Discovering Transcription Regulation Networks in Bioenergy-relevant Yeast Species

Veronika Dubinkina[1,2*] (vd6@illinois.edu), Ping-Hung Hsieh,[1,3] **Yasuo Yoshikuni,**[1,3] **Sergei Maslov,**[1,2] and **Huimin Zhao**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), Urbana, IL; [2]Department of Bioengineering and Carl Woese Institute of Genomic Biology (IGB), University of Illinois Urbana-Champaign, Urbana; [3]Lawrence Berkeley National Lab, Berkeley, CA

https://cabbi.bio

Project Goals: The goal of this project is to reconstruct Gene Regulatory Networks (GRN) for a set of bioenergy-relevant novel yeasts leveraging existing data on model yeast species. The resulting refined GRNs will be used to develop strategies for improving yield of pathways producing target bioenergy compounds.

Understanding gene regulatory networks (GRN) is critical for metabolic engineering of yeast species. It is known that transcription regulation undergoes rapid evolutionary rewiring leading to substantial divergence of regulatory networks across species. Thus, for poorly understood, non-model organisms, such as *Issatchenkia orientalis, Yarrowia lipolytica,* and *Rhodosporidium toruloides*, GRNs remain mostly unknown.

To facilitate the reconstruction of full GRNs for yeasts of interest we first identified all putative transcription factors (TFs) in their genomes. To do that we compiled a database of TF-specific protein domains from two different databases (DBD: Transcription factor prediction database [1] and Fungal Transcription Factor Database [2]) and performed a protein family domain search to identify all putative TF genes in new strains. We estimated that there are 306 putative TF in the *I. orientalis* genome, 336 in *Y. lipolytica*, and 398 in *R. toruloides*.

Since many transcription factor binding motifs (TFBS) and DNA-binding domains are known to be conserved, we used them to predict GRNs of novel species based on the existing data for the model yeast species *Saccharomyces cerevisiae*. We built a two-step bioinformatics pipeline for genome sequence analysis that transfers information about experimentally validated binding sites of TFs in *S. cerevisiae* to the species of interest. We used OrthoFinder along with the recently published collection of 332 yeast genomes [3] to identify putative TFs which are orthologous to some previously studied in *S. cerevisiae*. We then retrieved a set of motifs for these TFs in *S. cerevisiae* from the YEASTRACT+ database [4]. A MEME motif scan was used to identify putative targets of these TFs based on motif presence in the gene promoters of species of interest (see Table 1 for the summary of reconstructed GRNs). Notably the number of orthologous target genes for these TFs which were conserved between species is relatively small. We also analyzed existing RNA-seq data for *I. orientalis* in 12 different media conditions and partially confirmed our predicted GRNs.

We plan to collect more RNA-seq data which will help us to validate and curate reconstructed GRNs. Although we do not have motif binding information for most of the newly predicted TFs, it can be partially inferred by de novo motif discovery using clusters of co-expressed genes in

RNA-seq experiments and can be further confirmed by DAP-seq experiments. Overall, resulting GRNs will be incorporated into the integrated model of metabolic networks of novel yeasts and empower CABBI teams to predict genomic modifications that can improve production of target biofuel compounds.

Table 1. General statistics of reconstructed GRNs.

Strain	Total # of genes	# of putative TFs	# of TFs with known motifs	# targets	# links
I. orientalis SD108	4925	309	64	4361	20678
Y. lipolytica W29	7919	336	71	7320	62102
R. toruloides IFO0880	8221	398	50	7099	31295
S. cerevisiae	6725	307	118	6725	201974

References

1. Kummerfeld, Sarah K., and Teichmann, Sarah A. "DBD: a transcription factor prediction database." *Nucleic acids research* 34.suppl_1 (2006): D74-D81.
2. Park, Jongsun, et al. "FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors." *Bioinformatics* 24.7 (2008): 1024-1025.
3. Shen, Xing-Xing, et al. "Tempo and mode of genome evolution in the budding yeast subphylum." *Cell* 175.6 (2018): 1533-1545.
4. Monteiro, Pedro T., et al. "YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts." *Nucleic Acids Research* 48.D1 (2020): D642-D649.

***Funding statement**: This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*