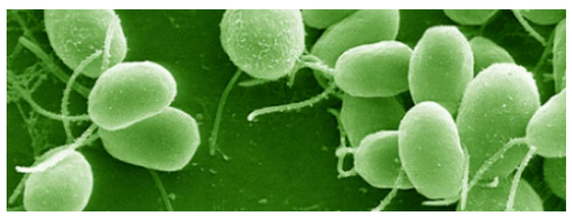
In this climate change defined era, the production of sustainable bioproduction crops that maintain optimal performance in diverse and fluctuating environments are needed. To improve bio-based crops and understand how plants respond to biotic and abiotic stresses, BNL's QPSI department aims to address the gap between plant genomes and their function through bioinformatic projects. The genome contains all the genes needed to build cells, tissues, and reproduce. Over the past two decades, scores of whole-genome sequences have been published, however the function of most genes is still unknown limiting synthetic biology approaches.

Many genes have relatives across the tree of life and this can be used to infer a protein's function. Assuming there is a gene relationship between two organisms, if the gene function from one organism is known then it can be used to understand the function of genes in the other related organism. Two model organisms, the yeast *Saccharomyces cerevisiae* and green alga *Chlamydomonas reinhardtii*, for such an approach are generally regarded as the most well understood eukaryotic organism in the stress response field (4,6).

A pertinent stress response field to the QPSI is arsenic, which is a ubiquitous toxic pollutant that universally exists in the environment and contaminates water, air and soil derived sources (3). When ingested by organisms, arsenic can inhibit gene function (i.e. photosynthesis) and induce apoptosis (5,6). Thus, it is imperative to understand the arsenic detoxification mechanisms of various organisms. This project aims to leverage phylogenetic relationships between the two microbial organisms and algae, fungi, and plant proteins, to contextualize functional inferences and predict protein function in response to arsenic stress.



1



2

Research Goals and Methods

1. Understand biological responses of *S. cerevisiae* to arsenic
   - A. Literature review looking at:
     - Sources and Forms of Arsenic Encountered in the Environment
     - Gene Expression Changes in Response to Arsenic
     - Different Arsenic Detoxification Pathways – Molecular and Physiological
   - B. Comparative Genome Analysis of *S. cerevisiae* arsenic detoxification proteins from literature review
     - Includes Protein Name, Uniprot ID, Function and Evidence (methods used to ascertain function), Protein Family, and Pfam/Interpro ID
2. Bioinformatic study of which genomes encode homologs of the *S. cerevisiae* proteins
   - A. Constructed Sequence Similarity Networks (SSNs) in Cytoscape for each *S. cerevisiae* protein
     - To identify isofunctional clusters
   - B. Phylogenetic trees built in iTOL for *S. cerevisiae* proteins and algae proteins within or linked to *S. cerevisiae* SSN clusters
     - Identified Swissprot Reviewed Proteins, and all *S. cerevisiae* plus *C. reinhardtii* proteins, and relevant MRCA nodes
     - Wanted to elucidate paralogs and orthologs
   - C. Phylum were classified in both SSNs and Phylogenetic Trees, while Kingdom were classified only in Phylogenetic Trees

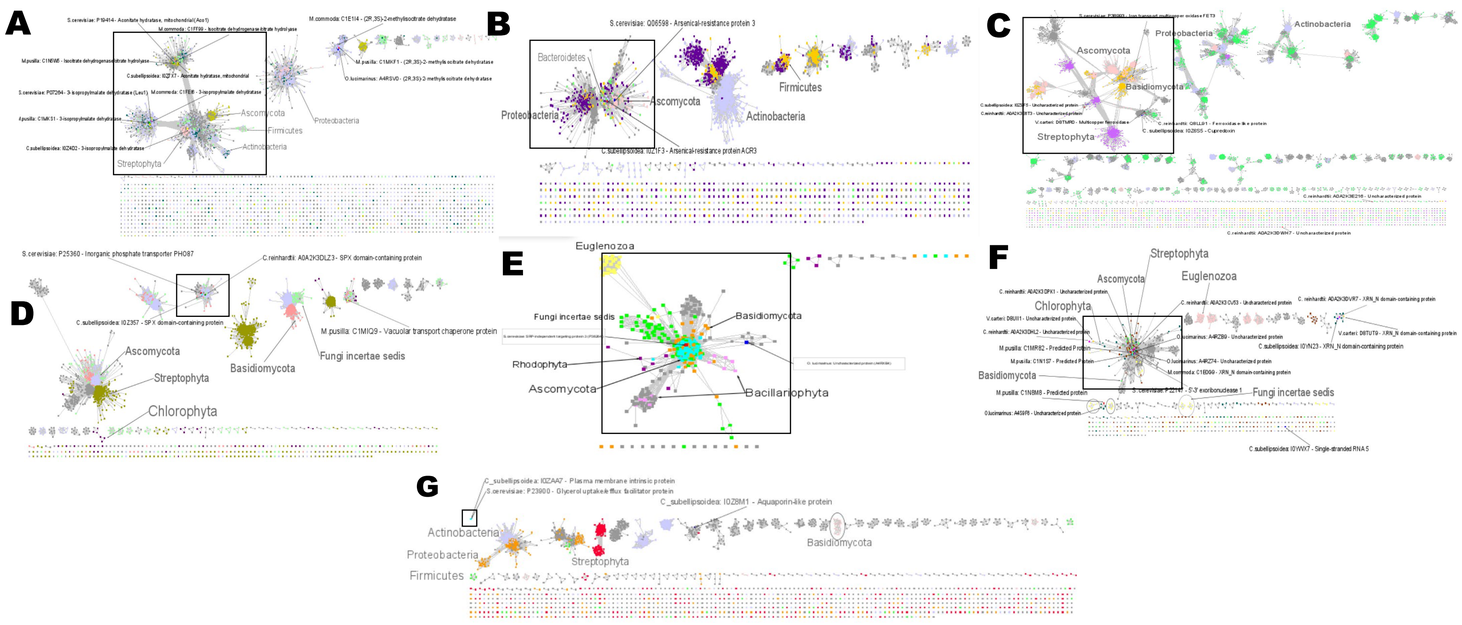


Figure 1: These SSNs contained arsenic detoxification algae proteins that were either linked or contained within the same isofunctional clusters as the *S. cerevisiae* proteins. Black boxes identify the location of these isofunctional cluster(s). The following SSNs are of the *S. cerevisiae* (A) Aco1 and Leu1, (B) Acr3, (C) Fet3, (D) Pho87, (E) Pho88, (F) Xrn1, (G) Fps1 proteins

Protein Functions

- **Aco1 and Leu1** = Overexpressed in presence of arsenate. Alleviates growth defects of arsenate that increase Fe toxicity
- **Acr3** = Plasma membrane protein induced by the presence of arsenic. Increases efflux of arsenite, reducing arsenic toxicity
- **Pho87 and Pho88** = Phosphate transporters that uptake arsenate. Inhibited in the presence of increased arsenate concentrations to increase tolerance
- **Fps1** = Bidirectional channel which mediates transport of substrates down the concentration gradient; major uptake and efflux of arsenite. Inhibited by arsenic exposure
- **Xrn1** = Is up-regulated by arsenate exposure and activates the Fet3 protein
- **Fet3** = mRNA levels are significantly decreased to arsenic exposure, repressing high affinity Fe uptake and eventual Fe deficiency but increasing arsenic resistance

Conclusions

- Of the 28 proteins identified from the literature review, only 10 SSNs could be generated
  - WHY: Job fails, Excessive file sizes, and lack of algae proteins
- From the 10 SSNs, seven SSNs had isofunctional cluster links between algae and the *S. cerevisiae* proteins
  - Aco1 and Leu1, Acr3, Fet3, Pho87, Pho88, Xrn1, Fps1
- Identification of paralogs and orthologs could not be concluded since UniRef groups were used
  - If UniRef90 or 50 were used, instead of UniRef100, then a single branch may represent multiple proteins instead of a single protein.
- Gene knockout experiments should be done to investigate the proteins preliminarily identified in the SSNs and phylogenetic trees

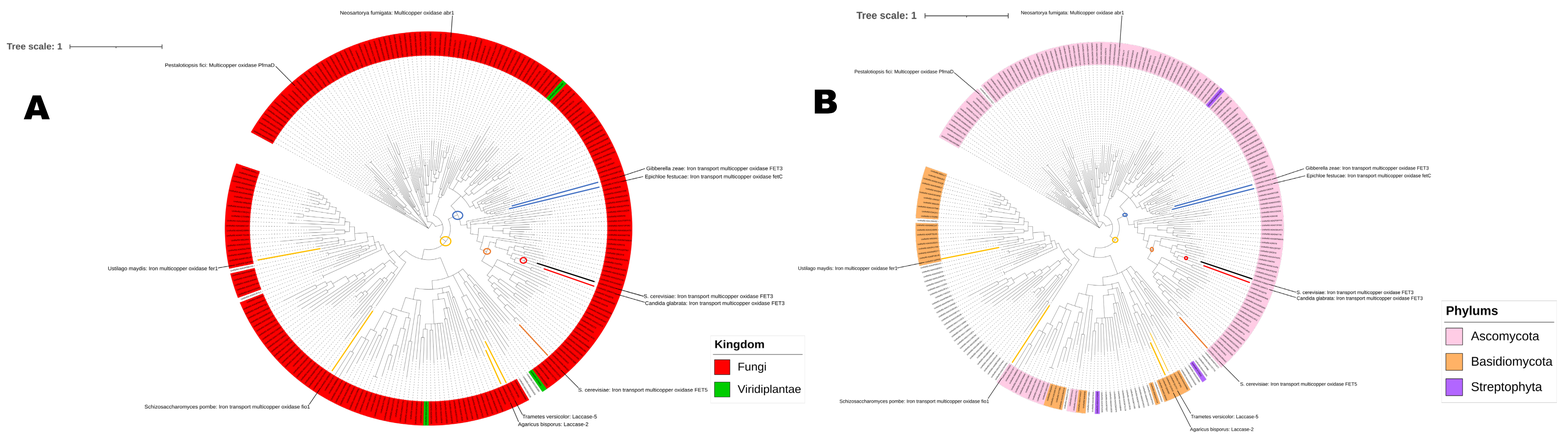


Figure 2: Along with the S. *cerevisiae*, each of the algae within isofunctional clusters identified in Figure 1 had phylogenetic trees constructed in iTOL. All *C. reinhardtii* and S. cerevisiae proteins, the protein used to generate the phylogenetic tree, all Swissprot Reviewed proteins, the most recent common ancestor (between the protein used to generate the trees and all the other identified proteins), and the top five most common kingdom or phylum were all identified in the phylogenetic trees. The solid black, thickened line represented the protein used to generate the phylogenetic tree. These are phylogenetic trees of the *S. cerevisiae* Fet3 protein with the (A) Kingdom and (B) Phylum identified.

Literature Cited

1. https://en.wikipedia.org/wiki/Saccharomyces_cerevisiae
2. http://astrobiology.com/2020/02/genome-sequence-of-chlamydomonas-reinhardtii-submitted-to-nasa-genelab-data-repository.html
3. Jin-Yong Chung, Seung-Do Yu and Young-Seoub Hong (2014). Environmental Source of Arsenic Exposure. Journal of Preventive Medicine and Public Health 47: 253-257.
4. Toshikazu Kaise, Shoko Fujiwara, Mikio Tsuzuki, Teruaki Sakurai, Tohru Saitoh and Chiyo Mastubara (1999). Accumulation of Arsenic in a Unicellular Alga *Chlamydomonas reinhardtii*. Applied Organometallic chemistry 13: 107-111.
5. Isao Kobayashi, Shoko Fujiwara, Kosuke Shimogawara, Toshikazu Kaise, Hideaki Usuda and Mikio Tsuzuki (2003). Insertional Mutagenesis in a Homologue of a Pi Transporter Gene Confers Arsenate Resistance on *Chlamydomonas*. Plant Cell Physiology 44(6): 597-606.
6. Claudina Rodrigues-Pousada, Frédéric Devaux, Soraia M. Caetano, Catarina Pimentel, Sofia da Silva, Ana Carolina Cordeiro and Catarina Amaral (2019). Yeast AP-1 like transcription factors (Yap) and stress response: a current overview. Microbial Cell 6(6): 267-285.

Acknowledgements

I wish to thank my mentor, Dr. Crysten Blaby, especially during this pandemic, for her professionalism and guidance during the SURP program. I would also like to thank the laboratory team for their suggestions and critique during lab meetings. Further, I would like to express my gratitude to the Office of Science Education, and all who helped plan and prepare this virtual internship. I appreciate the efforts of the National Science Foundation, regarding their financial support. Finally, I wish to acknowledge the virtual accommodation and kindness of Mr. Mel Morris, Brookhaven National Laboratory and the Biology Department.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

Genco, Garret