

ARTICLE

Oleaginous Yeast Biology Elucidated With Comparative Transcriptomics

Sarah J. Weintraub¹  | Zekun Li² | Carter L. Nakagawa¹ | Joseph H. Collins² | Eric M. Young^{1,2,3}

¹Department of Bioinformatics and Computational Biology, Worcester Polytechnic Institute, Worcester, Massachusetts, USA | ²Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts, USA | ³Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, Massachusetts, USA

Correspondence: Eric M. Young (ericyoung7@gmail.com)

Received: 3 April 2024 | **Revised:** 30 September 2024 | **Accepted:** 12 November 2024

Funding: This study was supported by a National Science Foundation CAREER Award (1944046 to E.Y.), and by a National Science Foundation National Research Training (NRT) grant, CEDAR (NRT-HDR-2021871).

Keywords: metabolic network | oleaginous | tolerance | transcriptomics | unsupervised machine learning | yeast

ABSTRACT

Extremophilic yeasts have favorable metabolic and tolerance traits for biomanufacturing- like lipid biosynthesis, flavinogenesis, and halotolerance – yet the connection between these favorable phenotypes and strain genotype is not well understood. To this end, this study compares the phenotypes and gene expression patterns of biotechnologically relevant yeasts *Yarrowia lipolytica*, *Debaryomyces hansenii*, and *Debaryomyces subglobosus* grown under nitrogen starvation, iron starvation, and salt stress. To analyze the large data set across species and conditions, two approaches were used: a “network-first” approach where a generalized metabolic network serves as a scaffold for mapping genes and a “cluster-first” approach where unsupervised machine learning co-expression analysis clusters genes. Both approaches provide insight into strain behavior. The network-first approach corroborates that *Yarrowia* upregulates lipid biosynthesis during nitrogen starvation and provides new evidence that riboflavin overproduction in *Debaryomyces* yeasts is overflow metabolism that is routed to flavin cofactor production under salt stress. The cluster-first approach does not rely on annotation; therefore, the coexpression analysis can identify known and novel genes involved in stress responses, mainly transcription factors and transporters. Therefore, this work links the genotype to the phenotype of biotechnologically relevant yeasts and demonstrates the utility of complementary computational approaches to gain insight from transcriptomics data across species and conditions.

1 | Introduction

Stress response is one of the least understood, but most important properties, of yeasts (Ribeiro, Bourbon-Melo, and Sá-Correia 2022; Shima et al. 2010; Attfield 1997). In biotechnological applications, an organism that survives and thrives under stresses has process design benefits such as tolerating inhibitory compounds in feedstocks, reducing heat removal requirements, mitigating contamination risk, and stimulating biosynthesis (Mukherjee et al. 2017; Deparis et al. 2017; Wallace 2014; Zhao and Bai 2009). Oleaginous,

flavinogenic, non-Saccharomyces yeasts not only have potentially beneficial metabolic capability to consume inexpensive biomass and produce commercially relevant classes of organic compounds, they have unique tolerance profiles (Adrio 2017; Chen et al. 2009; Shi and Zhao 2017; Sitepu et al. 2014; Santamauro et al. 2014). Here, we present a comparative study of potential yeast cell factories under stress, including nitrogen starvation, iron starvation, and salt stress. We focus on the oleaginous yeast *Yarrowia lipolytica* and the oleaginous, flavinogenic, salt-tolerant yeasts *D. hansenii* and *D. subglobosus*. Since these yeasts are from different lineages yet have

phenotypic similarity, comparing their transcriptomes could reveal underlying rules controlling stress response.

Oleaginous yeasts are of biotechnological interest because the fatty acids they produce can be used to produce oleochemicals (Spagnuolo, Yaguchi, and Blenner 2019; Yaguchi, Rives, and Blenner 2017; Lv et al. 2019). The most well-known is the yeast *Y. lipolytica*, which is a potential industrial host due to its halotolerance, robust secretion ability, and broad pH tolerance (Park and Ledesma-Amaro 2023; Wang et al. 2020; Madzak 2021; Bao et al. 2021; Miller and Alper 2019; Sun et al. 2022; Abdel-Mawgoud et al. 2018; Madzak 2018; Bankar, Kumar, and Zinjarde 2009). The *Y. lipolytica* PO1f strain is an engineered version of *Y. lipolytica* W29 that is widely used as a host for genetic engineering (Madzak et al. 2000; Jagtap et al. 2021; Liu and Alper 2014; Zhu et al. 2024; Cao et al. 2016). It is not naturally flavinogenic, but recent work has shown that riboflavin overproduction can be engineered (Wagner et al. 2018; Palmer et al. 2020; Lv et al. 2019; Sáez-Sáez et al. 2020). Lesser known, but gaining interest, are yeasts in the genus *Debaryomyces* (O'Brien et al. 2018; Defosse et al. 2018). Genomics has shown that *Debaryomyces* belongs to the CTG clade of alternative codon using yeasts, along with *Scheffersomyces stipitis* and *Candida albicans*, suggesting biotechnological and biomedical relevance. *D. hansenii* (formerly *C. famata*) is found in the production of surface-ripened cheese and in wild fermentations (Capece and Romano 2009; Sohier et al. 2009). Like *Y. lipolytica*, it is oleaginous and halotolerant. Yet, it is naturally flavinogenic (Dmytruk, Voronovsky, and Sibirny 2006; Voronovsky et al. 2004; Sibirny and Voronovsky 2009; Tsyrlunyk et al. 2020). *Debaryomyces subglobosus* (formerly *Candida famata* and *Candida flareri*) is also oleaginous, halotolerant, and flavinogenic and is most commonly isolated from processed food products (Dmytruk and Sibirny 2012). In particular, the dep8 strain of *D. subglobosus*, a mutagenized strain, has been shown to produce more than 20 g riboflavin/L in 200 h (Heefner et al. 1992). However, after several years of use as an industrial riboflavin producer, the dep8 strain is no longer used due to genetic instability (Abbas and Sibirny 2011). In contrast to *Y. lipolytica*, *Debaryomyces* yeasts have more robust growth on pentoses, common constituents in lignocellulosic biomass, and are extremely halotolerant (Navarrete et al. 2021; Hobot and Jennings 1981; Papouskova and Sychrova 2007). Even though *Debaryomyces* and *Yarrowia* share some similar phenotypes, they are both from different yeast lineages. *Y. lipolytica* is Dipodascaceae/Trichomonascaceae major clade, while *Debaryomyces* yeasts are in the CTG-Ser1 clade (Ochoa-Gutiérrez et al. 2022; Labella et al. 2019). *Y. lipolytica* and *Debaryomyces* last common ancestor is Saccharomycotina (Morel et al. 2015). This means that a comparison could highlight conserved and evolved genetic signatures of stress response.

One major issue with studying non-model organisms is their often incomplete or poorly annotated genomes. Whole genome assembly of resequenced genomes can improve genome quality (Collins et al. 2021), but annotation tools rely on inference to model organisms which can still be poor and miss unique biology. This limits the application of the bioinformatic tools that can analyze conditional gene expression data and connect genotype to phenotype. This is particularly true of network analysis, like metabolic network construction, which is highly

reliant on functional annotation (Blais, Chavali, and Papin 2013). One effective way to analyze poorly annotated transcriptomes is co-expression analysis, which can infer functions of unannotated genes (Ozaki et al. 2010; Farahbod and Pavlidis 2020; Heyer, Kruglyak, and Yooseph 1999). By clustering expression trends, patterns may be identified and gene function predicted (Agostinho et al. 2024; Oyelade et al. 2016). Further, these clusters often show significant enrichment in specific functional categories which can be used to identify roles of unknown genes (D'haeseleer 2005). Weighted gene co-expression network analysis (WGCNA) is a popular systems biology method used to identify clusters of highly correlated genes, termed modules, and to relate these modules to external traits or conditions (Wu et al. 2022; Zhu et al. 2019; Chen et al. 2024). It is a useful method for understanding the relationships between annotated genes and to identify key genes or pathways that are involved in different biological processes, such as the treatment or the response to a disease (Zhao, Ma, and Zou 2021; Lin et al. 2021; Li et al. 2014). However, without functional annotation, it simply clusters the data. Other algorithms, such as unsupervised machine learning, can effectively cluster data as well. Therefore, it is unclear if WGCNA or unsupervised machine learning is most appropriate for poorly annotated expression data.

The focus of this work was the biotechnologically relevant stresses of nitrogen starvation, iron starvation, and salt stress. Nitrogen starvation has been shown to improve the production of fatty-acid-derived molecules in oleaginous organisms, including *Y. lipolytica* (Evans and Ratledge 1984; Papanikolaou and Aggelis 2002; Ageitos et al. 2011; Blazcek et al. 2014; Pomraning et al. 2016; Bellou et al. 2016; Chen et al. 2022), *D. hansenii*, and other *Debaryomyces* yeasts are also oleaginous, but to our knowledge, no study has investigated the impact of nitrogen starvation (Merdinger and Devine Jr 1965; Yaguchi, Rives, and Blenner 2017). Iron starvation has been shown in multiple organisms to increase flavinogenesis, including *Debaryomyces* yeasts (Tanner Jr, Vojnovich, and Van Lanen 1945; Demain 1972; Fedorovich, Protchenko, and Lesuisse 1999; Knight et al. 2002; Voronovsky et al. 2004; Poli et al. 2014; Abghari and Chen 2014; Li and Alper 2016; de Figueiredo et al.). Flavinogenesis is the process by which cells produce flavin adenine dinucleotide (FAD), a vital coenzyme for many metabolic reactions, and associated precursors such as the B-vitamin riboflavin. This process is essential for oxidative metabolism and is tightly regulated at both the transcriptional and posttranslational levels (Albacar et al. 2023; Li 2014). Salt stress or osmotic pressure is a key stressor for microbes, and *Debaryomyces* yeasts are particularly salt tolerant. Tolerance is managed by several key processes, including membrane transporter expression (Bremer and Krämer 2019; Nakov et al. 2020; Hoffmann and Bremer 2016; Csonka 1989). Under high salt stress, yeasts increase glycerol biosynthesis through the glycerolipid pathway to maintain osmotic balance (Hohmann 2002; Stratford et al. 2019; Michán et al. 2013; Navarrete et al. 2021). Therefore, each of these stressors affects the metabolism of fatty acids and vitamins of biotechnological interest.

Transcriptomics is an effective tool to measure gene expression response to changes in environmental conditions and elucidate yeast biology (Brion et al. 2016). Transcriptomic analyses for

Y. lipolytica have revealed phosphofructokinase (PFK) and pyruvate kinase (PK) to be affected by nitrogen starvation (Morin et al. 2011; Pomraning et al. 2016). *D. hansenii* transcriptomics have shown DhRpn4 to activate the transcription of proteasomal genes using the ScRpn4 binding site and provide resistance to various stressors (Karpov et al. 2017). Additionally, sodium and potassium trigger different regulation and expression responses in *D. hansenii* highlighting specific cellular processes in halotolerance (Navarrete et al. 2022). Despite these impactful studies, the transcriptomes studied have not been well-annotated, thus a large amount of data has not been included in the analysis.

Therefore, this study correlates the phenotypes and gene expression patterns of three strains, *Y. lipolytica* Po1f (ATCC MYA-2613), *D. subglobosus* dep8 (ATCC 20849), and *D. hansenii* CBS767 (ATCC 36239) in response to $(\text{NH}_4)_2\text{SO}_4$ starvation, $\text{Fe}(\text{Cl}_3)$ starvation, and 1 M NaCl stress. To our knowledge, there are no existing methods to compare unannotated transcriptomes across multiple non-model yeast species subjected to multiple treatments. Therefore, we pursued two analysis approaches, “network-first” and “cluster-first.” In the network-first approach, we manually constructed a metabolic map using the most comprehensive annotation we could find. Then, we mapped the gene expression data under each condition to this network. In the cluster-first approach, we clustered the data before annotation with WGCNA (Langfelder and Horvath 2008) and unsupervised machine learning. Both clustering methods produced qualitatively the same results. Subsequent annotation and function prediction of unannotated genes with AlphaFold revealed the specific cellular processes involved in the stress response – specifically metabolism, cell membrane composition and transport, redox, and gene regulatory proteins. Overall, the network-first strategy presents a clear picture of metabolic regulation that explains the observed stress phenotypes, while the cluster-first strategy reveals unanticipated biological processes and new gene targets for tolerance engineering.

2 | Methods

2.1 | Yeast Strains and Strain Cultivation

The strains used in this study are *Y. lipolytica* Po1f (ATCC MYA-2613), *D. hansenii* CBS767 (ATCC 36239), and *D. subglobosus* dep8 (ATCC 20849) which was derived from ATCC 20755. Routine cultivation was performed in Yeast Peptone Dextrose (YPD) media for all yeast species. YPD was purchased from (Sunrise Science Products #1877) and consists of 10 g/L yeast extraction, 20 g/L casein peptone with 20 g/L D-glucose (Sigma-Aldrich #G7021).

All conditions were grown in the base medium of 0.79 g/L of Complete Synthetic Medium (CSM) Powder (Sunrise Science Products, 1001-010), glucose (Sigma-Aldrich, G7021) The control media was supplemented with 6.71 g/L YNB+Nitrogen Powder (Sunrise Science Products, 1501-100), and the stress conditions were grown with their respective addition of salt or subtraction of nitrogen or iron. The salt stress medium was supplemented with 6.71 g/L YNB+Nitrogen powder (Sunrise Science Products, 1500-100), with 1 M NaCl (Sigma-Aldrich,

S3014). Nitrogen starvation medium was supplemented with 1.71 g/L YNB Powder (Sunrise Science Products, 1500-100). Iron starvation medium was supplemented with 6.61 g/L YNB+Nitrogen- CaCl_2 - CuSO_4 - FeCl_3 Powder (Sunrise Science Products, 1548-250), 100 mg/L Calcium chloride dihydrate (Sigma-Aldrich, 223506), and 0.04 mg/L Copper (II) sulfate pentahydrate (Sigma-Aldrich, 209198).

2.2 | Cell Growth Evaluation

Cell growth was evaluated in a plate reader assay. Each yeast was precultured in seed cultures. For seed cultures, single colonies were incubated in a Falcon tube with 5 mL YPD in an Innova 44 shaker (New Brunswick Scientific Co. Inc. Edison, New Jersey, USA), *Y. lipolytica* and *D. subglobosus* incubated overnight at 30°C and *D. hansenii* incubated overnight at 25°C. The stock cultures were then diluted (1: 1000) in a 250 mL Erlenmeyer flask with 25 mL of fresh YPD until an $\text{OD}_{600 \text{ nm}}$ of at least 1.00 ± 0.02 was reached. The cells were washed twice by centrifugation at 500g for 10 min at room temperature and culture supernatant was discarded. The cells were then re-suspended in a 5 mL CSM medium (control, nitrogen starvation, iron starvation, and salt stress) and precultured overnight. Each culture was diluted to an initial optical density ($\text{OD}_{600 \text{ nm}}$) of 0.1 using the same medium and the 96-well black wall plate (Thermo Fisher Scientific #165305) was prepared with 200 μL appropriate cell cultures. Cultivation was carried out in the Synergy H1 Multi-Mode Reader (BioTek Instrument, Vermont, Winooski USA) at room temperature with orbital shaking for 48 h, and the growth rate was monitored by measuring the OD_{600} every 30 min.

2.3 | Quantification of Riboflavin

Riboflavin quantification was measured by fluorescence on the BioTek H1M plate reader using an excitation wavelength of 440 nm and an emission wavelength of 520 nm. A calibration curve was made to correlate the fluorescence emitted with riboflavin concentration. Serial dilutions of solution riboflavin at 10 mg/mL were performed.

2.4 | Nile Red Assay

The fermentation broth samples were mixed with a 20 μL solution of DMSO/water (1:1, v/v,) and 0.05 mg/mL Nile Red (Invitrogen, N1142) and analyzed by the BioTek H1M plate reader. Nile red fluorescence was measured with excitation at 530/25 and emission at 590/35. A kinetic reading was set for 60 min with 60 s intervals, and the optical position was set to the top 50% of the well. OD_{600} was also measured concurrently. The fluorescence data of the total lipids were normalized by dividing the fluorescence unit by the background OD_{600} values.

2.5 | Statistical Analysis

All growth and metabolite data were calculated as the mean value with corresponding standard deviation obtained from three

independent experiments. The statistical analysis showed in riboflavin and total lipids quantification section were performed using Student's *t*-test and *p*-value < 0.05 were considered significant.

2.6 | RNA Isolation, Library Creation, and Read Processing

RNA isolation was performed from shake flask cultures to ensure the quantity and quality of isolation. The preculture yeasts as previously described were inoculated in 50 mL of CSM medium (standard, nitrogen starvation, iron starvation, and high salt) in 125 mL shake flasks. These were grown to saturation (48 h). 1 mL culture broth was pelleted at 16,000g for 2 min, resuspended in 293 μ L 50 mM EDTA with 7.5 μ L Zymolyase, and incubated for 45 min at 37°C. The samples were pelleted again and resuspended in 600 μ L TriReagent. Total RNA was isolated using the Zymo Research Direct-zol RNA Miniprep Plus kit according to the manufacturer's instructions. Processing methods were conducted by the Joint Genome Institute (JGI) (Singan et al.) who created stranded RNAseq libraries and quantified using qPCR. Sequencing was performed using the Illumina NovaSeq S4 platform with 2 $\times$ 151 paired-end reads. Reads are available on PRJNA1028627 NCBI BioProject. JGI QC pipeline filtered and trimmed the raw fastq file reads. BBDuk (Bushnell, Rood, and Singer 2017) was used to evaluate raw reads for artifact sequences by kmer (25) matching, allowing 1 mismatch. Detected artifacts were trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads and those containing any Ns were removed. Quality trimming was performed using the Phred trimming method set at Q6. Finally, after trimming, reads below the length threshold were removed, either the minimum length 25 bases or $\frac{1}{3}$ of the original read length, whichever was longer.

2.7 | gDNA Isolation, Genome Sequencing, Assembly, and Annotation

Yeast cell cultures were grown to saturation in shake flasks in YPD. gDNA was isolated for next-gen sequencing using the Wizard Genomic DNA purification kit according to the manufacturer's instructions. Nanopore and Illumina DNA library and the loading of MinION and iSeq 100, respectively, were performed according to the manufacturer's instructions. *D. subglobosus* and *Y. lipolytica* reference genomes were assembled, polished, and annotated using Prymetime (Collins et al. 2021). *D. hansenii* reference genome was sequenced by Génolevures [Release Version: 2012/02/09]. *Y. lipolytica* PO1f was re-sequenced and published in Collins et al. (2021). *D. subglobosus* dep8 was sequenced and assembled by Prymetime and annotated using the JGI Fungal Annotation Pipeline (Kuo, Bushnell, and Grigoriev 2014). Raw genome sequences can be found on NCBI with the following accession IDs: *Y. lipolytica* PO1f – PRJNA1047134, *D. subglobosus* dep8 – PRJNA1047133.

2.8 | Transcriptome Alignment and Counting

Filtered reads from each library were aligned to the reference genomes using HISAT2 version 2.2.0 (Kim, Langmead, and

Salzberg 2015). JGI Annotation Pipeline was used to add additional annotation of functions to known and predicted proteins. featureCounts (Liao, Smyth, and Shi 2014) was used to generate the raw gene counts file using gff3 annotations. Only primary hits assigned to the reverse strand were included in the raw gene counts (-s 2 -p -primary options). Raw gene counts were used to evaluate the level of correlation between biological replicates using Pearson's correlation and determine which replicates would be used in the differential gene expression (DGE) analysis. One sample for *D. hansenii*'s (NH₄)₂SO₄ starvation condition correlated poorly with its replicate group and was excluded from the DGE analysis. Raw transcriptome sequences can be found on NCBI with the following accession ID: PRJNA1028627. All scripts and data are available on GitHub (<https://github.com/emyounglab/rnaseq-onboarding>).

2.9 | DGE Analysis and Annotation

DGE analysis was performed with EdgeR (Robinson, McCarthy, and Smyth 2010) and DESeq2 (Love, Huber, and Anders 2014) using the respective R packages. For DESeq2, regularized logarithmic transformation values were used to improve distancing and clustering for PCA and hierarchical cluster visualizations. ashR shrinkage estimator was used to help rank genes by log₂FC across genes. Only genes with FDR-adjusted *p*-values less than 0.05 with log₂FC greater than 1 or less than -1 for each condition were used for downstream analysis. collapseReplicates was used to sum read counts within the three technical replicates on each biological growth condition. PCAs and MA-plots were used to visualize transcriptomic data from shrunk log₂FC. UpSet (Lex et al. 2014) plots are an alternative for quantitative set intersection visualizations, using UpSetR (Conway, Lex, and Gehlenborg 2017), an R package, to visualize set intersections allowed us to further compare the differentially expressed gene (DEG) variation of organisms. In addition to JGI's annotation provided, further functional annotation was performed with ERGO (Overbeek et al. 2003), EggNOG-mapper (Huerta-Cepas et al. 2019), and WebMGA (Wu et al. 2011) to obtain Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000), Gene Ontology (GO) (Ashburner et al. 2000), and euKaryotic Orthologous Groups (KOG) database (Tatusov et al. 2003). All scripts and data are available on GitHub (<https://github.com/emyounglab/rnaseq-onboarding>).

2.10 | Clustering

2.10.1 | WGCNA

Weighted gene co-expression network analysis (WGCNA), performed using the R package (Langfelder and Horvath 2008), is a method for identifying relationships between genes based on their expression levels across different samples or conditions. The gene expression data was pre-processed by removing any missing or low-quality data, normalizing the data, and removing any highly correlated genes. Differentially expressed genes that were significant using FDR-adjusted *p*-values less than 0.05

with $\log_2FC$ greater than 1 or less than -1 were used for clustering, further referenced as significant DEGs. The gene co-expression network was constructed using Pearson's correlation coefficient, which was used to calculate the correlation between each pair of genes and construct a network of genes based on these correlations using $\text{minModuleSize} = 30$. The edges of the network were then weighted based on the strength of the relationship between the genes using the topological overlap measure (TOM). Finally, modules were determined by identifying groups of genes that are more highly connected to each other than to genes in other modules. These modules can be identified using clustering algorithms or other methods. All scripts and data are available on GitHub (<https://github.com/emyounglab/rnaseq-onboarding>).

2.10.2 | Unsupervised Machine Learning

The significant DEGs were clustered using the k-means algorithm implemented in Python (v3.9) (Van Rossum and Drake Jr 1995). To determine the ideal number of clusters, the within-group sum of squares (WCSS) method, also known as the elbow method, was employed. This analysis utilized Matplotlib (v3.6) (Hunter 2007) for data visualization. Both Lloyd's and Elkan's methods were utilized within the k-means algorithm, implemented through scikit-learn (v1.3) (Pedregosa et al. 2011). Dimensionality reduction techniques, including T-distributed Stochastic Neighbor Embedding (t-SNE) from scikit-learn and kernel Principal Component Analysis (kPCA), were applied. The implementation of kPCA involved the use of scikit-learn. For comparison purposes, different kernels within kPCA were evaluated using scikit-learn. SKLearn NearestNeighbors from scikit-learn was employed to identify cluster centers, while Matplotlib facilitated the visualization of these centers as representative points. All scripts and data are available on GitHub (<https://github.com/emyounglab/rnaseq-onboarding>).

K-means clustering was used to cluster the data. To identify the optimal number of clusters, the Within-Cluster Sum of Squares (WCSS) method, also known as the Elbow method, was used to sum the squared distance between each point and the centroid in a cluster to identify the elbow, indicating the appropriate number of clusters. The elbow is the point on the graph where the curve visibly bends, where the x-axis of the elbow represents the number of clusters and the y-axis of the elbow represents the Within Clusters Sum of Squares (WCSS) for each number of clusters. To implement the k-means algorithm, both Lloyd's and Elkan's methods were compared. K-means takes an input data set of size n and a specified number of clusters, k , and labels each of the n points into one of the k clusters based on the centroid to which it is closest. Lloyd's algorithm is an iterative method that alternates between assigning centroids and classifying points until the assignment no longer changes. We continued with Elkan's algorithm, as it is similar to Lloyd's, but it uses a triangular inequality to reduce the cost of k-means. In addition to k-means, different dimensionality reduction techniques were implemented using manifold learning to determine the most appropriate method for the data set. This included T-distributed Stochastic Neighbor Embedding (t-SNE), which focuses on the local structure of the data and can be useful for

datasets with multiple manifolds and kPCA, which is defined by a specified kernel. For a complete analysis, several kernels within kPCA were compared, including linear, radial basis function, sigmoid, and cosine. Once clusters were identified, SKLearn's (Pedregosa et al. 2011) NearestNeighbors was used to identify centers and then plot those as cluster representatives, similar to WGCNA's module eigengene.

3 | Results

3.1 | Initial Phenotyping

Growth, total lipids, and riboflavin was characterized for each species under the three stress conditions – $(\text{NH}_4)_2\text{SO}_4$ (nitrogen) starvation, $\text{Fe}(\text{Cl}_3)$ (iron) starvation, and NaCl (salt) stress – and control (Figure 1a). Growth curves, quantified by optical density at 600 nm over 48 h, are shown in Supporting Information S1: Figure 1. *D. hansenii* had nearly double the final optical density of *Y. lipolytica* in all conditions. *D. subglobosus* grew equivalent to *Y. lipolytica* except under salt stress, where it increased. This is evidence of the halophilic nature of some *Debaryomyces* isolates, even though this increase was not observed in *D. hansenii*. Total lipids are shown in Figure 1b, measured by Nile Red staining and reported as relative fluorescence units (RFU) per unit of optical density (absorbance at 600 nm). The *Debaryomyces* yeasts produced more lipids than *Y. lipolytica* in the control condition. Nitrogen starvation increased lipid production in *Y. lipolytica* nearly sixfold but decreased production in the *Debaryomyces* yeasts. Each species showed a decrease in lipids under iron starvation, resulting in roughly equal amounts. Under salt stress, *D. hansenii* produced the most lipids, *Y. lipolytica* produced slightly less, but *D. subglobosus* showed a sharp decrease in lipid production. This is the same condition in which *D. subglobosus* exhibited a growth increase. Riboflavin production is shown in Figure 1c, as measured by fluorescence calibrated to a standard curve (Supporting Information S1: Figure 2). *D. subglobosus* produced significantly more than *Y. lipolytica* and *D. hansenii* under all conditions. *D. subglobosus* under iron starvation produced the most riboflavin, while it produced the least riboflavin under salt stress. While not obvious in Figure 1c, *D. hansenii* produced significantly more riboflavin than *Y. lipolytica* under salt stress, with a p -value of 0.0006, and under iron starvation, with a p -value of 0.0013. These phenotyping results were used to evaluate the RNA-Seq results for these conditions.

3.2 | Transcriptomics and Differential Gene Expression

As shown in Figure 1a, total RNA was also isolated under each condition for each yeast species. RNA-Seq of these samples was performed in collaboration with the Department of Energy Joint Genome Institute. Genome sequences of each yeast were used to curate the reads (Methods). The *D. hansenii* CBS767 genome was obtained from Génolevures [Release version: 2012/02/09], the *Y. lipolytica* Po1f genome was the resequenced genome from Collins et al. (2021), and the *D. subglobosus* genome was a *de novo* assembly using the same approach as *Y. lipolytica*.

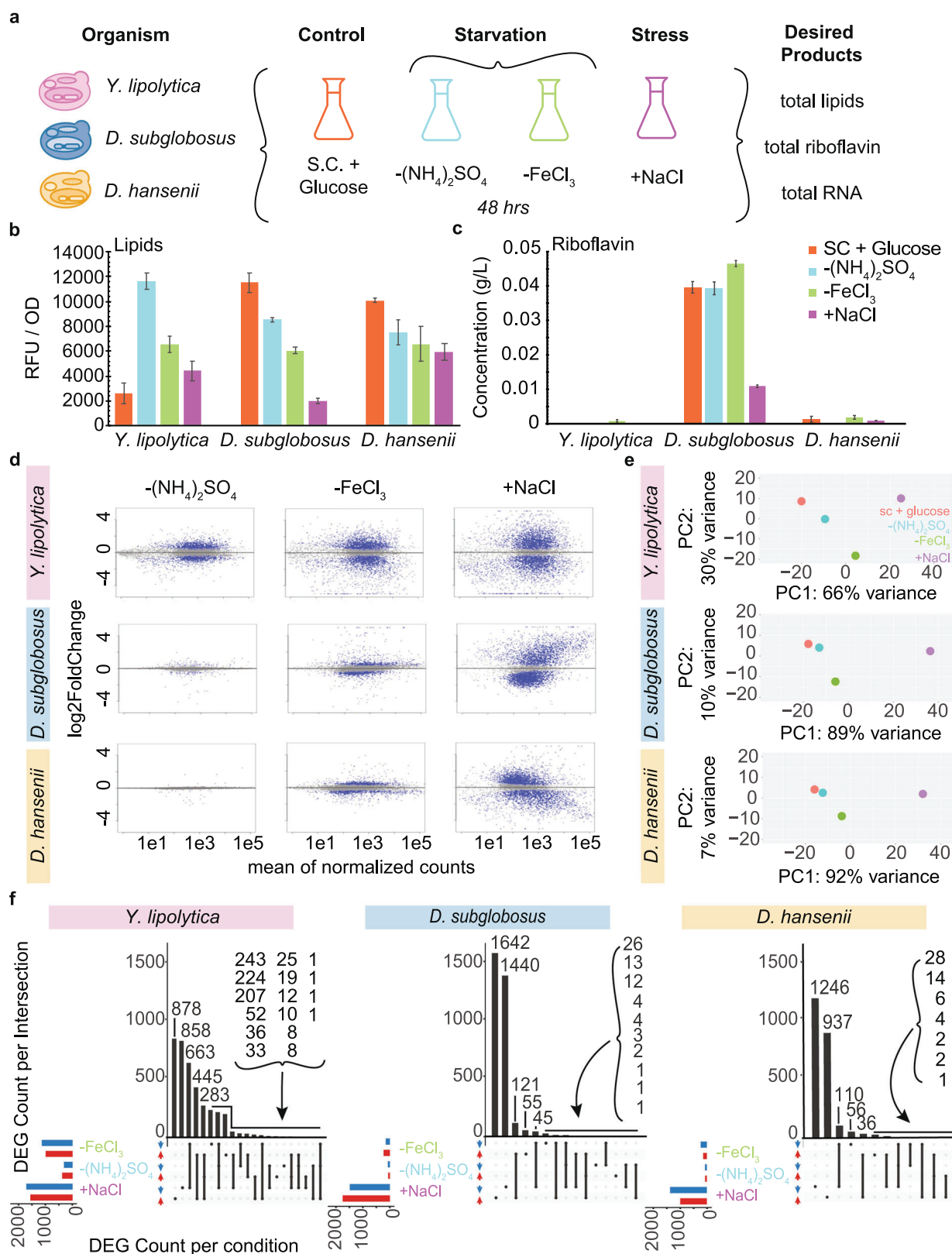


FIGURE 1 | Phenotyping and transcriptomic. (a) Overview of the conditions and organisms tested in this study. For each condition and organism, total lipids, riboflavin, and total RNA were extracted. (b) Total lipids per cell after 48 h cultivation as calculated by Nile Red stain (RFU) divided by the optical density (OD). (c) Riboflavin production after 48 h cultivation, as measured by correlating fluorescence to a standard curve. (d) MA plots of the RNA-Seq results, plotting log fold-change versus mean expression for each treatment against the control. (e) PCA plots of the RNA-Seq results. (f) UpSet plots of the RNA-Seq results, showing how many genes were upregulated and downregulated in each condition (bar chart in bottom left), and whether those genes were unique to a condition or shared across conditions (bar chart in upper right). Dots indicate appearing uniquely in a condition, and lines indicate a connection between conditions for each gene group. The number of genes in each group is shown next to each bar.

Genome quality statistics are included in Supporting Information S1: Tables 1 and 2. Differential gene expression was then calculated for each transcript using the control condition as the baseline.

The differential gene expression data was first visualized by an MA plot (Figure 1d). MA plots show the $\log_2$ fold change for each gene on the y-axis and the average expression for each gene on the x-axis (Love, Huber, and Anders 2014). Qualitatively, *Y. lipolytica* had many differentially expressed genes (DEGs) across all conditions, while *D. subglobosus* and *D. hansenii* had few DEGs under nitrogen and iron starvation. Salt stress had a large effect on the transcriptomes of all organisms, with both *Debaryomyces* species showing a large number of DEGs. Principal component analysis (PCA) was then applied to reduce the dimensionality of the datasets and compare variance, shown in Figure 1e. A similar pattern is observed across species, with control and nitrogen starvation relatively close, iron starvation close in PC1 but not in PC2, and salt stress close in PC2 but not PC1. This is particularly true for the *Debaryomyces* yeasts, where PC1 explains 90% of the variance. Finally, to get a sense of whether the DEGs were similar across conditions, the data was plotted using Upset plots (Figure 1f). Each Upset plot contains a bar chart in the bottom left showing the total number of DEGs in each condition. Blue bars are downregulated genes and red bars are upregulated genes. The right part of the plot shows the DEGs that are unique to each condition, and the DEGs that appear in two conditions. For all organisms, the salt stress condition resulted in the most DEGs – 878 downregulated and 858 upregulated DEGs in *Y. lipolytica*, 1246 downregulated and 937 upregulated genes in *D. hansenii*, and 1642 downregulated and 1440 upregulated genes in *D. subglobosus*. There were more intersections in *Y. lipolytica* than *D. hansenii* and *D. subglobosus*, suggesting more overlap between the groups. The largest intersection between conditions in *Y. lipolytica* was 663 DEGs, between downregulated salt stress and downregulated iron starvation, whereas there were only 110 DEGs in *D. hansenii* between the same conditions. The *Debaryomyces* yeasts did not have any other large groups outside of salt stress, corroborating the MA and PCA plots.

Taken together, this analysis points to a specialized response to salt stress in *Debaryomyces* yeasts, while *Yarrowia* more generally responds to stresses. Additionally, this shows that the increase in lipid production by *Y. lipolytica* under nitrogen starvation is accompanied by a large gene expression response that is not seen in *Debaryomyces*.

3.3 | Functional Annotation, Mapping, and Clustering

To map the DEGs to known genes, the RNA-Seq data was annotated using the Enzyme Commission (EC) number, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and Gene Ontology (GO) terms. However, these annotation methods did not annotate a large fraction of the DEGs in each organism. For example, EC number annotated only 843 (24%) in *Y. lipolytica*, 347 (15%) in *D. hansenii*, and 767 (32%) in *D. subglobosus*. KOG (euKaryotic Orthologous Groups), although

less widely adopted, yielded many more annotations. KOG annotated 2318 (67%) in *Y. lipolytica*, 2359 (72%) in *D. hansenii*, and 1159 (49%) in *D. subglobosus*. Therefore, KOG annotations were used to label heatmaps of the 10 highest and lowest DEGs in each condition (Supporting Information S1: Figures 3, 4, and 5 for *D. hansenii*, *D. subglobosus*, and *Y. lipolytica*, respectively). However, KOG is not compatible with many of the most common DEG analysis software packages. Therefore, a metabolic map of central carbon metabolism was constructed manually with the available annotations (network-first analysis), and the data was clustered with two annotation-free approaches – Weighted Gene Correlation Network Analysis (WGCNA) and k-means clustering (cluster-first analysis). WGCNA constructs a gene co-expression network by calculating pairwise correlations between gene expression profiles and then transforming these correlations into a weighted network, where edges represent the strength of co-expression relationships. By applying hierarchical clustering to this network, WGCNA identifies modules of genes that exhibit similar expression patterns across samples (Langfelder and Horvath 2008). WGCNA is typically used on well-annotated datasets, as the last two steps in WGCNA's method include functionally annotating the modules, and is not compatible with KOG annotations. Thus, we also applied a general unsupervised machine learning algorithm, k-means clustering (Pedregosa et al. 2011). The results from both clustering approaches were then separately annotated with KOG. Unknown transcripts with high differential gene expression across the clusters were then input to AlphaFold to attempt to infer function. Thus, two methods, one classical and one machine learning, were applied to identify functional classes beyond the metabolic enzymes captured in the network-first analysis and to infer the function of unannotated genes. The results of these analysis approaches are described in the following sections.

3.4 | Network-First Analysis

We manually reconstructed a metabolic map of central carbon metabolism, flavinogenesis, and lipid biosynthesis and then mapped KOG annotated transcripts to each gene in the network (Figure 2). The map is split into seven pathways (glycolysis, glycerolipid biosynthesis, fatty acid biosynthesis, fatty acid degradation, pentose phosphate, purine, and riboflavin). For each gene, a 3×3 heatmap is shown for the change in gene expression for each species and each condition, with red intensity corresponding to upregulation, gray to no significant change, and blue to downregulation.

In the glycolysis pathway, *Y. lipolytica* upregulates most genes in response to any stress, while the *Debaryomyces* yeasts show only a small, mixed response under salt stress. The exception to these trends is enolase (ENO), which *Y. lipolytica* downregulates and the *Debaryomyces* yeasts upregulate in response to salt stress. Glycerolipid biosynthesis, fatty acid biosynthesis, and fatty acid degradation all contribute to total lipids and the production of the key osmoprotectant glycerol (Xue, Chen, and Jiang 2017; Ohlrogge and Browse 1995). *Y. lipolytica* Po1f upregulates most genes in these three pathways across stress conditions, indicating a broad metabolic shift. The glycerol

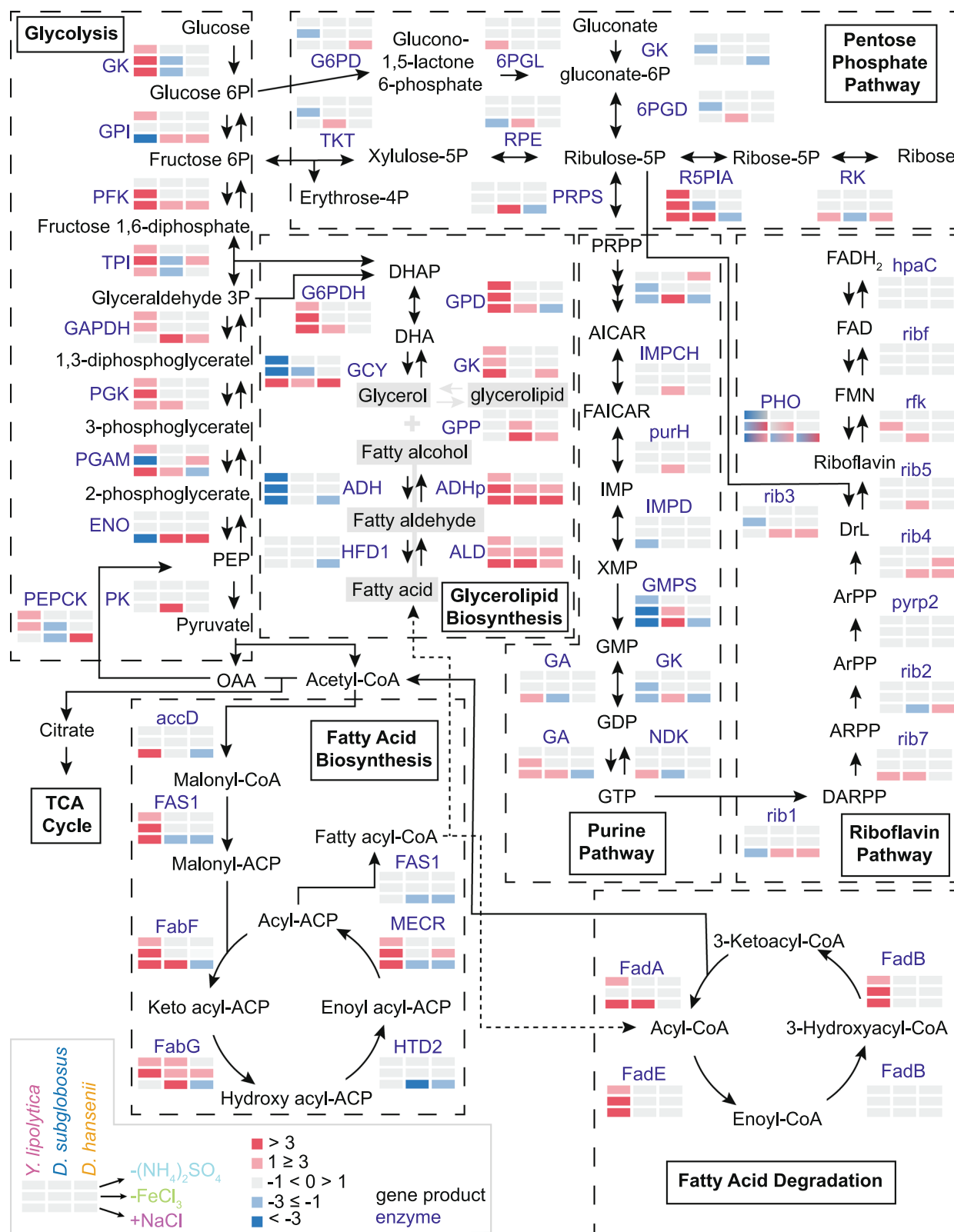


FIGURE 2 | Legend on next page.

node in particular exhibits upregulation of all genes that produce glycerol – both from glycerolipid and from glycolysis – with the greatest upregulation in the salt stress condition. The *Debaryomyces* yeasts do not dramatically change gene expression in fatty acid biosynthesis or degradation. Yet, several genes that make fatty molecules that are combined with glycerol to make glycerolipids are upregulated. These results show a different response in lipid and glycerol biosynthesis between *Yarrowia* and *Debaryomyces*, which corroborates the observed phenotypic difference, where *Yarrowia* increases lipid pools and *Debaryomyces* decreases lipid pools in response to stress (Figure 1b). The pentose phosphate pathway, the purine pathway, and the riboflavin pathway all contribute to the production of riboflavin and other cofactors. At the key node of ribulose-5P, *Y. lipolytica* upregulates R5PIA toward ribose production, and downregulates the purine pathway. In contrast, *D. subglobosus* upregulates the purine pathway and riboflavin biosynthesis in response to salt stress. This is an unexpected result, since *D. subglobosus* shows a massive decrease in riboflavin pools under salt stress (Figure 1c). Yet, *D. subglobosus* also upregulates the kinase that converts riboflavin to FMN (*rfl* in Figure 2) and downregulates the phosphatase that converts FMN to riboflavin (PHO in Figure 2). This indicates that salt stress does not cause *D. subglobosus* to repress flavinogenesis but rather it stimulates the production of flavin cofactors that draw down the riboflavin pools. Taken together, the network-first analysis supports the conclusions from the initial analysis of the datasets – that *Y. lipolytica* Po1f has a large change in gene expression response, while the *Debaryomyces* yeasts only alter the expression of a

small number of genes. It also provides insight that this general observation is due to *Y. lipolytica* broadly upregulating glycolysis, lipids, and glycerol pathways, while the *Debaryomyces* yeasts upregulate glycerol production from glycerolipids only while converting riboflavin to flavin cofactors in salt.

3.5 | Cluster-First Analysis

3.5.1 | WGCNA

WGCNA outputs a dendrogram of all genes organized in a Topological Overlap Matrix (TOM) that it uses for clustering. This is displayed in Figure 3a as a heatmap, with darker colors representing increasing gene similarity. WGCNA identified six modules in *Y. lipolytica*, three in *D. subglobosus*, and three modules in *D. hansenii* with a minimum module size of 30 genes (Supporting Information S1: Figure 6 and Table 3). In Figure 3a, the module assignments are represented by colored rectangles on the side and top of the TOM, next to the gene dendrogram. These colors are assigned specifically in descending order of module size (turquoise, blue, brown, yellow, green, and red). Modules with similar eigengene patterns were observed across organisms. Therefore, the modules were reorganized into three supertypes – Group 1 (upregulated in salt), Group 2 (downregulated in salt and other stresses), and Group 3 (upregulated in iron starvation). Figure 3b displays the modules grouped by organism and supertype. The color assigned by WGCNA and the module size are labeled on each eigengene

FIGURE 2 | Metabolic network with a heatmap of gene expression change for each step in each organism and each condition. The key at the bottom left illustrates the layout of each heatmap next to each metabolic step in the metabolic map. It shows that the left column is *Yarrowia lipolytica*, the middle column is *Debaryomyces subglobosus*, the right column is *D. hansenii*, while the top row is nitrogen starvation, the middle row is iron starvation, and the bottom row is salt stress. High $\log_2$ FC is marked in red, low $\log_2$ FC is marked in blue, while no significant change is marked in gray. Compounds are shown in black text, genes are in blue text. In the glycerolipid pathway, the highlighted genes indicate the different fatty molecules that can be combined with glycerol to make glycerolipids. Abbreviations, organized by pathway, are listed as follows. Glycolysis pathway: glucokinase (GK), glucose 6-phosphate, glucose 6-phosphate isomerase (GPI), fructose 6-phosphate, fructose 1,6-diphosphate, (TPI), glyceraldehyde 3-phosphate, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), 1,3-diphosphoglycerate, phosphoglycerate kinase (PGK), 3-phosphoglycerate, 2, 3-biphosphoglycerate-dependent phosphoglycerate mutase (PGAM), 2-phosphoglycerate, enolase (ENO), phosphoenolpyruvate (PEP), pyruvate kinase (PK), pyruvate, oxalacetate (OAA), phosphoenolpyruvate carboxykinase (PEPCK). Pentose phosphate pathway: glucose-6-phosphate 1-dehydrogenase (G6PD), glucono-1,5-lactone 6-phosphate, 6-phosphogluconolactonase (6PGL), gluconate 6-phosphate, gluconate, gluconokinase (GK), 6-phosphogluconate dehydrogenase (6PGD), ribulose 5-phosphate, transketolase (TKT), erythrose 4-phosphate, xylulose 5-phosphate, ribulose-phosphate 3-epimerase (RPE), ribose, ribokinase (RK), ribose 5-phosphate, ribose 5-phosphate isomerase A (R5PIA), ribose-phosphate pyrophosphokinase (PRPS). Glycerolipid biosynthesis pathway: glucose-6-phosphate dehydrogenase (G6PDH), dihydroxyacetone phosphate (DHAP), glycerol 3-phosphate dehydrogenase (GPD), dihydroxyacetone (DHA), glycerol 2-dehydrogenase (DCY), glycerol, glycerol 3-phosphate (GPP), glycerol kinase (GK), lipid, aldehyde dehydrogenase (ALD), fatty aldehyde dehydrogenase (HFD1), fatty aldehyde, alcohol dehydrogenase propanol-prefering (ADHp), alcohol dehydrogenase (ADH). Purine pathway: 5-phospho- α -D-ribose 1-diphosphate (PRPP), 5-aminoimidazole-4-carboxamide ribotide (AICAR), phosphoribosylaminoimidazolecarboxamide formyltransferase (IMPCH), 5'-Phosphoribosyl-5-formamido-4-imidazolecarboxamide (FAICAR), phosphoribosylaminoimidazolecarboxamide formyltransferase (purH), (IMPT), inosine 5'-phosphate dehydrogenase (IMPD), xanthosine 5'-phosphate (XMP), GMP synthase (GMPS), guanosine 5'-phosphate (GMP), golgi apyrase (GA), guanylate kinase (GK), guanosine diphosphate (GDP), nucleoside-diphosphate kinase (NDK), guanosine 5'-triphosphate (GTP). Riboflavin pathway: riboflavin 1 (rib1), 2,5-diamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one (DARPP), riboflavin 7 (rib7), 2,5-diamino-6-(5-phospho-D-ribitylamino)pyrimidin-4(3H)-one (ARPP), riboflavin 2 (rib2), 5-amino-6-(5'-phospho-D-ribitylamino)uracil (ArPP), 5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase (prp2), 5-amino-6-(D-ribitylamino)uracil (5-ArPP), (rib4), 6,7-dimethyl-8-(D-ribityl)lumazine (DrL) riboflavin 3 (rib3), riboflavin 5 (rib5), riboflavin, acid phosphatase (PHO), riboflavin kinase (*rfl*), flavin mononucleotide (FMN), riboflavin f (ribf), flavin adenine dinucleotide (FAD), flavin reductase (*hpaC*) ($FADH_2$). Lipid biosynthesis: Acetyl-CoA, acetyl-CoA carboxylase (*accD*), malonyl-CoA, lipid synthase (FAS1), malonyl-ACP, 3-oxoacyl-ACP synthase II (FabF), ketoacyl-ACP, 3-oxoacyl-ACP reductase (FabG), hydroxyacyl-ACP, 3-hydroxyacyl-thioester dehydratase (HTD2), enoyl acyl-ACP, mitochondrial enoyl-ACP reductase (MECR), acyl-ACP, fatty acyl-ACP synthase (FAS), fatty acyl-ACP. lipid degradation: acyl-CoA, (FadE), enoyl-CoA, enoyl-CoA hydratase (*echA*), 3-hydroxyacyl-CoA, 3-hydroxyacyl-CoA dehydrogenase (HADH), 3-ketoacyl-CoA.

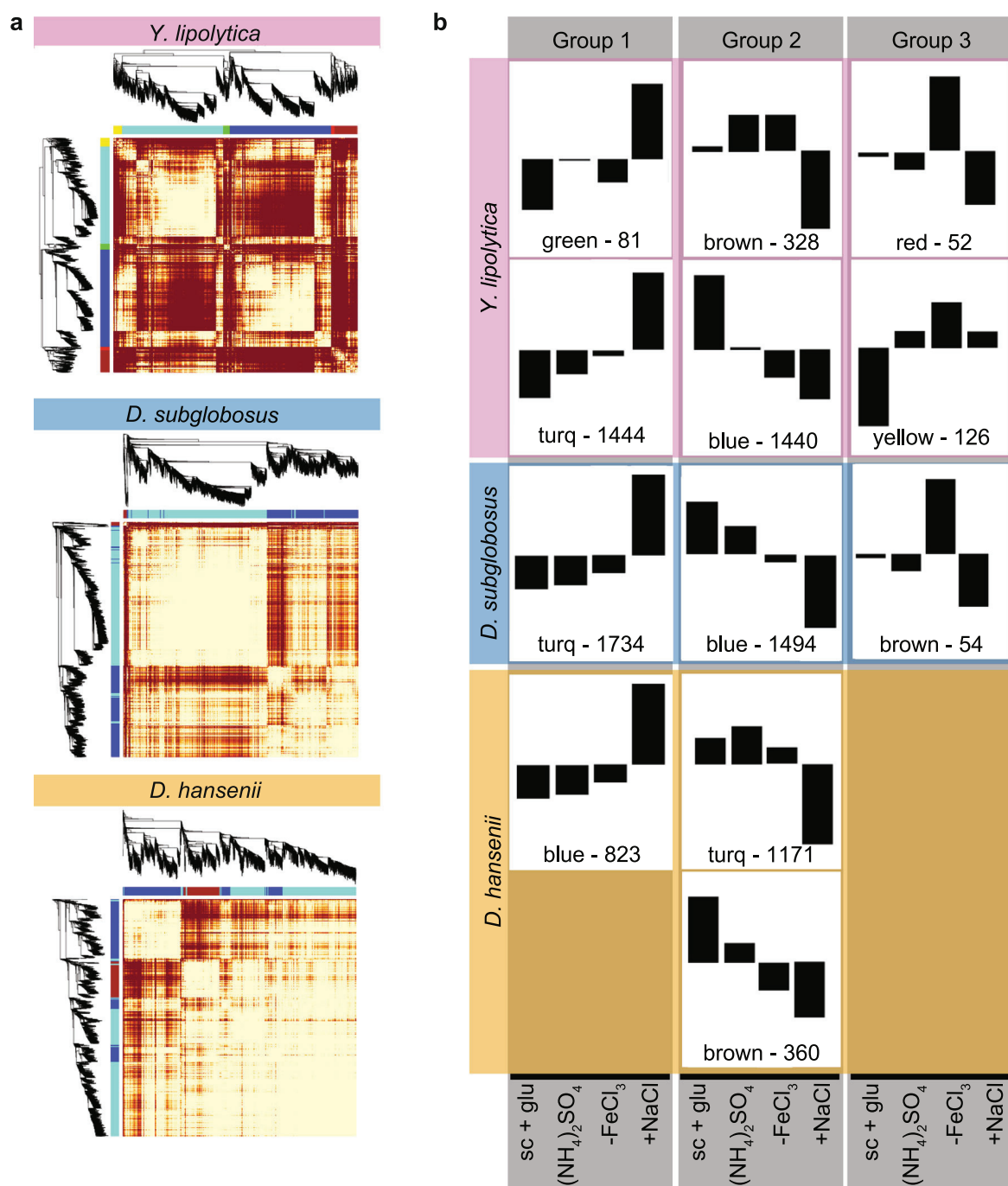


FIGURE 3 | Clustering with WGCNA. (a) Topological overlap matrix output from WGCNA, showing a gene dendrogram used for module determination, the colored rectangles showing final module assignment, and a heatmap comparing all genes where darker color represents more expression similarity. (b) The clusters produced by WGCNA are grouped by organism and supertype. The bar charts plot the module eigengenes for each condition. WGCNA module names are assigned in descending order from the largest cluster for each organism, and the number of genes in each cluster is shown next to the name.

plot. It is interesting to note that even though *Y. lipolytica* showed a response across stress conditions, these responses resulted in similar eigengene patterns to the other organisms. Furthermore, with WGCNA it was possible to illustrate once again the challenges of annotation for these organisms. Within the clusters, over 75–85% of genes were unannotated with EC number (Supporting Information S1: Table 4), while only 23–50% were unannotated with KOG (Supporting Information S1: Table 5).

In *Y. lipolytica*, six modules were identified (Figure 3b), two of which were in supertype Group 1 (upregulated in salt) – turquoise and green. Turquoise is the largest module with 1444 genes, 1108 of which were annotated by KOG. In turquoise, the three highest annotated DEGs in salt stress were a predicted transporter ($\log_2\text{FC}$ 8.78), dihydroxyacetone kinase/glycerone kinase ($\log_2\text{FC}$ 6.01) (Molin, Norbeck, and Blomberg 2003), and another predicted transporter ($\log_2\text{FC}$ 5.76). This is expected, since ion transport and glycerol biosynthesis are two

essential components of salt tolerance in yeast. WGCNA also calculates “module membership,” which defines how well each gene fits into a module by comparing its expression with the module’s main pattern. Among the top 10 annotated genes by module membership in turquoise were ribulose kinase and related carbohydrate kinases (3.00), F-actin capping protein (3.57), transcription initiation factor IIF (2.03), protein tyrosine phosphatase-like protein PTPLA (2.61), malonyl-CoA:ACP transacylase (2.79), thymidylate kinase/adenylate kinase (2.34), histone H3 (Lys4) methyltransferase complex (3.34), histone deacetylase complex (iron starvation 2.23, salt stress 3.87), methylthioadenosine phosphorylase (MTAP) (iron starvation 2.63, salt stress 3.87), and vacuolar H⁺-ATPase V0 sector (2.68). These genes point to the observed increase in lipid biosynthesis, as well as increases in translation and epigenetic processes that could explain the broad shifts in gene expression patterns. In the green module, only three of the highest DEGs in salt were annotated: cyclin B (3.49, nitrogen starvation 2.46, iron starvation 2.83), reductase (3.17), and multicopper oxidase (3.09). Cyclin B is involved in cell cycle regulation (Epstein and Cross 1992), and the fact that it was upregulated across all stress conditions also points to the broader response *Y. lipolytica* had to stress. Reductase proteins likely manage oxidative stress by reducing reactive oxygen species (ROS) (Herrero et al. 2008) and multicopper oxidase proteins are essential for iron homeostasis and oxidative stress management (Solomon, Sundaram, and Machonkin 1996), therefore these genes could be overexpressed because of the generally increased metabolism seen for *Y. lipolytica* in the salt condition, which would generate more ROS.

Two modules in *Y. lipolytica* were in supertype Group 2 (downregulated in salt) – blue and brown. The blue module was the second largest cluster with 1440 genes and 892 annotated, exhibiting high expression under control but low expression under iron starvation and salt stress. The brown module, containing 328 genes with 183 annotated, had high expression in nitrogen starvation and iron starvation, but low expression under salt stress. For the blue module, the three lowest DEGs in salt were cobalamin synthesis protein (nitrogen starvation –4.55, iron starvation –7.05, salt stress –8.18), an unannotated gene (nitrogen starvation –2.4, iron starvation –3.62, salt stress –7.98), and a Fe²⁺/Zn²⁺ regulated transporter (nitrogen starvation –2.21, iron starvation –10.40, salt stress –7.49). These genes are involved in vitamin B12 synthesis (Balabanova et al. 2021), potential stress response roles, and metal ion transport (Cyert and Philpott 2013), respectively. The lowest salt stress DEG in the brown module was a thyroid hormone receptor-associated protein complex, subunit TRAP230 (–9.34), indicating its potential role in transcription regulation under stress (Fondell et al. 1999). The next lowest DEG was a molecular chaperone (small heat-shock protein Hsp26/Hsp42) with a log₂FC value of –6.94, crucial for protein folding and protection under stress (Verghese et al. 2012). It is interesting to note that while elevated in the starvation conditions, these genes are severely downregulated in salt. This indicates that there are different tolerance mechanisms at play between the stress conditions.

The final two modules in *Y. lipolytica* were in supertype Group 3 (upregulated in iron starvation). The yellow

module, with 126 genes, had an eigengene pattern showing upregulation in all stress conditions compared to control, while the red module, with 52 genes, showed exclusive upregulation in iron starvation. In the yellow module, two of the top five DEGs under iron starvation were unannotated (log₂FC values of 3.99 and 3.08), while the annotated genes included a phospholipid methyltransferase (3.36), a sister chromatid cohesion complex subunit PDS5 (3.25), and an FGF receptor-activating protein 1 (3.04), indicating that membrane synthesis, chromosome segregation, and signaling are activated under iron-limited conditions (Kanipes and Henry 1997; Hartman et al. 2000; Kinoshita, Minshull, and Kirschner 1995). In the red module, the highest significantly annotated DEG was a “protein similar to a predicted member of the intramitochondrial sorting protein family” (2.49), pointing to the known importance of mitochondrial function under iron starvation (Stojanovski et al. 2012). There were also four unannotated genes with log₂FC values ranging from 3.57 to 2.51.

WGCNA identified three modules in *D. subglobosus* (Figure 3b). Turquoise was the largest and contained 1734 genes, 1220 of which were annotated. It is in supertype Group 1 (upregulated in salt). The top eight salt log₂FC values (8.01, 7.81, 6.91, 6.63, 6.52, 6.25, 5.82, and 5.42) were unannotated. Among the annotated genes, Cu²⁺/Zn²⁺ superoxide dismutase (SOD1) showed high upregulation with log₂FC values of 5.22 and 4.87, suggesting its importance in managing oxidative stress induced by high salt levels (Fujii et al. 1996). Another gene, the DNA-binding protein jumonji/RBP2/SMCY with a JmJc domain, had a log₂FC of 4.73, indicating its role in chromatin remodeling and gene expression regulation under salt stress (Christensen et al. 2007).

The blue module in *D. subglobosus* comprised 1494 genes with 1100 annotated and was in supertype Group 2 (downregulated in salt). The lowest salt stress log₂FC values were for biotin synthase (Ahmad and Rose 1962) (–4.78), a transcription factor NSD1 (Fang et al. 2021) (–3.35), and an amino acid transporter (Grenson 1992) (–4.34). These genes were also among the lowest annotated DEGs in the blue cluster under iron and nitrogen starvation: biotin synthase (–4.14 iron, –4.30 nitrogen), transcription factor NSD1 (–6.52 iron, –4.08 nitrogen), and the amino acid transporter (–4.30 iron, –3.58 nitrogen). This genes may be more generally involved in stress response than the representative eigengene pattern suggests for this cluster.

The smallest module in *D. subglobosus* was brown which contained 54 genes, 37 of which were annotated. It is in supertype Group 3 (upregulated in iron starvation). The top three DEGs (log₂FC values of 11.20, 11.09, and 10.99), were unannotated. Among the top 25 DEGs, 18 were transporters with log₂FC values ranging from 8.88 to 3.95, likely involved in the uptake and redistribution of nutrients and ions under iron-limited conditions. Of the genes with the highest module membership, seven were amino acid transporters (Grenson 1992) and the others included urea and uridine transporters (Vickers et al. 2000), which implicate nitrogen metabolism, nucleotide metabolism, and general cellular transport processes in iron stress response.

In *D. hansenii*, WGCNA identified three modules (Figure 3b). Two – turquoise and brown – are in supertype Group 2 (downregulated in salt). Turquoise contained 1171 genes, 577 of which were annotated. Among the top 50 genes with the most module membership, 32 were annotated. Yet, the lowest three DEGs were unannotated. The fourth lowest, with a $\log_2FC$ of -4.34 , was identified as a synaptic vesicle transporter SVOP, part of the major facilitator superfamily, highlighting its role in cellular processes affected by high salt levels (Li et al. 2020). Brown contained 360 genes, with 215 annotations. This module differentiated between control and nitrogen starvation conditions (high) from iron starvation and salt stress (low). Under iron starvation, five genes had $\log_2FC$ values greater than 2, including a urea transporter (2.8), two amino acid transporters (2.67 and 2.05), an ornithine aminotransferase (2.66), and an unannotated gene (2.28), highlighting changes in nitrogen metabolism and transport (Cooper and Sumrada 1975; Bianchi et al. 2019; Dougherty et al. 1993). The five lowest salt $\log_2FC$ values included three unannotated genes (-5.07 , -4.29 , and -4.06), and two annotated genes: a WD repeat- and FYVE domain-containing protein 4 (-4.07) and a predicted Zn-dependent hydrolase (-3.97). These findings underscore the importance of these genes in salt stress responses and suggest that the unannotated genes could represent novel components of the organism's stress response mechanisms (Humphrey and Enoch 1998; Zhang et al. 2021; Hernick and Fierke 2005).

The second largest module in *D. hansenii*, blue, was in supertype Group 1 (upregulated in salt). It comprised 823 genes, 367 of which were annotated. The five genes with the highest module membership included an ABC transporter, enoyl-(Acyl carrier protein) reductase, actin-related protein, lysophospholipase, and a phosphate transporter, all of which have roles in stress responses (Jungwirth and Kuchler 2006; Khan et al. 2010; Sun et al. 2023; Barman et al. 2018; Ahmad et al. 2021). The highest salt $\log_2FC$ was an unannotated gene with a value of 3.85, followed by a copper amine oxidase (2.38 $\log_2FC$) and an amidase (2.22 $\log_2FC$).

3.5.2 | Unsupervised Machine Learning

Since clustering was done before annotation, we explored whether unsupervised machine learning could be used to cluster the data, and whether that clustering was similar to the WGCNA result. A straightforward k-means clustering model was implemented (Methods). To determine k , we used WCSS which selected $k = 4$ for all three organisms, since 4 is at the inflection point where fewer groups produce high variance, and more groups do not decrease variance appreciably (Figure 4a). This was confirmed by testing different k values manually ($k = 2, 3, 4, 6, 8, 12, 15$). Because clustering was performed using unsupervised learning, there was no validation set. Instead, comparison against the WGCNA modules was used to measure effectiveness of the clustering. Therefore, in contrast to WGCNA, which found six modules for *Y. lipolytica* and three each for the *Debaryomyces* yeasts, k-means clustering was fixed at four for each. Even with this discrepancy in module number, the gene expression trends within the k-means modules still map to the three supertypes (Figure 4b and Supporting

Information S1: Figure 7). The values plotted in the representative bar charts for each cluster in Figure 4b have normalized average DESeq2 values rather than eigengene correlation values. They clearly show groups that are upregulated in salt (supertype Group 1), downregulated in salt (supertype Group 2), and upregulated in iron starvation (supertype Group 3). The k-means results were then annotated, and a similar comparison between EC number and KOG is presented in Supporting Information S1: Table 6.

The k-means clusters with many genes had very high similarity to the WGCNA results, but similarity decreased for the smaller clusters below 500 DEGs. For example, the *D. subglobosus* k-means C3 (754 DEGs) had a 93% match to the WGCNA blue module (1494 DEGs) and the *D. subglobosus* k-means C2 (692 DEGs) had a 82% match to the WGCNA turquoise module (1734 DEGs). However, for the *D. hansenii* k-means C1 (742 DEGs), there was only a 28% match to the WGCNA brown module (360 DEGs) and the *Y. lipolytica* k-means C0 (901 DEGs) only had a 20% match to the WGCNA brown module (328 DEGs). Given this observation, we ran WGCNA again, but with much larger minimum module sizes (300, 500, and 700, results shown in Supporting Information S1: Tables 7, 8, and 9 and Figures 8, 9, and 10, respectively). With a minimum module size of 300, clustering was similar to 30, except that it removed the smaller modules with unique trends from *Y. lipolytica* and *D. subglobosus*. Increasing to 500 and then 700 did not change modules for *Y. lipolytica* and *D. subglobosus*, but flattened *D. hansenii* to a single module.

We then analyzed each cluster within the supertype groups, similar to WGCNA. *Y. lipolytica* had one cluster, C1 (945 genes), in supertype Group 1 (upregulated in salt). The highest salt DEGs include a glyoxylate/hydroxypyruvate reductase (D-isomer-specific 2-hydroxy acid dehydrogenase superfamily) ($\log_2FC$ 3.10), a glycosylphosphatidylinositol anchor synthesis protein ($\log_2FC$ 3.06), a class II aldolase/adducin N-terminal domain protein ($\log_2FC$ 3.05), a cytochrome b5 ($\log_2FC$ 3.05), an acyl-CoA oxidase ($\log_2FC$ 3.01), an alcohol dehydrogenase, class IV ($\log_2FC$ 3.01), an mRNA splicing factor ($\log_2FC$ 2.99) and unannotated proteins (3.07, 3.10, 3.03, 3.00). This indicates increased activity in metabolic pathways such as glycolysis, the glyoxylate cycle, and fatty acid β -oxidation, which was observed in the metabolic network analysis and known to be part of stress adaptation (Zhang et al. 2020). This also indicates enhanced synthesis of GPI anchors and activity of cytochrome b5 which contribute to maintaining cell membrane structure and function, crucial for protecting the cell against osmotic stress (Kinoshita and Fujita 2015). Additionally, upregulation of enzymes involved in detoxification, such as alcohol dehydrogenase, and those maintaining redox balance, such as cytochrome b5, help the cell manage oxidative stress induced by high salinity (Gostinčar and Gunde-Cimerman 2018). Finally, the increased expression of mRNA splicing factors indicates efficient and accurate expression of stress-responsive genes (Dutertre et al. 2011).

Y. lipolytica had two clusters in supertype Group 2 (downregulated in salt): C0 (901 genes) and C3 (734 genes). C0 consisted of genes downregulated in salt as well as the other stress conditions. Most of the lowest salt DEGs are unannotated, yet have high $\log_2FC$ values ranging between -8.17 and -6.51 . Of the lowest 10 DEGs, only three are annotated as a Fe^{2+}/Zn^{2+}

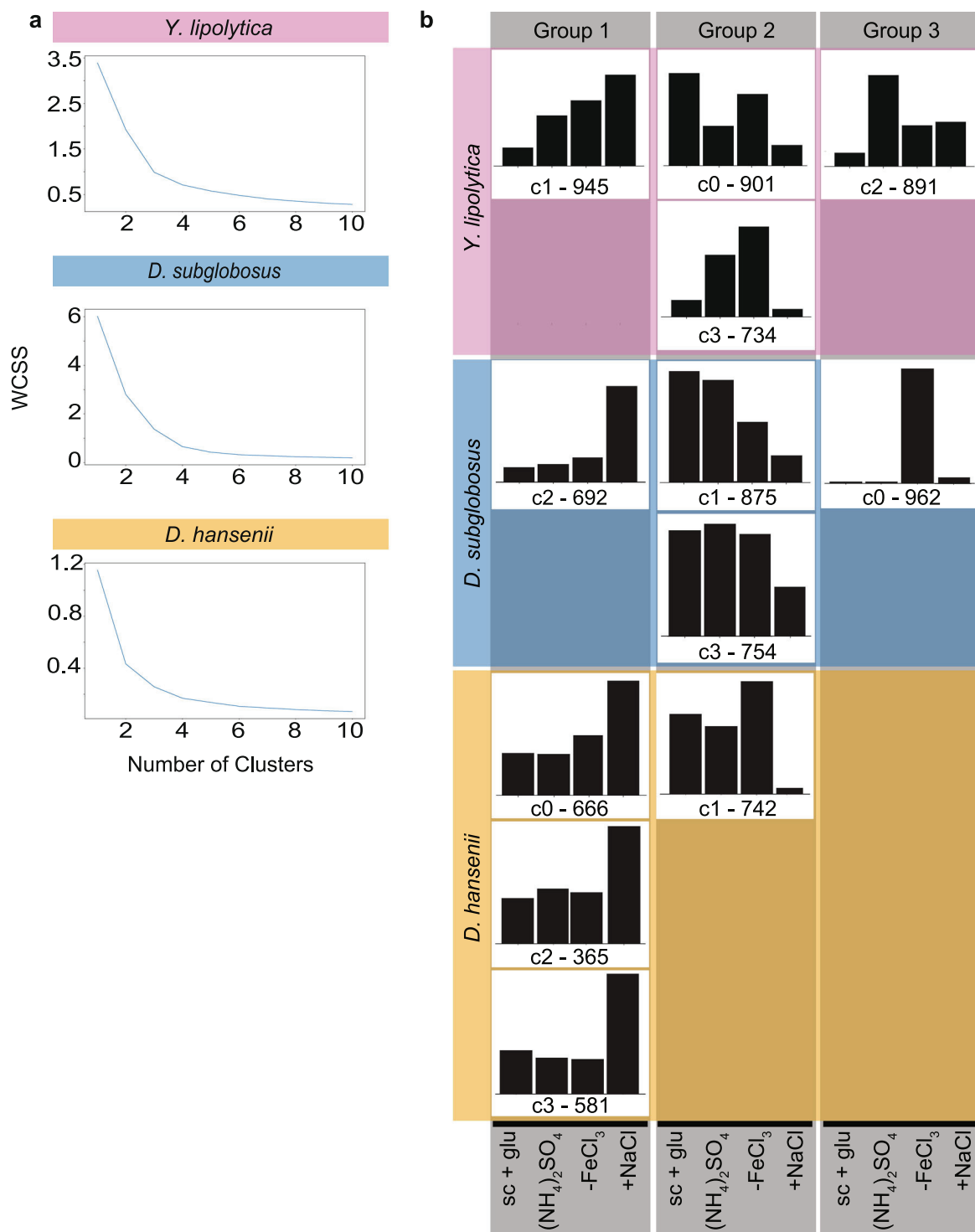


FIGURE 4 | K-means clustering. (a) Finding the optimal k value for each organism by WCSS. For each organism, $k = 4$ was the inflection point that both minimized clusters and WCSS. (b) The clusters produced by k -means, grouped by organism and supertype. The bar charts display the values of module centers, essentially the average gene expression, in each condition. The cluster name along with the number of genes in each cluster is displayed below the bar chart.

regulated transporter (−7.49), molecular chaperone (small heat-shock protein Hsp26/Hsp42) (−6.94), and a heterogeneous nuclear ribonucleoprotein R (RRM superfamily) (−6.32). The significant downregulation of these genes suggests adaptive responses including decreased metal ion uptake to prevent oxidative stress, reduced protein repair capacity leading to potential cellular damage, and impaired RNA metabolism contributing to overall cellular dysfunction (Shcherbik and

Pestov 2019; Morimoto et al. 1997). The $\text{Fe}^{2+}/\text{Zn}^{2+}$ regulated transporter is also the lowest DEG in nitrogen starvation (−10.40). Other annotated DEGs in C0 that were also low in nitrogen starvation included a predicted lipase (−6.95), a monocarboxylate transporter (−6.05), and a manganese superoxide dismutase (−5.90). The significant downregulation of these genes suggests adaptive responses including reduced metal ion uptake to prioritize nitrogen assimilation, decreased

lipid and monocarboxylate metabolism to conserve energy, and lowered capacity to manage oxidative stress, potentially increasing cellular oxidative damage (Baxter et al. 2007; d'Oria et al. 2022; Kaur et al. 2006; Allen et al. 2008). In contrast to C0, cluster C3 in *Y. lipolytica* had downregulated genes in salt relative to control, but upregulated genes in the other stress conditions. This pattern was unique to *Y. lipolytica*. The most downregulated annotated DEGs in C3 are a thyroid hormone receptor-associated protein complex, subunit TRAP230 (−9.34), a permease of the major facilitator superfamily (−7.41), a 60S ribosomal protein L19 (−7.33), a synaptic vesicle transporter SVOP and related transporters (major facilitator superfamily) (−6.44), and a thioredoxin binding protein TBP-2/VDUP1 (−6.41). The reduction in TRAP230 and ribosomal protein L19 suggests a broad suppression of transcriptional and translational activities, conserving energy and resources for essential stress response pathways (Yamasaki and Anderson 2008; Fondell et al. 1999). Downregulating MFS permeases and SVOP-related transporters indicates a strategy to minimize cellular transport activity, which can help maintain osmotic and ionic balances under high salinity (Zhang et al. 2016). Changes in TBP-2/VDUP1 expression reflect an adjustment of the cellular redox state, crucial for managing oxidative stress induced by salt (Nishinaka et al. 2001).

Y. lipolytica had one cluster in supertype Group 3, cluster C2 (891 genes). This cluster had upregulated gene expression across all stress conditions, but particularly in nitrogen starvation. Of the 10 most upregulated nitrogen starvation DEGs, only two were annotated. These included proteins containing BTB/POZ and Kelch domains, involved in regulatory/signal transduction processes (T'Jampens et al. 2002) (2.20), and a WD repeat protein (Humphrey and Enoch 1998) (2.11). These are both regulatory proteins (Nacak et al. 2006; LiD 2001).

D. subglobosus had one cluster in supertype Group 1 (upregulated in salt), C2 (692 genes). Only 4 of the top 15 DEGs were annotated: Cu²⁺/Zn²⁺ superoxide dismutase SOD1 (5.22 and 4.87), 60S ribosomal protein L29 (4.87), and a translocase of outer mitochondrial membrane complex, subunit TOM7 (4.41). The upregulation of SOD1 indicates ROS detoxification (Hasanuzzaman et al. 2021). TOM7 and L29 are involved in the import of proteins into mitochondria (Johnston et al. 2002) and facilitating enhanced protein synthesis (Contreras-Martinez et al. 2012), respectively.

D. subglobosus had two clusters in supertype Group 2 (downregulated in salt), C1 (875 genes) and C3 (754 genes). In C1, the transcription factor NSD1 and related SET domain protein were the lowest DEGs for salt (−3.35), with the next lowest annotated DEG being a predicted transporter (−2.91). NSD1 and related SET domain proteins were also the most downregulated in nitrogen and iron starvation (−4.08 and −6.52, respectively). In C3, similarly, all conditions share the most downregulated annotated DEG which is a biotin synthase (−4.78 in salt stress, −4.30 in nitrogen starvation, and −4.14 in iron starvation). Downregulating biotin synthase reflects a shift in metabolic priorities, where the cell focuses on survival and stress adaptation rather than growth and proliferation (Lazar et al. 2017). One other interesting finding in this cluster is in iron starvation, of the highest 15 upregulated DEGs, 13 of them are transporters, with log₂FC values ranging from 7.98 to 3.95,

highlighting the critical role of transporters in the cellular response to iron deficiency (Yang, Lin, and Schmidt 2010).

The fourth *D. subglobosus* cluster, C0 (962 genes), was in supertype Group 3 (upregulated in iron starvation). Once again, many of the most highly DEGs were unannotated, and the sequences matched the unannotated genes in the WGCNA brown module. The top five annotated DEGs were transporters, specifically an ammonia permease (8.88), copper amine oxidase (6.26), amino acid transporters (3.29 and 2.04), and a predicted transporter (2.18).

D. hansenii had three clusters in supertype Group 1 (upregulated in salt): C0 (666 genes), C2 (365 genes), and C3 (581 genes). These clusters were all similar, with minor variations in differential gene expression in the nitrogen starvation and iron starvation conditions. As shown in the WCSS chart, although $k = 4$ minimized variance for the *D. hansenii* datasets, the variance for $k = 2$ and $k = 3$ was also quite small, therefore it is possible that all three clusters could be combined. In C0, only L-carnitine dehydratase/alpha-methylacyl-CoA racemase (log₂FC value of 3.28) was annotated among the top five DEGs under salt stress. This enzyme plays a vital role in fatty acid metabolism, supporting energy production and membrane integrity under high salt conditions (Kong et al. 2020). In C2, the highest salt DEGs include the transcription factor TMF (5.70), which regulates stress-responsive gene expression (Slabaugh and Brandizzi 2011), the CCAAT-binding factor (5.42), which modulates gene expression related to stress adaptation (Chakravarti et al. 2017), and isocitrate lyase (5.26), an enzyme involved in the glyoxylate cycle for energy production (Honer Zu Bentrop et al. 1999). Nitrogen starvation slightly increased average expression in C2. Nitrogen starvation's highest annotated DEGs include a ferric reductase (1.67) that facilitates iron uptake by converting ferric iron (Fe³⁺) to ferrous iron (Fe²⁺), crucial for managing nitrogen metabolism (Schmidt 1999). In C3, the highest salt DEGs include the actin-binding protein SLA2 (2.64), which aids in cytoskeleton organization and cell morphogenesis (Holtzman, Yang, and Drubin 1993), flavonol reductase (2.53), involved in flavonoid biosynthesis for stress mitigation (Pollastri and Tattini 2011), and aldehyde dehydrogenase (2.48), which detoxifies aldehydes generated under salt stress (Dong et al. 2017).

The fourth *D. hansenii* cluster, C1 (742 genes), was in supertype Group 2 (downregulated in salt). The lowest annotated salt DEGs include the mitochondrial import inner membrane translocase (−4.15), cytochrome c (−4.09), and ubiquinone oxidoreductase (−4.01), all involved in mitochondrial function and energy production (Meyer, Bömer, and Pratej 1997). In contrast, under high iron stress, cluster C1 has several upregulated transporters, including transporters of urea (2.8), amino acids (2.67, 2.05) and ammonium (1.70), which enhance nitrogen assimilation and metabolism (Shakoury-Elizeh et al. 2004), reflecting an adaptive response to iron starvation availability.

3.5.3 | Analysis Across Supertypes Uncovers Transporters, Conserved Genes, and Unknown Genes

Two of the three cluster supertypes (Group 1 and Group 2) displayed a unique response to salt, which corroborates our initial PCA analysis showing that the salt stress condition contributed the

majority of the variance in the datasets. Across all three organisms and clusters, Group 1 (upregulated in salt) contained 4,082 genes, 2730 of which were annotated and Group 2 (downregulated in salt) contained 4793 genes, 2967 of which were annotated. Many of these are metabolic genes, yet other common groups were ribosomal proteins, transcription factors, genes involved in glycerolipid biosynthesis, and nearly 10% were transporters. The novel transporter B5RUG0, identified in *D. hansenii* by Navarrete et al. (Navarrete et al. 2022), was found in this set. Expression in *D. hansenii* increased 1.2 log₂FC under salt stress. Using BLAST, a similar gene jgi.plDebSub1149526 was found in *D. subglobosus*, but it was not significantly regulated. *Y. lipolytica* did not have a homologous gene. However, one hit was found in each organism when “cation_efflux,” a family and domain search term found on InterPro for B5RUG0, was used as a search term on the annotated genomes. The *D. hansenii* hit matched B5RUG0, the *D. subglobosus* hit was upregulated 1.5 log₂FC in salt stress, and the *Y. lipolytica* hit had −1.7 log₂FC in salt stress. It is interesting to note the differential regulation of putative cation transporters related to B5RUG0 between the extremely salt tolerant *Debaryomyces* yeasts and *Y. lipolytica*. It could be that simply changing the gene regulatory patterns in *Y. lipolytica* could lead to higher salt tolerance in this organism. The specific annotations can be found in Supporting Information S1: Tables 10–12 or in the published data files on GitHub (<https://github.com/emyounglab/rnaseq-onboarding>).

Then, a search over all supertypes found that there were 46 DEGs in both *Debaryomyces* species, but not in *Yarrowia* (Table 1). Of these 46, 21 were poorly annotated or unannotated in both organisms, and 4 were only identified in one organism. These conserved genes unique to these organisms may contribute to the phenotypic similarity observed between the *Debaryomyces* yeasts, but not *Yarrowia*, particularly the specific salt response.

Finally, since so many unannotated genes were differentially expressed across all supertype groups, we used AlphaFold (Jumper et al. 2021) to predict possible functions of the highest DEGs. In *Y. lipolytica* cluster C0 (downregulated in salt), the lowest five unannotated genes were predicted to be an atrophin (−8.46, a transcriptional corepressor), a pirin protein (−8.17, a redox-sensitive modulator of gene expression), a cobalamin synthesis protein (−7.86, vitamin B12 biosynthesis), a phosphatidylinositol transfer protein (−5.75, involved in lipid trafficking), and the fifth (−5.88) had no prediction. In *D. subglobosus* cluster C0 (upregulated in iron starvation), the three highest unannotated genes were predicted to be a carbon-nitrogen hydrolase (11.20, involved in catabolic processes), a 2-hydroxy-palmitic acid diosygenase mpol (11.09, involved in iron-dependent fatty acid degradation), and an ino 80 subunit (10.99, a chromatin remodeling protein). While currently time and computationally expensive, these AlphaFold results point to the potential of machine learning to not only cluster transcriptomics data but infer the function of unknown transcripts as well.

4 | Discussion

This study compared the phenotypes and transcriptomes of three oleaginous yeasts, *Y. lipolytica*, *D. hansenii*, and *D. subglobosus*, under three stress conditions, NaCl stress, (NH₄)₂SO₄

starvation, and Fe(Cl₃) starvation. We showed that the *Debaryomyces* yeasts extensively and exclusively rewired their transcriptomes when exposed to salt, while *Y. lipolytica* Po1f responded uniquely to each condition tested.

Phenotyping showed that *Debaryomyces* lipid and riboflavin production declined under salt stress. Custom metabolic network analysis based on KOG annotations showed that the lipid decrease was due to increased degradation of glycerolipids to glycerol and free fats. This is expected, since glycerol is implicated as a key osmoprotectant. Furthermore, the metabolic network analysis showed that the riboflavin decrease was not due to downregulation of riboflavin biosynthesis. Rather, the downstream pathway was upregulated, appearing to draw riboflavin to flavin cofactor. This could be because additional redox equivalents are needed to power the increased production of glycerol (Becker, Zhu, and Moxley 2011). Thus, this study implicates flavin cofactors along with glycerol production as a key mechanism that *Debaryomyces* yeasts use to tolerate salt. *Yarrowia*, without this mechanism, appears to generate redox equivalents through increased glycolysis. This is further confirmed by the clustering analysis, which showed many redox-active and glycerol-active genes in the supertype Group 1 (upregulated in salt).

Clustering implicated many genes in the stress response of the three yeasts, many of which were annotated with KOG. The strongest responses observed were in the salt stress condition. The DEGs in the salt-responsive clusters were involved in many cell processes, including metabolism, transport, membrane composition and trafficking, redox, and gene expression. Taken together, these point to a concerted salt stress response where not only glycerol biosynthesis is upregulated, but the cell membrane composition is completely changed, both in transporter expression and lipid composition. To coordinate this effort, many transcription factors and chromatin remodelers are differentially expressed under salt stress. This analysis yields many more gene targets in addition to the transporter discovered in *D. hansenii* (Navarrete et al. 2022) that could be used to engineer salt tolerance in many yeasts. Furthermore, because clustering was performed on the entire data set, unannotated genes were significant DEGs in all clusters. Feeding the transcript sequences for a subset of these unannotated genes into AlphaFold allowed prediction of function for some, showing that improving annotation with machine learning could discover even more mechanisms involved in stress tolerance.

These insights were not possible without both a “network-first” and “cluster-first” approach to analyze the RNA-Seq datasets. The network-first analysis was essential in highlighting metabolic responses, such as the observed increase in *Y. lipolytica* Po1f lipid biosynthesis in response to nitrogen starvation and the novel observation that flavinogenesis not repressed under salt stress in *Debaryomyces*, rather upregulation of flavin cofactor production genes draw down the riboflavin pools. The cluster-first analysis was essential for elucidating the involvement of transporters, cell membranes, and transcription factors in the stress response in addition to the metabolic genes. Furthermore, cluster-first analysis was essential for determining the most important unannotated genes. Thus, we propose both

TABLE 1 | Genes of interest uniquely found in *Debaryomyces* and not *Yarrowia lipolytica*.

<i>D. subglobosus</i>	KOG	<i>D. hansenii</i>
jgi.plDebsub1l210935	Porin/voltage-dependent anion-selective channel protein	jgi.plDebha1l2840
jgi.plDebsub1l212290	Mitochondrial/chloroplast ribosomal protein L2	jgi.plDebha1l3198
jgi.plDebsub1l213021	Copper chaperone	jgi.plDebha1l3906
jgi.plDebsub1l25534	WASP-interacting protein VRP1/WIP, contains WH2 domain	jgi.plDebha1l1786
jgi.plDebsub1l4916	Mitochondrial/chloroplast ribosomal protein S14/S29	jgi.plDebha1l2727
jgi.plDebsub1l183764	Mitochondrial/chloroplast ribosomal protein S12	jgi.plDebha1l1714
jgi.plDebsub1l209950	Ubiquinone biosynthesis methyltransferase COQ5	jgi.plDebha1l2594
jgi.plDebsub1l210147	Peroxisomal biogenesis protein peroxin	jgi.plDebha1l1922
jgi.plDebsub1l210748	Transcription factor MBF1	jgi.plDebha1l669
jgi.plDebsub1l210873	Mitochondrial import inner membrane translocase, subunits TIM10/TIM12	jgi.plDebha1l2893
jgi.plDebsub1l212276	60s ribosomal protein L23	jgi.plDebha1l3182
jgi.plDebsub1l212517	Mitochondrial/chloroplast ribosomal protein S19	jgi.plDebha1l3420
jgi.plDebsub1l214388	60S ribosomal protein L14/L17/L23	jgi.plDebha1l1069
jgi.plDebsub1l215345	Mitochondrial/chloroplast ribosomal protein L33-like	jgi.plDebha1l2205
jgi.plDebsub1l215533	DNA-directed RNA polymerase, subunit RPB7.0	jgi.plDebha1l2328
jgi.plDebsub1l217249	Glycolate oxidase	jgi.plDebha1l1414
jgi.plDebsub1l217687	Chitinase	jgi.plDebha1l4734
jgi.plDebsub1l24442	Sof1-like rRNA processing protein (contains WD40 repeats)	jgi.plDebha1l1880
jgi.plDebsub1l28804	Predicted E3 ubiquitin ligase	jgi.plDebha1l1561

network and clustering analysis is necessary for analyzing the transcriptomes of poorly annotated species.

Our cluster-first analysis used unsupervised k-means clustering and weighted gene co-expression network analysis (WGCNA). The unsupervised k-means clustering performed similarly to WGCNA – both produced similar clusters and highlighted transcription factors and transporters as important parts of the observed rewiring, particularly to salt stress. This means that machine learning algorithms that cluster effectively, are simple to use, and do not need genome annotation, can be useful tools for transcriptomic analysis. However, while the large clusters found by k-means were very similar to the WGCNA clusters, k-means did not find the smaller clusters (below 500 genes) that WGCNA did. Thus, while unsupervised machine learning captured the most prevalent gene expression trends, the flexibility of WGCNA to alter the number of clusters for the best fit may be useful in elucidating more complex gene expression patterns in other applications. However, with the majority of genes being regulated by salt across all conditions, both methods produced very similar results.

In summary, our analysis provides a new foundation for investigating the stress response of nonconventional yeasts and implicates many genes and cellular processes involved in tolerance. This was only possible through a combined network analysis and clustering approach, leveraging machine learning, that provided both a comprehensive metabolic overview as well as annotation-agnostic gene groups for gene discovery and correlating genotype to phenotype. Therefore, this combined approach is effective for comparing multiple organisms under multiple conditions to gain further insight into adaptations and

identify candidate genes that can confer desirable biological traits.

Author Contributions

Sarah J. Weintraub, Joseph H. Collins, and Eric M. Young contributed to the conceptualization of the study. Zekun Li conducted fermentation experiments. Joseph H. Collins performed genome sequencing and preparation of RNA samples. Sarah J. Weintraub and Carter L. Nakagawa analyzed the data. Sarah J. Weintraub, Zekun Li, and Eric M. Young contributed to manuscript preparation.

4.1 | Acknowledgments

The authors would like to acknowledge the contributions of Anna Lipzen, Stephen Mondo, and Igor Grigoriev for technical services for transcriptomics at the DOE Joint Genome Institute. The transcriptomics work (proposal 10.46936/10.25585/60001246) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. The Department of Energy operated under Contract No. DE-AC02-05CH11231. Prymetime was based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) under Finding Engineering Linked Indicators (FELIX) program contract N66001-18-C-4507. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. Phenotyping by Z.L. was supported by a National Science Foundation CAREER Award (1944046). S.J.W. was supported by a National Science Foundation National Research Training (NRT) grant, CEDAR (NRT-HDR-

2021871). Transcriptomic sample preparation was supported by WPI startup funds.

Data Availability Statement

RNA-Seq reads are available on PRJNA1028627 NCBI BioProject, and the Joint Genome Institute. Genome sequences can be found on NCBI with the following accession IDs: *Y. lipolytica* PO1f - PRJNA1047134, *D. subglobosus* dep8 - PRJNA1047133. All data is available on GitHub (<https://github.com/emyounglab/rnaseq-onboarding>).

The data that support the findings of this study are openly available in NCBI at <https://submit.ncbi.nlm.nih.gov/>.

References

- Abbas, C. A., and A. A. Sibirny. 2011. "Genetic Control of Biosynthesis and Transport of Riboflavin and Flavin Nucleotides and Construction of Robust Biotechnological Producers." *Microbiology and Molecular Biology Reviews* 75: 321–360.
- Abdel-Mawgoud, A. M., K. A. Markham, C. M. Palmer, N. Liu, G. Stephanopoulos, and H. S. Alper. 2018. "Metabolic Engineering in the Host *Yarrowia lipolytica*." *Metabolic Engineering* 50: 192–208.
- Abghari, A., and S. Chen. 2014. "*Yarrowia lipolytica* as an Oleaginous Cell Factory Platform for Production of Fatty Acid-Based Biofuel and Bioproducts." *Frontiers in Energy Research* 2: 21.
- Adrio, J. L. 2017. "Oleaginous Yeasts: Promising Platforms for the Production of Oleochemicals and Biofuels." *Biotechnology and Bioengineering* 114: 1915–1920.
- Ageitos, J. M., J. A. Vallejo, P. Veiga-Crespo, and T. G. Villa. 2011. "Oily Yeasts as Oleaginous Cell Factories." *Applied Microbiology and Biotechnology* 90: 1219–1227.
- Agostinho, S. P., M. A. Branco, D. E. S. Nogueira, et al. 2024. "Unsupervised Analysis of Whole Transcriptome Data From Human Pluripotent Stem Cells Cardiac Differentiation." *Scientific Reports* 14: 3110.
- Ahmad, F., and A. H. Rose. 1962. "The Role of Biotin in the Regulation of Enzyme Synthesis in Yeast." *Archives of Biochemistry and Biophysics* 97: 302–308.
- Ahmad, I., A. Rawoof, K. Islam, J. Momo, and N. Ramchiary. 2021. "Identification and Expression Analysis of Phosphate Transporter Genes and Metabolites in Response to Phosphate Stress in *Capsicum annum*." *Environmental and Experimental Botany* 190: 104597.
- Albacar, M., A. Zekhnini, J. Pérez-Valle, J. L. Martínez, A. Casamayor, and J. Ariño. 2023. "Transcriptomic Profiling of the Yeast *Komagataella phaffii* in Response to Environmental Alkalinization." *Microbial Cell Factories* 22: 63.
- Allen, A. E., J. LaRoche, U. Maheswari, et al. 2008. "Whole-Cell Response of the Pennate Diatom *Phaeodactylum tricornutum* to Iron Starvation." *Proceedings of the National Academy of Sciences* 105: 10438–10443.
- Ashburner, M., C. A. Ball, J. A. Blake, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25: 25–29.
- Attfield, P. V. 1997. "Stress Tolerance: The Key to Effective Strains of Industrial Baker's Yeast." *Nature Biotechnology* 15: 1351–1357.
- Balabanova, L., L. Averianova, M. Marchenok, O. Son, and L. Tekutyeva. 2021. "Microbial and Genetic Resources for Cobalamin (Vitamin B12) Biosynthesis: From Ecosystems to Industrial Biotechnology." *International Journal of Molecular Sciences* 22: 4522.
- Bankar, A. V., A. R. Kumar, and S. S. Zinjarde. 2009. "Environmental and Industrial Applications of *Yarrowia lipolytica*." *Applied Microbiology and Biotechnology* 84: 847–865.
- Bao, W., Z. Li, X. Wang, et al. 2021. "Approaches to Improve the Lipid Synthesis of Oleaginous Yeast *Yarrowia lipolytica*: A Review." *Renewable and Sustainable Energy Reviews* 149: 111386.
- Barman, A., D. Gohain, U. Bora, and R. Tamuli. 2018. "Phospholipases Play Multiple Cellular Roles Including Growth, Stress Tolerance, Sexual Development, and Virulence in Fungi." *Microbiological Research* 209: 55–69.
- Baxter, C. J., H. Redestig, N. Schauer, et al. 2007. "The Metabolic Response of Heterotrophic Arabidopsis Cells to Oxidative Stress." *Plant Physiology* 143: 312–325.
- Becker, D. F., W. Zhu, and M. A. Moxley. 2011. "Flavin Redox Switching of Protein Functions." *Antioxidants & Redox Signaling* 14: 1079–1091.
- Bellou, S., I.-E. Triantaphyllidou, P. Mizerakis, and G. Aggelis. 2016. "High Lipid Accumulation in *Yarrowia lipolytica* Cultivated Under Double Limitation of Nitrogen and Magnesium." *Journal of Biotechnology* 234: 116–126.
- Bianchi, F., J. S. van't Klooster, S. J. Ruiz, and B. Poolman. 2019. "Regulation of Amino Acid Transport in *Saccharomyces cerevisiae*." *Microbiology and Molecular Biology Reviews* 83: 10–1128.
- Blais, E. M., A. K. Chavali, and J. A. Papin. 2013. "Linking Genome-Scale Metabolic Modeling and Genome Annotation." *Systems Metabolic Engineering: Methods in Molecular Biology* 985: 61–83.
- Blazcek, J., A. Hill, L. Liu, et al. 2014. "Harnessing *Yarrowia lipolytica* Lipogenesis to Create a Platform for Lipid and Biofuel Production." *Nature Communications* 5: 3131.
- Bremer, E., and R. Krämer. 2019. "Responses of Microorganisms to Osmotic Stress." *Annual Review of Microbiology* 73: 313–334.
- Brion, C., D. Pflieger, S. Souali-Crespo, A. Friedrich, and J. Schacherer. 2016. "Differences in Environmental Stress Response Among Yeasts Is Consistent With Species-Specific Lifestyles." *Molecular Biology of the Cell* 27: 1694–1705.
- Bushnell, B., J. Rood, and E. Singer. 2017. "Bbmerge-Accurate Paired Shotgun Read Merging Via Overlap." *PLoS One* 12: e0185056.
- Cao, X., Y.-B. Lv, J. Chen, T. Imanaka, L.-J. Wei, and Q. Hua. 2016. "Metabolic Engineering of Oleaginous Yeast *Yarrowia lipolytica* for Limonene Overproduction." *Biotechnology for Biofuels* 9: 1–11.
- Capece, A., and P. Romano. 2009. "Pecorino Di Filiano' Cheese as a Selective Habitat for the Yeast Species, *Debaryomyces hansenii*." *International Journal of Food Microbiology* 132: 180–184.
- Chakravarti, A., K. Camp, D. S. McNabb, and I. Pinto. 2017. "The Regulation of the *Candida albicans* Oxidative Stress Response by the ccaat-Binding Factor." *PLoS One* 12: e0170649.
- Chen, J., L. Zhang, Y. Liu, et al. 2024. "RNA-Seq-Based WGCNA and Association Analysis Reveal the Key Regulatory Module and Genes Responding to Salt Stress in Wheat Roots." *Plants* 13: 274.
- Chen, X., Y. He, L. Liu, X. Zhu, B. Sen, and G. Wang. 2022. "Nitrogen Starvation Enhances the Production of Saturated and Unsaturated Fatty Acids in *Aurantiochytrium* Sp. PKU# SW8 by Regulating Key Biosynthetic Genes." *Marine Drugs* 20: 621.
- Chen, X., Z. Li, X. Zhang, F. Hu, D. D. Ryu, and J. Bao. 2009. "Screening of Oleaginous Yeast Strains Tolerant to Lignocellulose Degradation Compounds." *Applied Biochemistry and Biotechnology* 159: 591–604.
- Christensen, J., K. Agger, P. A. Cloos, et al. 2007. "Rbp2 Belongs to a Family of Demethylases, Specific for Tri-And Dimethylated Lysine 4 on Histone 3." *Cell* 128: 1063–1076.
- Collins, J. H., K. W. Keating, T. R. Jones, et al. 2021. "Engineered Yeast Genomes Accurately Assembled From Pure and Mixed Samples." *Nature Communications* 12: 1–15.
- Contreras-Martinez, L. M., J. T. Boock, J. S. Kostecki, and M. P. DeLisa. 2012. "The Ribosomal Exit Tunnel as a Target for Optimizing Protein Expression in *Escherichia coli*." *Biotechnology Journal* 7: 354–360.
- Conway, J. R., A. Lex, and N. Gehlenborg. 2017. "UpSet: An R Package for the Visualization of Intersecting Sets and Their Properties."

- Bioinformatics 33: 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- Cooper, T. G., and R. Sumrada. 1975. "Urea Transport in *Saccharomyces cerevisiae*." *Journal of Bacteriology* 121: 571–576.
- Csonka, L. N. 1989. "Physiological and Genetic Responses of Bacteria to Osmotic Stress." *Microbiological Reviews* 53: 121–147.
- Cyert, M. S., and C. C. Philpott. 2013. "Regulation of Cation Balance in *Saccharomyces cerevisiae*." *Genetics* 193: 677–713.
- Defosse, T., Y. Le Govic, V. Courdavault, et al. 2018. "Yeasts From the Ctg Clade (Candida Clade): Biology, Impact in Human Health, and Biotechnological Applications." *Journal de Mycologie Medicale* 28: 257–268.
- Demain, A. L. 1972. "Riboflavin Oversynthesis." *Annual Reviews in Microbiology* 26: 369–388.
- Deparis, Q., A. Claes, M. R. Foulquié-Moreno, and J. M. Thevelein. 2017. "Engineering Tolerance to Industrially Relevant Stress Factors in Yeast Cell Factories." *FEMS yeast research* 17: fox036.
- D'haeseleer, P. 2005. "How Does Gene Expression Clustering Work?" *Nature Biotechnology* 23: 1499–1501.
- Dmytruk, K. V., and A. A. Sibirny. 2012. "*Candida famata* (*Candida flareri*)." *Yeast* 29: 453–458.
- Dmytruk, K. V., A. Y. Voronovsky, and A. A. Sibirny. 2006. "Insertion Mutagenesis of the Yeast *Candida famata* (*Debaryomyces hansenii*) by Random Integration of Linear DNA Fragments." *Current Genetics* 50: 183–191.
- Dong, Y., H. Liu, Y. Zhang, et al. 2017. "Comparative Genomic Study of aldH Gene Superfamily in Gossypium: A Focus on *Gossypium hirsutum* Under Salt Stress." *PLoS One* 12: e0176733.
- d'Oria, A., G. Courbet, B. Billiot, et al. 2022. "Drought Specifically Downregulates Mineral Nutrition: Plant Ionic Content and Associated Gene Expression." *Plant Direct* 6: e402.
- Dougherty, K. M., D. A. Swanson, L. C. Brody, and D. Valle. 1993. "Expression and Processing of Human Ornithine- δ -Aminotransferase in *Saccharomyces cerevisiae*." *Human Molecular Genetics* 2: 1835–1840.
- Dutertre, M., G. Sanchez, J. Barbier, L. Corcos, and D. Auboeuf. 2011. "The Emerging Role of Pre-Messenger RNA Splicing in Stress Responses: Sending Alternative Messages and Silent Messengers." *RNA Biology* 8: 740–747.
- Epstein, C. B., and F. R. Cross. 1992. "Clb5: A Novel B Cyclin From Budding Yeast With a Role in S Phase." *Genes & Development* 6: 1695–1706.
- Evans, C. T., and C. Ratledge. 1984. "Effect of Nitrogen Source on Lipid Accumulation in Oleaginous Yeasts." *Microbiology* 130: 1693–1704.
- Fang, Y., Y. Tang, Y. Zhang, et al. 2021. "The h3k36me2 Methyltransferase nsd1 Modulates h3k27ac at Active Enhancers to Safeguard Gene Expression." *Nucleic Acids Research* 49: 6281–6295.
- Farahbod, M., and P. Pavlidis. 2020. "Untangling the Effects of Cellular Composition on Coexpression Analysis." *Genome Research* 30: 849–859.
- Fedorovich, D., O. Protchenko, and E. Lesuisse. 1999. "Iron Uptake by the Yeast *Pichia guilliermondii*. Flavinogenesis and Reductive Iron Assimilation Are Co-Regulated Processes." *Biomaterials* 12: 295–300.
- Fondell, J. D., M. Guermah, S. Malik, and R. G. Roeder. 1999. "Thyroid Hormone Receptor-Associated Proteins and General Positive Cofactors Mediate Thyroid Hormone Receptor Function in the Absence of the Tata Box-Binding Protein-Associated Factors of tfiid." *Proceedings of the National Academy of Sciences* 96: 1959–1964.
- Fujii, J., T. Myint, A. Okado, H. Kaneto, and N. Taniguchi. 1996. "Oxidative Stress Caused by Glycation of Cu, Zn-Superoxide Dismutase and Its Effects on Intracellular Components." *Nephrology Dialysis Transplantation* 11: 34–40.
- Gostinčar, C., and N. Gunde-Cimerman. 2018. "Overview of Oxidative Stress Response Genes in Selected Halophilic Fungi." *Genes* 9: 143.
- Grenson, M. 1992. "Amino Acid Transporters in Yeast: Structure, Function and Regulation." In J. J. H. M. De Pont, (Ed.), *New Comprehensive Biochemistry*, vol. 21, 219–245. Elsevier.
- Hartman, T., K. Stead, D. Koshland, and V. Guacci. 2000. "Pds5p Is an Essential Chromosomal Protein Required for Both Sister Chromatid Cohesion and Condensation in *Saccharomyces cerevisiae*." *Journal of Cell Biology* 151: 613.
- Hasanuzzaman, M., M. R. H. Raihan, A. A. C. Masud, et al. 2021. "Regulation of Reactive Oxygen Species and Antioxidant Defense in Plants Under Salinity." *International Journal of Molecular Sciences* 22: 9326.
- Heefner, D. L., C. A. Weaver, M. J. Yarus, and L. A. Burdzinski. 1992. Method for Producing Riboflavin With *Candida famata*. US Patent 5164303.
- Hernick, M., and C. A. Fierke. 2005. "Zinc Hydrolases: The Mechanisms of Zinc-Dependent Deacetylases." *Archives of Biochemistry and Biophysics* 433: 71–84.
- Herrero, E., J. Ros, G. Belli, and E. Cabisco. 2008. "Redox Control and Oxidative Stress in Yeast Cells." *Biochimica et Biophysica Acta (BBA)-General Subjects* 1780: 1217–1235.
- Heyer, L. J., S. Kruglyak, and S. Yooseph. 1999. "Exploring Expression Data: Identification and Analysis of Coexpressed Genes." *Genome Research* 9: 1106–1115.
- Hobot, J. A., and D. H. Jennings. 1981. "Growth of *Debaryomyces hansenii* and *Saccharomyces cerevisiae* in Relation to pH and Salinity." *Experimental Mycology* 5: 217–228.
- Hoffmann, T., and E. Bremer. 2016. "Management of Osmotic Stress by *Bacillus subtilis*: Genetics and Physiology." In F. J. de Bruijn (Ed.), *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*: 657–676. <https://doi.org/10.1002/9781119004813.ch63>.
- Hohmann, S. 2002. "Osmotic Stress Signaling and Osmoadaptation in Yeasts." *Microbiology and Molecular Biology Reviews* 66: 300–372.
- Holtzman, D. A., S. Yang, and D. G. Drubin. 1993. "Synthetic-Lethal Interactions Identify Two Novel Genes, sla1 and sla2, That Control Membrane Cytoskeleton Assembly in *Saccharomyces cerevisiae*." *Journal of Cell Biology* 122: 635–644.
- Honer Zu Bentrup, K., A. Miczak, D. L. Swenson, and D. G. Russell. 1999. "Characterization of Activity and Expression of Isocitrate Lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*." *Journal of Bacteriology* 181: 7161–7167.
- Huerta-Cepas, J., D. Szklarczyk, D. Heller, et al. 2019. "EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses." *Nucleic Acids Research* 47: D309–D314.
- Humphrey, T., and T. Enoch. 1998. "Sum1, a Highly Conserved wd-Repeat Protein, Suppresses sm Checkpoint Mutants and Inhibits the Osmotic Stress Cell Cycle Response in Fission Yeast." *Genetics* 148: 1731–1742.
- Hunter, J. D. 2007. "Matplotlib: A 2d Graphics Environment." *Computing in Science & Engineering* 9: 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Jagtap, S. S., A. A. Bedekar, V. Singh, Y.-S. Jin, and C. V. Rao. 2021. "Metabolic Engineering of the Oleaginous Yeast *Yarrowia lipolytica* polf for Production of Erythritol From Glycerol." *Biotechnology for Biofuels* 14: 1–17.
- Johnston, A. J., J. Hoogenraad, D. A. Dougan, et al. 2002. "Insertion and Assembly of Human tom7 into the Preprotein Translocase Complex of the Outer Mitochondrial Membrane." *Journal of Biological Chemistry* 277: 42197–42204.

- Jumper, J., R. Evans, A. Pritzel, et al. 2021. "Highly Accurate Protein Structure Prediction With AlphaFold." *Nature* 596: 583–589.
- Jungwirth, H., and K. Kuchler. 2006. "Yeast abc Transporters – A Tale of Sex, Stress, Drugs and Aging." *FEBS Letters* 580: 1131–1138.
- Kanehisa, M., and S. Goto. 2000. "Kegg: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28: 27–30.
- Kanipes, M. I., and S. A. Henry. 1997. "The Phospholipid Methyltransferases in Yeast." *Biochimica et Biophysica Acta (BBA)–Lipids and Lipid Metabolism* 1348: 134–141.
- Karpov, D. S., E. N. Grineva, A. T. Leinsoo, et al. 2017. "Functional Analysis of *Debaryomyces hansenii* rpn4 on a Genetic Background of *Saccharomyces cerevisiae*." *FEMS Yeast Research* 17: fow098.
- Kaur, A., M. Pan, M. Meislin, M. T. Facciotti, R. El-Gewely, and N. S. Baliga. 2006. "A Systems View of Haloarchaeal Strategies to Withstand Stress From Transition Metals." *Genome Research* 16: 841–854.
- Khan, S., S. N. Nagarajan, A. Parikh, et al. 2010. "Phosphorylation of Enoyl-Acyl Carrier Protein Reductase Inha Impacts Mycobacterial Growth and Survival." *Journal of Biological Chemistry* 285: 37860–37871.
- Kim, D., B. Langmead, and S. L. Salzberg. 2015. "Hisat: A Fast Spliced Aligner With Low Memory Requirements." *Nature Methods* 12: 357–360.
- Kinoshita, N., J. Minshull, and M. W. Kirschner. 1995. "The Identification of Two Novel Ligands of the fgf Receptor by a Yeast Screening Method and Their Activity in Xenopus Development." *Cell* 83: 621–630.
- Kinoshita, T., and M. Fujita. 2015. "Biosynthesis of gpi-Anchored Proteins: Special Emphasis on gpi Lipid Remodeling." *Journal of lipid research* 57: 6–24.
- Knight, S. A. B., E. Lesuisse, R. Stearman, R. D. Klausner, and A. Dancis. 2002. "Reductive Iron Uptake by *Candida albicans*: Role of Copper, Iron and the tup1 Regulator." *Microbiology* 148: 29–40.
- Kong, G., H. Lee, Q. Tran, et al. 2020. "Current Knowledge on the Function of α -Methyl Acyl-CoA Racemase in Human Diseases." *Frontiers in Molecular Biosciences* 7: 153.
- Kuo, A., B. Bushnell, and I. V. Grigoriev. 2014. "Fungal Genomics: Sequencing and Annotation." In Francis M. Martin (Ed.), *Advances in Botanical Research*, vol. 70, 1–52. Elsevier.
- Labella, A. L., D. A. Oplente, J. L. Steenwyk, C. T. Hittinger, and A. Rokas. 2019. "Variation and Selection on Codon Usage Bias Across an Entire Subphylum." *PLoS Genetics* 15: e1008304.
- Langfelder, P., and S. Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9: 559.
- Lazar, N., A. Fay, M. Nandakumar, et al. 2017. "Control of Biotin Biosynthesis in Mycobacteria by a Pyruvate Carboxylase Dependent Metabolic Signal." *Molecular Microbiology* 106: 1018–1031.
- Lex, A., N. Gehlenborg, H. Strobel, R. Vuilleumot, and H. Pfister. 2014. "Upset: Visualization of Intersecting Sets." *IEEE Transactions on Visualization and Computer Graphics (InfoVis)* 20: 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>.
- Li, B., L. C. Tsoi, W. R. Swindell, et al. 2014. "Transcriptome Analysis of Psoriasis in a Large Case-Control Sample: RNA-Seq Provides Insights Into Disease Mechanisms." *Journal of Investigative Dermatology* 134: 1828–1838.
- Li, F., J. Eriksen, J. Finer-Moore, et al. 2020. "Ion Transport and Regulation in a Synaptic Vesicle Glutamate Transporter." *Science* 368: 893–897.
- Li, H., and H. S. Alper. 2016. "Enabling Xylose Utilization in *Yarrowia lipolytica* for Lipid Production." *Biotechnology Journal* 11: 1230–1240.
- Li, S. 2014. "Transcriptional Control of Flavonoid Biosynthesis: Fine-Tuning of the myb-bhlh-wd40 (mbw) Complex." *Plant Signaling & Behavior* 9: e27522.
- Liao, Y., G. K. Smyth, and W. Shi. 2014. "Featurecounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics* 30: 923–930.
- LiD, R. 2001. "Wd-Repeatproteins: Structure Characteristics, Biological Function, and Their Involvement in Human Diseases." *CellMolLifeSci* 58: 2085–2097.
- Lin, W., Y. Wang, Y. Chen, Q. Wang, Z. Gu, and Y. Zhu. 2021. "Role of Calcium Signaling Pathway-Related Gene Regulatory Networks in Ischemic Stroke Based on Multiple wgcna and Single-Cell Analysis." *Oxidative Medicine and Cellular Longevity* 2021: 8060477.
- Liu, L., and H. S. Alper. 2014. "Draft Genome Sequence of the Oleaginous Yeast *Yarrowia lipolytica* po1f, a Commonly Used Metabolic Engineering Host." *Genome Announcements* 2: 10–1128.
- Love, M. I., W. Huber, and S. Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for Rna-Seq Data With Deseq2." *Genome Biology* 15: 1–21.
- Lv, Y., M. Marsafari, M. Koffas, J. Zhou, and P. Xu. 2019. "Optimizing Oleaginous Yeast Cell Factories for Flavonoids and Hydroxylated Flavonoids Biosynthesis." *ACS Synthetic Biology* 8: 2514–2523.
- Madzak, C. 2018. "Engineering *Yarrowia lipolytica* for Use in Biotechnological Applications: A Review of Major Achievements and Recent Innovations." *Molecular Biotechnology* 60: 621–635.
- Madzak, C. 2021. "*Yarrowia lipolytica* Strains and Their Biotechnological Applications: How Natural Biodiversity and Metabolic Engineering Could Contribute to Cell Factories Improvement." *Journal of Fungi* 7: 548.
- Madzak, C., B. Tréton, S. Blanchin-Roland, et al. 2000. "Strong Hybrid Promoters and Integrative Expression/Secretion Vectors for Quasi-Constitutive Expression of Heterologous Proteins in the Yeast *Yarrowia lipolytica*." *Journal of Molecular Microbiology and Biotechnology* 2: 207–216.
- Merdinger, E., and E. M. Devine, Jr. 1965. "Lipids of *Debaryomyces hansenii*." *Journal of Bacteriology* 89: 1488–1493.
- Meyer, W., U. Bömer, and E. Pratje. 1997. "Mitochondrial Inner Membrane Bound pet1402 Protein Is Rapidly Imported Into Mitochondria and Affects the Integrity of the Cytochrome Oxidase and Ubiquinol-Cytochrome C Oxidoreductase Complexes." *Biological Chemistry* 378: 1373–1379.
- Michán, C., J. L. Martínez, M. C. Alvarez, M. Turk, H. Sychrova, and J. Ramos. 2013. "Salt and Oxidative Stress Tolerance in *Debaryomyces hansenii* and *Debaryomyces fabryi*." *FEMS Yeast Research* 13: 180–188.
- Miller, K. K., and H. S. Alper. 2019. "*Yarrowia lipolytica*: More Than an Oleaginous Workhorse." *Applied Microbiology and Biotechnology* 103: 9251–9262.
- Molin, M., J. Norbeck, and A. Blomberg. 2003. "Dihydroxyacetone Kinases in *Saccharomyces cerevisiae* Are Involved in Detoxification of Dihydroxyacetone." *Journal of Biological Chemistry* 278: 1415–1423.
- Morel, G., L. Sterck, D. Swennen, et al. 2015. "Differential Gene Retention as an Evolutionary Mechanism to Generate Biodiversity and Adaptation in Yeasts." *Scientific Reports* 5: 11571.
- Morimoto, R. I., M. P. Kline, D. N. Bimston, and J. J. Cotto. 1997. "The Heat-Shock Response: Regulation and Function of Heat-Shock Proteins and Molecular Chaperones." *Essays in Biochemistry* 32: 17–29.
- Morin, N., J. Cescut, A. Beopoulos, et al. 2011. "Transcriptomic Analyses During the Transition From Biomass Production to Lipid Accumulation in the Oleaginous Yeast *Yarrowia lipolytica*." *PLoS One* 6: e27966.
- Mukherjee, V., D. Radecka, G. Aerts, K. J. Verstrepen, B. Lievens, and J. M. Thevelein. 2017. "Phenotypic Landscape of Non-Conventional Yeast Species for Different Stress Tolerance Traits Desirable in Bioethanol Fermentation." *Biotechnology for Biofuels* 10: 1–19.

- Nacac, T. G., K. Leptien, D. Fellner, H. G. Augustin, and J. Kroll. 2006. "The btb-Kelch Protein Iztr-1 Is a Novel Golgi Protein That Is Degraded Upon Induction of Apoptosis." *Journal of Biological Chemistry* 281: 5065–5071.
- Nakov, T., K. J. Judy, K. M. Downey, E. C. Ruck, and A. J. Alverson. 2020. "Transcriptional Response of Osmolyte Synthetic Pathways and Membrane Transporters in a Euryhaline Diatom During Long-Term Acclimation to a Salinity Gradient." *Journal of Phycology* 56: 1712–1728.
- Navarrete, C., A. T. Frost, L. Ramos-Moreno, M. R. Krum, and J. L. Martínez. 2021. "A Physiological Characterization in Controlled Bioreactors Reveals a Novel Survival Strategy for *Debaryomyces hansenii* at High Salinity." *Yeast* 38: 302–315.
- Navarrete, C., B. J. Sánchez, S. Savickas, and J. L. Martínez. 2022. "Debaryomics: An Integrative-Omics Study to Understand the Halophilic Behaviour of *Debaryomyces hansenii*." *Microbial Biotechnology* 15: 1133–1151.
- Nishinaka, Y., H. Masutani, H. Nakamura, and J. Yodoi. 2001. "Regulatory Roles of Thioredoxin in Oxidative Stress-Induced Cellular Responses." *Redox Report* 6: 289–295.
- O'Brien, C. E., C. G. McCarthy, A. E. Walshe, et al. 2018. "Genome Analysis of the Yeast *Diutina Catenulata*, a Member of the *Debaryomycetaceae*/ *Metschnikowiaceae* (Ctg-Ser) Clade." *PLoS One* 13: e0198957.
- Ochoa-Gutiérrez, D., A. M. Reyes-Torres, I. de la Fuente-Colmenares, et al. 2022. "Alternative Cug Codon Usage in the Halotolerant Yeast *Debaryomyces hansenii*: Gene Expression Profiles Provide New Insights Into Ambiguous Translation." *Journal of Fungi* 8: 970.
- Ohlroge, J., and J. Browse. 1995. "Lipid Biosynthesis." *Plant Cell* 7: 957.
- Overbeek, R., N. Larsen, T. Walunas, et al. 2003. "The ergo tm Genome Analysis and Discovery System." *Nucleic Acids Research* 31: 164–171.
- Oyelade, J., I. Isewon, F. Oladipupo, et al. 2016. "Clustering Algorithms: Their Application to Gene Expression Data." *Bioinformatics and Biology Insights* 10: BBI-S38316.
- Ozaki, S., Y. Ogata, K. Suda, et al. 2010. "Coexpression Analysis of Tomato Genes and Experimental Verification of Coordinated Expression of Genes Found in a Functionally Enriched Coexpression Module." *DNA research* 17: 105–116.
- Palmer, C. M., K. K. Miller, A. Nguyen, and H. S. Alper. 2020. "Engineering 4-coumaroyl-coa Derived Polyketide Production in *Yarrowia lipolytica* Through a β -Oxidation Mediated Strategy." *Metabolic Engineering* 57: 174–181.
- Papanikolaou, S., and G. Aggelis. 2002. "Lipid Production by *Yarrowia lipolytica* Growing on Industrial Glycerol in a Single-Stage Continuous Culture." *Bioresource Technology* 82: 43–49.
- Papouskova, K., and H. Sychrova. 2007. "The Co-Action of Osmotic and High Temperature Stresses Results in a Growth Improvement of *Debaryomyces hansenii* Cells." *International journal of Food Microbiology* 118: 1–7.
- Park, Y.-K., and R. Ledesma-Amaro. 2023. "What Makes *Yarrowia lipolytica* Well Suited for Industry?" *Trends in Biotechnology* 41: 242–254.
- Pedregosa, F., G. Varoquaux, A. Gramfort, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–2830.
- Poli, J. S., M. A. N. da Silva, E. P. Siqueira, V. M. Pasa, C. A. Rosa, and P. Valente. 2014. "Microbial Lipid Produced by *Yarrowia lipolytica* qu21 Using Industrial Waste: A Potential Feedstock for Biodiesel Production." *Bioresource Technology* 161: 320–326.
- Pollastri, S., and M. Tattini. 2011. "Flavonols: Old Compounds for Old Roles." *Annals of Botany* 108: 1225–1233.
- Pomraning, K. R., Y.-M. Kim, C. D. Nicora, et al. 2016. "Multi-Omics Analysis Reveals Regulators of the Response to Nitrogen Limitation in *Yarrowia lipolytica*." *BMC Genomics* 17: 1–18.
- Ribeiro, R. A., N. Bourbon-Melo, and I. Sá-Correia. 2022. "The Cell Wall and the Response and Tolerance to Stresses of Biotechnological Relevance in Yeasts." *Frontiers in Microbiology* 13: 953479.
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. "Edger: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26: 139–140.
- Sáez-Sáez, J., G. Wang, E. R. Marella, S. Sudarsan, M. C. Pastor, and I. Borodina. 2020. "Engineering the Oleaginous Yeast *Yarrowia lipolytica* for High-Level Resveratrol Production." *Metabolic Engineering* 62: 51–61.
- Santamauro, F., F. M. Whiffin, R. J. Scott, and C. J. Chuck. 2014. "Low-Cost Lipid Production by an Oleaginous Yeast Cultured in Non-Sterile Conditions Using Model Waste Resources." *Biotechnology for Biofuels* 7: 1–11.
- Schmidt, W. 1999. "Mechanisms and Regulation of Reduction-Based Iron Uptake in Plants." *New Phytologist* 141: 1–26.
- Shakoury-Elizeh, M., J. Tiedeman, J. Rashford, et al. 2004. "Transcriptional Remodeling in Response to Iron Deprivation in *Saccharomyces cerevisiae*." *Molecular Biology of the Cell* 15: 1233–1243.
- Shcherbik, N., and D. G. Pestov. 2019. "The Impact of Oxidative Stress on Ribosomes: From Injury to Regulation." *Cells* 8: 1379.
- Shi, S., and H. Zhao. 2017. "Metabolic Engineering of Oleaginous Yeasts for Production of Fuels and Chemicals." *Frontiers in Microbiology* 8: 2185.
- Shima, J., A. Ando, and T. Nakamura, et al. 2010. "Environmental Stress Tolerance of Yeast: Importance in Industrial Uses and Molecular Mechanisms." *Nippon Shokuhin Kagaku Kogaku Kaishi=Journal of the Japanese Society for Food Science and Technology* 57: 225–231.
- Sibirny, A. A., and A. Y. Voronovsky. 2009. "*Candida famata* (*Debaryomyces hansenii*)." In Satyanarayana, T., G. Kunze, (Eds.), *Yeast Biotechnology: Diversity and Applications*. Springer. https://doi.org/10.1007/978-1-4020-8292-4_5.
- Sitepu, I., T. Selby, T. Lin, S. Zhu, and K. Boundy-Mills. 2014. "Carbon Source Utilization and Inhibitor Tolerance of 45 Oleaginous Yeast Species." *Journal of Industrial Microbiology and Biotechnology* 41: 1061–1070.
- Slabaugh, E., and F. Brandizzi. 2011. "Membrane-Tethered Transcription Factors Provide a Connection Between Stress Response and Developmental Pathways." *Plant Signaling & Behavior* 6: 1210–1211.
- Sohier, D., A.-S. Le Dizes, D. Thuault, C. Neuveglise, E. Coton, and S. Casaregola. 2009. "Important Genetic Diversity Revealed by Inter-ltr pcr Fingerprinting of *Kluyveromyces marxianus* and *Debaryomyces hansenii* Strains From French Traditional Cheeses." *Dairy Science and Technology* 89: 569–581.
- Solomon, E. I., U. M. Sundaram, and T. E. Machonkin. 1996. "Multi-copper Oxidases and Oxygenases." *Chemical Reviews* 96: 2563–2606.
- Spagnuolo, M., A. Yaguchi, and M. Blenner. 2019. "Oleaginous Yeast for Biofuel and Oleochemical Production." *Current Opinion in Biotechnology* 57: 73–81.
- Stojanovski, D., M. Bohnert, N. Pfanner, and M. van der Laan. 2012. "Mechanisms of Protein Sorting in Mitochondria." *Cold Spring Harbor Perspectives in Biology* 4: a011320.
- Stratford, M., H. Steels, M. Novodvorska, D. B. Archer, and S. V. Avery. 2019. "Extreme Osmotolerance and Halotolerance in Food-Relevant Yeasts and the Role of Glycerol-Dependent Cell Individuality." *Frontiers in Microbiology* 9: 3238.
- Sun, M.-L., T.-Q. Shi, L. Lin, R. Ledesma-Amaro, and X.-J. Ji. 2022. "Advancing *Yarrowia lipolytica* as a Superior Biomanufacturing Platform by Tuning Gene Expression Using Promoter Engineering." *Bioresource Technology* 347: 126717.
- Sun, Y., M. Shi, D. Wang, et al. 2023. "Research Progress on the Roles of Actin-Depolymerizing Factor in Plant Stress Responses." *Frontiers in Plant Science* 14: 1278311.

- Tanner, Jr, F., C. Vojnovich, and J. Van Lanen. 1945. "Riboflavin Production by *Candida* Species." *Science* 101: 180–181.
- Tatusov, R. L., N. D. Fedorova, J. D. Jackson, et al. 2003. "The Cog Database: An Updated Version Includes Eukaryotes." *BMC Bioinformatics* 4: 1–14.
- T'Jampens, D., L. Devriendt, V. De Corte, J. Vandekerckhove, and J. Gettemans. 2002. "Selected Btb/Poz-Kelch Proteins Bind atp." *FEBS Letters* 516: 20–26.
- Tsyrlunyk, A. O., Y. A. Andreieva, J. Ruchala, et al. 2020. "Expression of Yeast Homolog of the Mammal bcrp Gene Coding for Riboflavin Efflux Protein Activates Vitamin B2 Production in the Flavinoogenic Yeast *Candida famata*." *Yeast* 37: 467–473.
- Van Rossum, G., and F. L. Drake, Jr. 1995. *Python Reference Manual*. Centrum voor Wiskunde en Informatica Amsterdam.
- Verghese, J., J. Abrams, Y. Wang, and K. A. Morano. 2012. "Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (*Saccharomyces cerevisiae*) as a Model System." *Microbiology and Molecular Biology Reviews* 76: 115–158.
- Vickers, M. F., S. Y. Yao, S. A. Baldwin, J. D. Young, and C. E. Cass. 2000. "Nucleoside Transporter Proteins of *Saccharomyces cerevisiae*: Demonstration of a Transporter (fui1) With High Uridine Selectivity in Plasma Membranes and a Transporter (fun26) With Broad Nucleoside Selectivity in Intracellular Membranes." *Journal of Biological Chemistry* 275: 25931–25938.
- Voronovsky, A. Y., C. A. Abbas, K. V. Dmytruk, et al. 2004. "Candida famata (*Debaryomyces hansenii*) DNA Sequences Containing Genes Involved in Riboflavin Synthesis." *Yeast* 21: 1307–1316.
- Wagner, J. M., L. Liu, S.-F. Yuan, M. V. Venkataraman, A. R. Abate, and H. S. Alper. 2018. "A Comparative Analysis of Single Cell and Droplet-Based facts for Improving Production Phenotypes: Riboflavin Overproduction in *Yarrowia lipolytica*." *Metabolic engineering* 47: 346–356.
- Wallace, V. 2014. "Improving Stress Tolerance in Industrial *Saccharomyces cerevisiae* Strains for Ethanol Production From Lignocellulosic Biomass." Dissertation thesis. Lund University.
- Wang, J., R. Ledesma-Amaro, Y. Wei, B. Ji, and X.-J. Ji. 2020. "Metabolic Engineering for Increased Lipid Accumulation in *Yarrowia lipolytica*- A Review." *Bioresource Technology* 313: 123707.
- Wu, S., Z. Zhu, L. Fu, B. Niu, and W. Li. 2011. "Webmga: A Customizable Web Server for Fast Metagenomic Sequence Analysis." *BMC Genomics* 12: 1–9.
- Wu, Y., Y. Wang, H. Shi, H. Hu, L. Yi, and J. Hou. 2022. "Time-Course Transcriptome and wgcna Analysis Revealed the Drought Response Mechanism of Two Sunflower Inbred Lines." *PLoS One* 17: e0265447.
- Xue, L.-L., H.-H. Chen, and J.-G. Jiang. 2017. "Implications of Glycerol Metabolism for Lipid Production." *Progress in Lipid Research* 68: 12–25.
- Yaguchi, A., D. Rives, and M. Blenner. 2017. "New Kids on the Block: Emerging Oleaginous Yeast of Biotechnological Importance." *AIMS Microbiology* 3: 227.
- Yamasaki, S., and P. Anderson. 2008. "Reprogramming mRNA Translation During Stress." *Current Opinion in Cell Biology* 20: 222–226.
- Yang, T. J., W.-D. Lin, and W. Schmidt. 2010. "Transcriptional Profiling of the *Arabidopsis* Iron Deficiency Response Reveals Conserved Transition Metal Homeostasis Networks." *Plant Physiology* 152: 2130–2141.
- Zhang, J., X. Du, X. Zhou, D. Jin, J. Miao, and X. Liu. 2021. "An fyve-Domain-Containing Protein, psfp1, Is Involved in Vegetative Growth, Oxidative Stress Response and Virulence of *Phytophthora Sojae*." *International Journal of Molecular Sciences* 22: 6601.
- Zhang, X., F. Zink, F. Hezel, et al. 2020. "Metabolic Substrate Utilization in Stress-Induced Immune Cells." *Intensive Care Medicine Experimental* 8: 28.
- Zhang, Y., Y. Li, Y. Zhang, et al. 2016. "Quantitative Proteomics Reveals Membrane Protein-Mediated Hypersaline Sensitivity and Adaptation in Halophilic Nocardiosis *Xinjiangensis*." *Journal of Proteome Research* 15: 68–85.
- Zhao, X., and F. Bai. 2009. "Mechanisms of Yeast Stress Tolerance and Its Manipulation for Efficient Fuel Ethanol Production." *Journal of Biotechnology* 144: 23–30.
- Zhao, Y., T. Ma, and D. Zou. 2021. "Identification of Unique Transcriptomic Signatures and Hub Genes Through RNA Sequencing and Integrated wgcna and ppi Network Analysis in Nonerosive Reflux Disease." *Journal of Inflammation Research* 14: 6143–6156.
- Zhu, M., H. Xie, X. Wei, et al. 2019. "Wgcna Analysis of Salt-Responsive Core Transcriptome Identifies Novel Hub Genes in Rice." *Genes* 10: 719.
- Zhu, Y., J. Liu, L. Sun, M. Liu, Q. Qi, and J. Hou. 2024. "Development of Genetic Markers in *Yarrowia lipolytica*." *Applied Microbiology and Biotechnology* 108: 14.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.