Database README

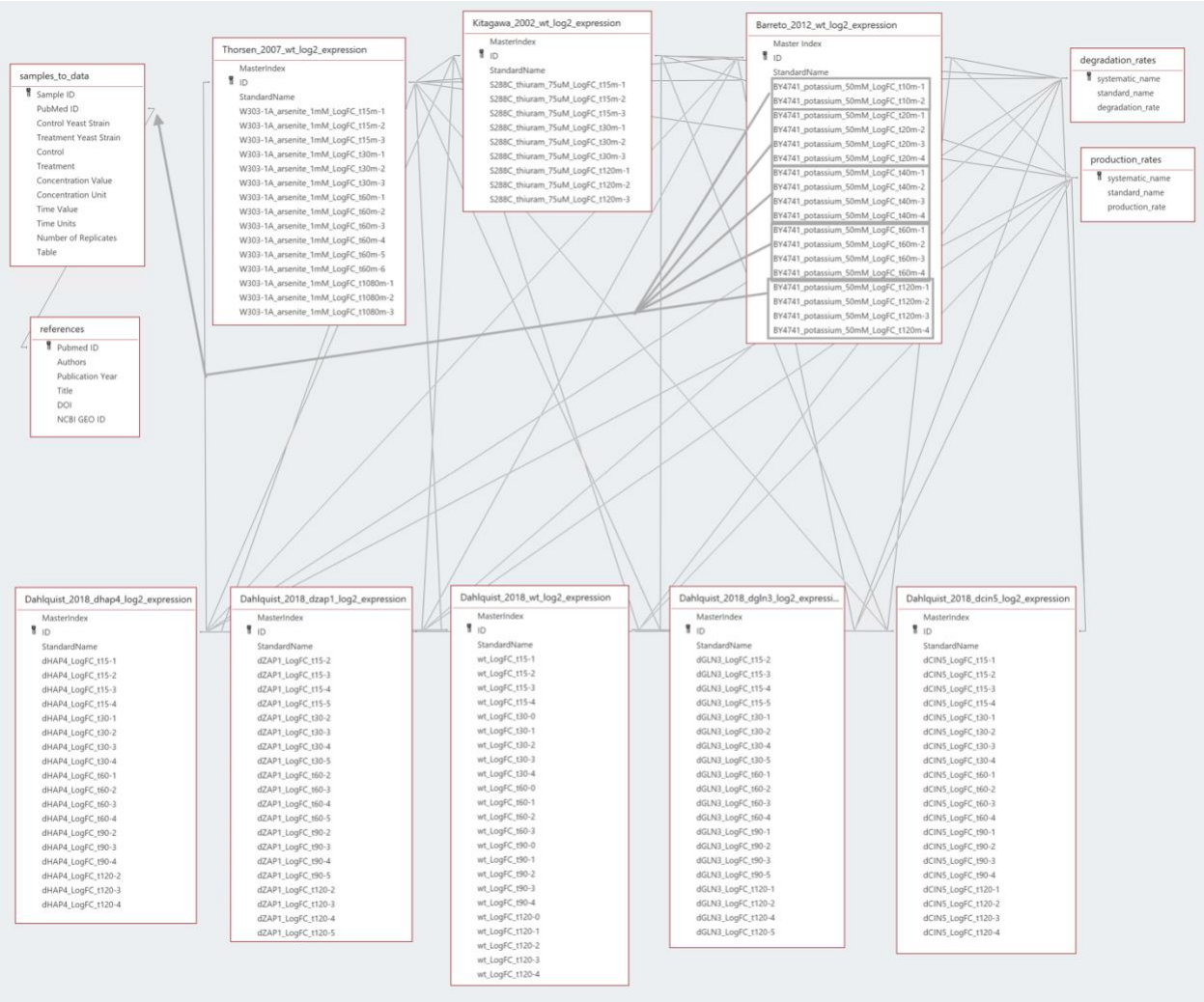

Fig. 1: Diagram of database schema is shown above. The title of each table is shown at the top of each box, with the title of each field listed underneath. Lines represent relationships between tables and fields.

Database Structure

The database consists of eight tables containing gene expression data, two tables with degradation and production rates, and two reference tables containing metadata about the sources of the data. There are three tables with data that the Biological Databases class created. These tables are called “Thorsen_2007_wt_log2_expression”, “Kitagawa_2002_wt_log2_expression”, “Barreto_2012_wt_log2_expression”. These tables (and the “Dahlquist_2018_..._log2_expresson” sheets) consist of a Master Index, ID (systematic name), standard name, and gene expression data sheets. The names of the gene expression data sheets follow a standard nomenclature ([Yeast Strain]_LogFC_t[Time Point]-[Replicate Number]). The values in brackets (and the brackets) are replaced with the actual field names. For example, a field labeled “wt_LogFC_t15-3” would indicate that the field contains data log-fold expression data from a wild type yeast strain and is the third replicate of the 15th timepoint. The

“degradation_rates” and “production_rates” tables contain the “systematic_name”, “standard_name”, and either “degradation_rate” or “production_rate” fields, depending on the table. The “samples_to_data” table contains information regarding each table. The fields consist of “Sample ID” (the PubMed ID + index), “PubMed ID” (the PubMed ID of the article from which the data was derived), “Control Yeast Strain”, “Treatment Yeast Strain”, “Control” (the control treatment), “Treatment”, “Concentration Value”, “Concentration Unit”, “Time Value”, “Time Unit”, “Number of Replicates”, and “Table” (the table name in which the data is located). Each entry in this table is linked to the fields which correspond to the appropriate data set and timepoints via the entries in the “Table” field. The “references” table is linked to this data via the “PubMed ID” fields contained in both tables. This table contains information regarding the articles from which the data was derived. The fields contained in this table are “PubMed ID”, “Authors”, “Publication Year”, “Title”, “DOI”, and “NCBI GEO ID” (the accession number of the dataset in the NCBI Gene Expression Omnibus database).

Database Data Sources

Data in each table in the database came from the following sources:

Thorsen_2007_wt_log2_expression:

Thorsen, M., Lagniel, G., Kristiansson, E., Junot, C., Nerman, O., Labarre, J., & Tamás, M. J. (2007). Quantitative transcriptome, proteome, and sulfur metabolite profiling of the Saccharomyces cerevisiae response to arsenite. *Physiological genomics*, 30(1), 35-43. DOI: 10.1152/physiolgenomics.00236.2006

Kitagawa_2002_wt_log2_expression:

Kitagawa, E., Takahashi, J., Momose, Y., & Iwahashi, H. (2002). Effects of the pesticide thiuram: genome-wide screening of indicator genes by yeast DNA microarray. *Environmental science & technology*, 36(18), 3908-3915. DOI: 10.1021/es015705v

Barreto_2012_wt_log2_expression:

Barreto, L., Canadell, D., Valverde-Saubí, D., Casamayor, A., & Ariño, J. (2012). The short-term response of yeast to potassium starvation. *Environmental microbiology*, 14(11), 3026-3042. DOI: 10.1111/j.1462-2920.2012.02887.x

Dahlquist_2018_..._log2_expresson:

unpublished Dahlquist lab data