Elucidating Mechanisms and Genotypes Underlying Robust Phenotypes in *Yarrowia lipolytica*

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Caleb M. Walker
May 2021

Copyright © 2021 by Caleb M. Walker.
All rights reserved.

Dedicated to my parents

Jane and Jeff

And my grandparents

Jean, Fred and Judy

ABSTRACT

Robustness is an important phenotype for bioenergy microbes to acquire but is difficult to engineer. Hence, tools for engineering microbial robustness are critical to unlock novel phenotypes for innovative bioprocessing strategies. The oleaginous yeast, *Yarrowia lipolytica*, is an exceptionally robust microbe that can tolerate stressful environments, assimilate a wide range of substrates, and produce high-value chemicals. In this doctoral dissertation, the impacts of systems biology and metabolic engineering to reveal mechanisms and identify genotypes- underlying robust phenotypes are addressed.

The first approach employs adaptive laboratory engineering to generate a platform strain by which to study superior robust mechanisms. This approach generated the most solvent-tolerant microorganism reported to date, identified sterols are critical for solvent tolerance in *Y. lipolytica*, and reverse engineered high solvent tolerance by increasing sterol biosynthesis via overexpression of the sterol transcription factor (Chapter I). Similarly, short-term adaptation of *Y. lipolytica* to depolymerized plastic waste improved utilization of all hydrophobic substrates tested and acquired a mechanism enabling cellular adhesion to the hydrophobic layer (Chapter V).

A more elaborate approach is developed using temporal RNA-sequencing data to directly predict genotypes underling robust phenotypes. Here, a novel concept was established that ranks differentially expressed genes by their co-expression connectivity to predict top-performing genetic targets conferring robustness. This methodology identified genes that would normally be overlooked by common differential expression approaches yet proved to have the highest prediction of genetic targets conferring solvent tolerance in *Y. lipolytica* (Chapter III).

We approached engineering an absent genotype (e.g., thiamine auxotroph) using comparative genomics with a thiamine prototroph strain and discovered the missing gene for thiamine biosynthesis (Chapter II). Bioinformatics in conjunction with metabolic engineering elucidated a promoter tightly regulated by thiamine concentration that enabled strong expression of the missing gene and restoration of thiamine biosynthesis in *Y. lipolytica* (Chapter II).

Finally, we exploited the phenotypic diversity exhibited by non-conventional *Y. lipolytica* isolates to reveal key processes and regulators underlying xylose metabolism and lipid accumulation or degradation from biomass hydrolysate (Chapter IV).

In conclusion, the outcome of this research is to develop strategies to elucidate and establish genetic tractability of robust phenotypes.

TABLE OF CONTENTS

INTRODUCTION 1
Robustness. 1
Outline of Dissertation 4
CHAPTER I Exceptional Solvent Tolerance in Yarrowia lipolytica Is Enhanced by Sterols 9
Abstract 10
Introduction 10
Results 11
Generate robust Y. lipolytica mutants via adaptive laboratory evolution as a basis for elucidating IL-tolerant mechanism 11
Elucidate IL-responsive physiology and metabolism in Y. lipolytica strains 14
Sterol biosynthesis is one key IL-responsive process to improve IL-tolerance in Y. lipolytica 18
Discussion 20
Methods 22
Strains and plasmids 22
Analytical methods 24
Acknowledgements 29
References 30
Appendix 38
CHAPTER II Understanding and eliminating the detrimental effect of thiamine deficiency on the oleaginous yeast *Yarrowia lipolytica.* 60
Abstract 61
Introduction 61
RESULTS 63
The effects of thiamine deficiency in Y. lipolytica 63
Restoring thiamine prototrophy in Y. lipolytica 66
DISCUSSION 69
MATERIALS AND METHODS 71
Plasmids and strains 71
Media and culturing conditions 73
Analytical methods 73
ACKNOWLEDGEMENTS 77
References 78
Appendix 86
CHAPTER III Co-expression Connectivity Predicts Genetic Targets Underlying Solvent Tolerance in *Yarrowia lipolytica* 98
Abstract 99
Introduction 99
Results 101
Genotype of the IL-tolerant evolved strain YlCW001 101

Basal gene expression differences between the WT and MT strains without IL exposure 101
IL-responsive gene expression 102
Enhanced IL-responsive gene expression 103
Discussion 106
Materials and Methods. 108
Strains 108
Plasmids 108
Medium and culturing conditions 109
Analytical methods 109
References 112
Appendix 114
Chapter IV Exploring Proteomes of Robust *Yarrowia lipolytica* Isolates Cultivated in Biomass Hydrolysate Reveal Key Processes Impacting Complex Sugar Utilization, Lipid Accumulation, and Degradation 125
ABSTRACT 126
IMPORTANCE 127
INTRODUCTION 128
RESULTS 130
Comparative genomics reveals unique genotypes of undomesticated Yarrowia strains 130
Proteomic analysis reveals key processes impacting sugar utilization and lipid degradation 131
DISCUSSION 136
MATERIALS AND METHODS 138
Strains 138
Medium and culturing conditions 138
Analytical methods 140
Bioinformatics and data analysis 141
ACKNOWLEDGEMENT 143
References 144
Appendix 149
Chapter V Harnessing *Yarrowia lipolytica* for Biological Upcycling of Plastic Waste into Citric Acid and Lipids 165
Abstract 166
Introduction 166
Materials and Methods 168
Medium and culturing conditions 168
Analytical methods 169
Results 170
Growth and adaptation of Y. lipolytica on synthetic oil. 170
Stability analysis of synthetic oil. 171
Lipid and citric acid production from synthetic oi 171
Discussion 172

References 175
Appendix 177
CONCLUSION AND FUTURE DIRECTION 183
VITA 186

LIST OF TABLES

Table 1- 1. Untargeted LCMS summary of perturbed pathways with pathway significance factor > 0.58. 38
Table 1- 2. Locus tags of the sterol pathway genes and sterol transcription factor. 40
Table 1- 3. List of primers used in this study. Bold italics are AscI and AvrII sequences. 41
Table 1- 4. List of plasmids and strains used in this study. 43

Table 2-1 List of plasmids and strains. 86
Table 2-2 List of primers. 86
Table 2- 3. Protein expression of thiamine-dependent enzymes at both exponential-growth time points. 87
Table 2- 4. List of putative thiamine-regulated genes in *Y. lipolytica*. Abbreviations: PP: *P. pastoris*, SC: *S. cerevisiae*, and SP: *S. pombe*. 88

Table 3- 1. Primers used for genetic target overexpression. 114
Table 4- 1. Singleton genes of YB392, YB419, YB420, YB566 and YB567. 149
Table 4- 2. Exclusively shared genes between xylose consuming strains YB566 and YB567. 150
Table 4- 3. Exclusively shared genes between lipid accumulating strains YB392, YB419 and YB420. 151
Table 4- 4. Proteomics analysis of xylose metabolism and transporters. 151
Table 4- 5. Proteomics analysis of lipid metabolism, lipid regulators, lipase and NADPH generating proteins. 152

LIST OF FIGURES

Figure 1- 1. Evolutionary engineering of IL-tolerant *Y. lipolytica* strains. Step 1: Adaptive laboratory evolution of *Y. lipolytica* on various concentrations of [EMIM][OAc]. Step 2: Single mutant isolation on plates without IL. Step 3: Mutant characterization to elucidate the underlying mechanism of solvent tolerance. 44

Figure 1- 2. (A) Adaptive laboratory evolution of *Y. lipolytica* grown with increasing concentrations of [EMIM][OAc]. (B, C) Irreversibility testing of the evolved *Y. lipolytica* strain YlCW001 in various concentrations of [EMIM][OAc]. (D, E) Broad tolerance test for growth of *Y. lipolytica* wildtype and mutant at various concentrations of 0.6M and 1.09M ILs. Error bars represent the standard deviation of three independent biological replicates repeated twice (n=6). Abbreviations: [EMIM][OAc]: 1-ethyl-3-methylimidizolium acetate; [EMIM][Cl]: 1-ethyl-3-methylimidizolium chloride; [EMIM][Br]: 1-ethyl-3-methylimidizolium bromide acetate; [BMIM][OAc]: 1-butyl-3-methylimidazolium; [BMIM][Cl]: 1-butyl-3-methylimidizollium chloride; [BMIM][Br]: 1-butyl-3-methylimidazolium bromide; and [AMIM][Cl]: 1-allyl-3-methylimidizolium chloride. 45

Figure 1- 3. SEM imaging of the wildtype and YlCW001 cells exposed to (A, B) no IL, (C, D) 18% (v/v) [EMIM][OAc], (E, F) 0.3M [BMIM][OAc], and (G, H) 0.6M [BMIM][OAc]. (I) IL-responsive pathways perturbed with respect to wildtype *Y. lipolytica* growing in medium without IL with a psf score > 0.58. Biological triplicates were performed for all biological conditions and results were combined from hydrophilic metabolomics extraction (negative ionization, n = 3), and hydrophobic lipidomics extractions (negative ionization, n = 3; positive ionization, n = 3). 47

Figure 1- 4. Glycerophospholipid and fatty acid reorganization in wildtype *Y. lipolytica* and YlCW001 in 0% and 8% (v/v) [EMIM][OAc]. (A) PC: Phosphatidylcholine; (B) PI: phosphatidylinositol; (C) PA: phosphatidic acid; (D) PG: phosphatidyl glycerol; (E) PS: phosphatidylserine; (F) PE: phosphatidylethanolamine; (G) CL: cardiolipin. (H) Fatty acid distributions. (I) Chemical structures of glycerophospholipid backbone and headgroup species. All error bars represent standard deviation of biological triplicates (n=3) and statistical significance was calculated using one-way analysis of variance (ANOVA) with Holm-Sidak correction against control group, wildtype in 0% IL (degrees of freedom, 3). Symbols: “*”: *p*-value < 0.05; “**”: *p*-value < 0.01; “***”, *p*-values < 0.001. 49

Figure 1- 5. (A) Sterol biosynthesis pathway in *Y. lipolytica*. Fluconazole is an antifungal drug that inhibits sterol 14-demethylase (Ster4) (B) Differential gene expression of the steroid pathway as compared to the wildtype strain in medium without IL (Ster1, YALI0A10076g; Ster2, YALI0E15730g; Ster3, YALI0F04378g; Ster4, YALI0B05126g; Ster5, YALI0B23298g; Ster6, YALI0F11297g; Ster7, YALI0C22165g; Ster8, YALI0B17644g; Ster9, YALI0F08701g; Ster10-1, YALI0E32065g; Ster10-2, YALI0B17204g; Ster11, YALI0D20878g; Ster12, YALI0A18062g; Ster13, YALI0D19206g). (C) Effect of fluconazole inhibiting the steroid pathway on cell growth relative to wildtype in non-IL, non-fluconazole media. Error bars represent standard deviation from three technical replicates repeated twice per biological condition (n = 6) (D) Relative ergosterol content of

wildtype and YlCW001 strains in 0% and 8% (v/v) [EMIM][OAc]. (E) Toxicity plate assay of *Y. lipolytica* strains expressing the sterol biosynthesis and transcription factor genes in presence of various concentrations of [EMIM][OAc]. Error bars represent standard deviation of biological triplicates (n =3) and statistical significance was calculated between the wildtype and YlCW001 in 8% IL using the Student's t-test (t = -4.244 with 4 degrees of freedom). Abbreviation: "**", *p*-value = 0.013. 51
Figure 1- 6. Mechanisms of IL interference and tolerance in wildtype *Y. lipolytica* and YlCW001. (A, D) No IL exposure. (B, E) IL exposure and interference. (C, D) Cellular membrane response. 53
Figure 1- 7. Effect of 0.3M [BMIM][OAc] on growth of *Y. lipolytica* wildtype and YlCW001. Each data point is an average ± standard deviation of three independent biological replicates. 54
Figure 1- 8. SEM images of wildtype Y. lipolytica cells incubated for 6 hours in (A) no IL, (B) 2% (v/v) [EMIM][OAc], (C) 8% (v/v) [EMIM][OAc] and (D) 12% (v/v) [EMIM][OAc]. 55
Figure 1- 9. Effect of various concentrations [EMIM][OAc] on the chitin content in the membranes of the wildtype and YlCW001. Each experiment was conducted at least twice for each biological condition in technical replicates, i.e., wildtype 0%, n = 18; wildtype 2%, n = 6; wildtype 5%, n = 15; wildtype 8%, n = 12; YlCW001 0%, n = 25; YlCW001 2%, n = 6; YlCW001 5%, n = 15; YlCW001 8%, n = 29; YlCW001 10%, n = 6. Each error bar represents standard deviation of technical replicates and statistical significance was calculated using one-way analysis of variance (ANOVA) with Holm-Sidak correction against all pairwise groups (degree of freedom=8, F=18.53). Asterisks represent statistical significance against respective strain in 0% IL where *, p-values < 0.01, **p-values < 0.001. 56
Figure 1- 10. Fatty acid molar ratios of wildtype and YlCW001 in 0% and 8% (v/v) [EMIM][OAc]. (A) Molar ratio of C16:C18 fatty acid moieties. (B) Molar ratio of (C18:2):(C18:1) fatty acid moieties. Each error bar represents standard deviation of 3 biological replicates and statistical significance was calculated using one-way analysis of variance (ANOVA) with Holm-Sidak correction against control group, wildtype in 0% IL (degree of freedom = 3). *P values < 0.05, **P values < 0.01, ***P values < 0.001. 57
Figure 1- 11. (A) Growth kinetics and (B) ergosterol content of SterTF overexpression in 8% (v/v) [EMIM][OAc] relative to wildtype in 0% and 8% (v/v) [EMIM][OAc] at 36 h. Error bars represent the average and standard deviation of biological triplicates. Statistical significance was calculated between the wildtype and SterTF-expressing strains in 8% (v/v) [EMIM][OAc] using one-way analysis of variance (ANOVA) with Holm-Sidak correction. 58
Figure 1- 12. Transcription factor affinity prediction (TRAP) for sterol biosynthesis genes. TRAP was performed for 1050bp nucleotide sequences upstream of each sterol biosynthesis gene and pTEF promoter as negative control. 59

Figure 2- 1. Metabolic map of thiamine-dependent enzymes (green) in relationship to central (black) and peripheral (blue) pathways. Abbreviations: TPP: thiamine pyrophosphate dependent enzymes, PDH: pyruvate dehydrogenase enzyme complex, KGDH: a-ketoglutarate dehydrogenase, TKL: transketolase, AHAS: acetolactate synthase, BCKDC: branched chain α-ketoacid dehydrogenase, TPP: thiamine pyrophosphate, AAs: amino acids, TCA: citrate (Krebs) cycle, KGA: alpha-ketoglutarate.dehydrogenase. 89
Figure 2- 2. Growth characterization of the thiamine-auxotrophic *Y. lipolytica* YlSR001 in 0 (red), 0.5 (blue) and 400 (purple) µg/L thiamine. (A) Scheme of thiamine depletion design experiments. (B) Cell growth profiles. (C) Glucose consumption profiles. (D) KGA production profiles. (E) Pyruvate production profiles. 90
Figure 2- 3. Proteomics of the thiamine-auxotrophic *Y. lipolytica* YLSR001 grown in 0 (red) and 400 (purple) µg/L thiamine. (A) Growth profiles and proteomics samples indicated by arrows. (B) Venn diagram representing upregulated and downregulated proteins at both exponential time points. (C) Thiamine-responsive proteins on metabolic pathways. Bar graphs represent the percentage of proteins upregulated in 0 (red) or 400 (purple) µg/L thiamine for each pathway (D) Electron transport chain and ATPase. (E) Thiamine metabolism of *Y. lipolytica*. Enzymes: Thi4P (YALI0A09768p, thiamine thiazole synthase), Thi11p (YALI0E04224p, 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase), Thi6p (YALI0C15554p, bifunctional hydroxyethylthiazole kinase/thiamine-phosphate diphosphorylase), PHO (YALI0A12573p, acid phosphatase), Thi80p (YALI0E21351p, thiamine diphosphokinase), Thi13p (4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase), and ADK (YALI0F26521p, adenylate kinase). Abbreviations: PDH: pyruvate dehydrogenase enzyme complex, KGDH: a-ketoglutarate dehydrogenase, TKL: transketolase, AHAS: acetolactate synthase, BCKDC: branched chain α-ketoacid dehydrogenase, TPP: Thiamine pyrophosphate, HET-P: hydroxyethylthiazole phosphate, HMP-PP: hydroxymethylpyrimidine pyrophosphate, TP: thiamine monophosphate, TPP: thiamine pyrophosphate, TPPP: thiamine triphosphate. 91
Figure 2- 4. Thiamine-responsive promoter characterization. (A) Chromosomal gene expression of thiamine-responsive genes using rt-PCR. (B) hrGFP protein expression. (C, D) Sensitivity of thiamine-responsive promoters in presence of increasing thiamine concentrations. 93
Figure 2- 5. Expression of pP3-scTHI13 in YlSR1006 restores the *de novo* thiamine biosynthesis in *Y. lipolytica*. (A) Cell growth profiles. (B) Glucose consumption profiles. (C) KGA production profiles. (D) Pyruvate production profiles. Growth characterization was conducted in 0 (red), 0.5 (blue) and 400 (purple) µg/L thiamine. Abbreviations: KGA, alpha-ketoglutarate. 94
Figure 2- 6. Lipid production is influenced by thiamine availability. (A) Lipid accumulation profiles for the thiamine-auxotrophic wildtype YlSR001 and thiamine-prototrophic strain YlSR1006 in 0 and 400 µg/L thiamine. (B) Lipid accumulation profiles for thiamine-auxotrophic and prototrophic strains in lipid production (C:N = 100) media with 0 and 400 µg/L thiamine. Statistical significance was calculated

with one-way analysis of variance (ANOVA) with Holm-Sidak correction between 0 and 400 µg/L thiamine for each strain respectively. Symbols: "*": p-value < 0.05.. 95
Figure 2- 7. A comprehensive model explains detrimental consequences of thiamine deficiency on metabolism and cellular growth of *Y. lipolytica*..... 96
Figure 2- 8. Constitutive expression of missing thiamine gene does not effectively restore thiamine prototrophy in Y. lipolytica. Wildtype is in blue and 6 replicates of YlSR005 growth without thiamine (0 µg/L thiamine) in orange..... 97
Figure 3- 1. Mutated genes of IL-tolerant mutant (MT, YlCW001) *Y. lipolytic*a which include 68 variants across 40 genes that change amino acid residues. (A) Chromosomal location of mutated genes depicting log2 fold changes at the mid-exponential RNA-seq sample between the MT in 8% [EMIM][OAc] verses WT in 0% IL (purple), WT in 8% IL (blue) and MT in 0% IL (green). (B) Types of variants in the mutated genes. (C) Gene ontology of the mutated genes..... 116
Figure 3- 2. Basal differences in gene expression between the MT and WT in 0% IL. (A) Growth of MT and WT in 0% IL. (B) Distribution of classified genes. (C) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition..... 117
Figure 3- 3. IL-responsive gene expression of the WT strain. (A) Growth of WT in 8% and 0% IL. (B) Distribution of classified genes. (C) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition..... 118
Figure 3- 4. IL-responsive gene expression of the MT strain. (A) Growth of MT in 8% and 0% IL. (B) Distribution of classified genes. (C) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition..... 119
Figure 3- 5. Enhanced IL-responsive gene expression of the MT strain verses the WT strain in 8% IL. (A) Growth of MT and WT in 8% IL. (B) Distribution of classified genes. (C) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition.. 120
Figure 3- 6. (A) Correlation (Co-expression) network of classified genes between the MT and WT in 8% IL. (B) Subnetwork of co-expressed genes upregulated by the MT in 8% IL illustrating gene connectivity by degree. (C) Genetic targets selected by degree, fold change or gene ontology..... 121
Figure 3- 7. Individual overexpression of genetic targets in 11% IL. (A) Box and whisker plot for each group of genetic targets depicting all replicate maximum OD 600nm for all associated genes contained within each group. (B) Max OD600nm for genetic targets cultured in 11% IL. *p-value ≤ 0.05 (n ≥ 9 for each gene and n=27 for the empty vector control). 123
Figure 3- 8. Gene classification methodology. Step 1: For each pairwise set of biological conditions, gene TPM values are floored to a minimum of 5, averaged,

and converted into log_2 scale. Step 2: Fold change (X and Y) and regulation (Z) scores are calculated between biological conditions. Step 3: Genes are classified based on the scores of X, Y and Z into upregulated or increasing for a biological condition, changed regulation, or no change. 124
Figure 4- 1. (A) Phylogenetic tree of *Yarrowia lipolytica* isolates with 20 closest neighbor species. (B) Pangenome of *Y. lipolytica* reference strains CLIB122122 and CBS7504 (W29) and non-conventional strains YB392, YB419, YB420, YB566 and YB567. Bold, strain; parenthesis, total genes; outer petals, singleton genes, middle petals, uniquely shared genes between lipid producers (blue) and xylose consumers (green); middle circle, core genes. 154
Figure 4- 2. (A) Workflow of bioreactor characterization of *Yarrowia lipolytica* isolates CBS7504 (B, D) and YB420 (C, E) in 50% SGH. Dotted lines, time of glucose depletion where xylose is the sole remaining carbon source. 156
Figure 4- 3. Proteomic analysis of *Y. lipolytica* (A-C) CBS7504 and (D-F) YB420 strains. (A, D) Number of proteins with significant abundance changes relative to S1. (B, E) Heatmap of significant proteins with changed abundance relative to S1. (C, F) Venn diagram illustrating the number of proteins with significant abundance changes between samples..... 158
Figure 4- 4. Pentose phosphate pathway proteome of (A) CBS7504 and (B) YB420. Increased protein abundance during xylose assimilation, red; decreased protein abundance during xylose assimilation, blue; pathways, rectangles; proteins, bold font; metabolites, plain text. (C) All transporters and (D) carbohydrate transporters with significant protein abundance changes in the xylose assimilation phase. CBS7504, blue; YB420, orange. Abbreviations: TKL (transketolase, YALI0D02277g); PRS1 (ribose-phosphate pyrophosphokinase, YALI0B13552g); Xyl1 (xylose reductase, YALI0D07634g); Xyl2 (xylitol dehydrogenase, YALI0E12463g); Xyl3 (xylulokinase, YALI0F10923g); ZWF1 (NADP+-dependent glucose-6-phosphate dehydrogenase, YALI0E22649g); SOL3 (6-phosphogluconolactonase, YALI0E11671g); RPIB (ribose 5-phosphate isomerase, YALI0F01628g); TAL (transaldolase, YALI0F15587g); RPE1 (ribulose-phosphate 3-epimerase, YALI0C11880g); GND1 (6-phosphogluconate dehydrogenase, YALI0B15598g)..... 159
Figure 4- 5. (A) Schematic of growth characterization phenotypes for CBS7504 and YB420 when xylose was the sole remaining carbon source. Proteomic analysis of lipid metabolism showing increased (red) and decreased (blue) protein abundance when xylose was the sole remaining carbon source for CBS7504 (B) and YB420 (C). Proteins, circles; metabolites, regular font; pathways, bold font. Abbreviations: TAG (triacylglycerol); DHAP (dihydroxyacetone phosphate); Gly-3P (glycerol-3-phosphate); LPA (lysophosphatidic acid); PA, phosphatidic acid; DAG, diacylglycerol; PL, phospholipid; TAG, triacylglycerol; FFA, free fatty acid; TCA, tricarboxylic acid cycle; GPD1 (glycerol-3-phosphate dehydrogenase, YALI0B02948g); SCT1 (bifunctional glycerol-3-phosphate/glycerone-phosphate O-acyltransferase, YALI0C00209g); ALE1 (lysophosphatidate acyltransferase, YALI0F19514g); SLC1 (1-acyl-sn-glycerol-3-phosphate acyltransferase,

YALI0E18964g); PAP (phosphatidic acid phosphohydrolase, YALI0D27016g); LPP1 (diacylglycerol diphosphate phosphatase/phosphatidate phosphatase , YALI0B14531g); DGA1 (acyl-CoA dependent diacylglycerol acyltransferase I, YALI0E32769g); DGA2 (acyl-CoA dependent diacylglycerol acyltransferase II, YALI0D07986g); TGL3 (triacylglycerol lipases 3, YALI0D17534g); TGL4 (triacylglycerol lipase 4, YALI0F10010g); YJU3 (monoglycerol lipase, YALI0C14520g); FAA1 (long-chain acyl-CoA synthetase, YALI0D17864g); FAS1 (fatty acid synthase 1,YALI0B15059g); FAS2 (fatty acid synthase 2; YALI0B19382g); ACC1 (acetyl-CoA carboxylase, YALI0C11407g); PDH1, PDH2; ACL1 (ATP-citrate lyase, YALI0E34793g); ME (malic enzyme, YALI0E18634g); MnDH1 (Mannitol dehydrogenase, YALI0B16192g); UGA2 (Succinate semialdehyde dehydrogenase, YALI0F26191g); SOL3 (6-Phosphogluconolactonase, YALI0E11671g); ZWF1 (NADP+-dependent glucose-6-phosphate dehydrogenase, YALI0E22649g); POT1 (3-oxyacyl-thiolase, YALI0E18568g); MFE1 (multifunctional β-oxidation protein, YALI0E15378g); PEX10 (peroxin-10, YALI0C01023g); POX1 (acyl-coenzyme A oxidase 1, YALI0E32835g); POX2 (acyl-coenzyme A oxidase 2, YALI0F10857g); POX3 (acyl-coenzyme A oxidase 3, YALI0D24750g); POX4 (acyl-coenzyme A oxidase 4, YALI0E27654g); POX5 (acyl-coenzyme A oxidase 5, YALI0C23859g); POX6 (acyl-coenzyme A oxidase 6, YALI0E06567g); YAS2 (HLH transcription factor, YALI0E32417g); POR1 (YALI0D12628g); SNF1 (YALI0D02101g); SAK1 (YALI0D08822g); REI1 (YALI0B08734g); HSF1 (YALI0E13948g); UPC2 (YALI0B15818g)..... 161

Figure 4- 6. Regulator proteins with significant protein abundance changes in the xylose assimilation phase. CBS7504, blue; YB420, orange..... 164

Figure 5- 1. (A) Condensation flow reactor scheme to generate DP. (B) Images and (C) composition of DP oil. 178

Figure 5- 2. Growth of wildtype *Y. lipolytica* in (A) culture tubes and (B) baffled flasks. (C) Short-term adaptation of *Y. lipolytica* in DP oil to generate adapted strain. (D) Growth comparison of wildtype and adapted *Y. lipolytica* in glucose, hexadecane and DP oil. (E) Image of adapted *Y. lipolytica* cultured in DP oil after centrifugation and microscopic image of adhering cells. *p-value < 0.001 using one way analysis of variance Holm-Sidak method. 179

Figure 5- 3. (A) Workflow to generate DP oil stored at RT or -20°C for stability characterization. (B) Growth of adapted *Y. lipolytica* on DP oil stored at RT or -20°C after 3, 6, 9 and 14 weeks. (C) Images of DP oil stored in RT at week 0 and week 14. *p-value < 0.001 using one way analysis of variance Holm-Sidak method. 181

Figure 5- 4. Production of citric acid and lipids from DP oil using adapted *Y. lipolytica*. (A) Growth, (B) citric acid production, (C) lipid accumulation, (D) and DP oil consumption profiles. (E) Fluorescent image of adapted *Y. lipolytica* cultured in DP oil and stained with Bodipy (as described in lipid quantification method). 182

INTRODUCTION

The genetic diversity of microbial organisms offers a plethora of desirable phenotypes exploitable for bioprocessing. The selection of microorganism(s) for bioproduction heavily depends on the strain's native phenotype(s) and genetic tractability to the required phenotype(s) to realize economic bioprocessing in a given environment, e.g., carbon source and chemical production. To this end, genetic tractability of desirable phenotypes enables greater flexibility for host organism selection. Accordingly, a microorganism displaying a novel phenotype could be used for biotransformation if the other required phenotypes are genetically tractable. For example, *Saccharomyces cerevisiae* produces high titers of ethanol from glucose but insufficiently assimilates 5-carbon (C5) sugars (e.g., xylose, arabinose) (1) and has poor tolerance of biomass inhibitors (e.g., furfural, HMF) (2). For *S. cerevisiae* to be selected for biotransformation of biomass hydrolysate, C5 sugars assimilation and tolerance of biomass inhibitors are required phenotypes. While C5 biomass sugar assimilation is genetically tractable and can be genetically engineered, tolerance of biomass inhibitors is largely unknown (3), obstructing economic bioprocessing with *S. cerevisiae*. For these reasons, linking genotype(s) to phenotype(s) sanctions efficient and economic biotransformation developmental strategies.

Robustness.

Microbial robustness encompasses the ability to withstand or overcome adverse conditions which are critical phenotypes to unlock and enable innovative and strategical biotransformation processes. These robust phenotypes can include resistance to chemical inhibitors (e.g., inhibitors arising from substrate processing, metabolic intermediates, or secreted products), cell growth on diverse or multiple substrates (e.g., biomass hydrolysate, hydrophobic substrates), overcoming nutrient limitations (e.g., essential vitamin or amino acid auxotroph), or withstanding stressful environments (e.g., pH, salinity, temperature). For example, assimilation of biomass-derived sugars necessitates microbial tolerance of the inhibitors released in biomass pretreatment steps (e.g., furfural) (4). Efficient production of biofuels also demands cellular resistance to high

concentrations of the alcohol (e.g., butanol) or diesel (e.g., fatty acid methyl esters) produced (5). *In situ* separation and recovery of hydrophobic, toxic products requires microbial resilience in the presence of organic solvents (e.g., alkanes, ionic liquids) (6).

Though robustness is broad in scope, engineering any individual robust phenotype is particularly challenging due to the complexity of toxic/inhibitory modes of action, biological resistance mechanisms, and genetic regulation controlling cellular metabolism and physiology. For instance, metabolic pathway engineering (e.g., substrate assimilation or product formation) is generally more predictable (i.e., genetically tractable), enabling a more rational approach (7) than robustness engineering. In fact, much more is known about how cells respond to chemicals (e.g., stress proteins, efflux pumps and altered membrane compositions) than how to develop more tolerant strains (4). To this end, tools for enhancing and engineering microbial tolerance to growth inhibitors are imperative to unlock novel phenotypes with important implications.

To rationally engineer robustness, knowledge of tolerant mechanisms or inhibitor modes of action are required. It is therefore critical to elucidate superior metabolic processes responsible for enhanced tolerance to chemical inhibition. One of the most frequently described consequence(s) of chemical inhibition is characterized by membrane damage or reorganization (7). For example, realizing that alcohols primarily inhibit cell growth by membrane leakage allowed engineering more-robust *Escherichia coli* strains by overexpressing a heterologous cis-trans isomerase to help incorporate trans unsaturated fatty acids into the membrane and restore (i.e., decrease) membrane fluidity (8). Additionally, pinpointing membrane damage as the major toxic mechanism of common bioproducts enabled rational engineering of phospholipid composition to enhance robustness to membrane damaging chemicals (i.e., fatty acids, furfural and acetate) (9). These studies highlight the importance of elucidating key metabolic processes responsible for enhanced tolerance of chemical inhibition to aide rational design of robust phenotypes.

In contrast to rational engineering, the most successful methods to improve complex phenotypes employ selection-based approaches such as metabolic evolution (7). In these cases, a strain is challenged by — and optimized isolates are selected

directly from — the inhibiting environment until the desired phenotype is achieved or the strain's performance plateaus. Once the desired phenotype has been successfully evolutionary-engineered, the strain can either be used as the host organism for bioproduction or the desired property is transferred to a different production host. The fundamental requirement of reverse engineering is genetic tractability, that is, the genotype(s) must be linked to the robust phenotype. Consequently, successful identification of the genetic basis for selection-based phenotypes (i.e., nontargeted strain improvement) has rarely been accomplished, preventing reverse engineering of the desired phenotype (10).

As forementioned, microbial robustness is often a complex phenotype controlled by multifaceted genotypes that are either intractable or unknown. A logical approach toward engineering robustness is to obtain or generate a strain with enhanced phenotypic performance by which to study superior metabolic and genotypic tolerance mechanisms. To enhance a complex phenotype with minimal genotypic information, three methods are commonly employed: random mutagenesis, site-directed evolution and adaptive laboratory evolution (ALE). While site-directed evolution requires an enzymatic target (11) and genome-wide random mutagenesis can result in unsuccessful, non-beneficial or lethal mutations (12), ALE exploits genome-wide natural selection to generate tactical mutations that benefit phenotypic performance without prior knowledge of genetic targets (13). Perhaps the most advantageous aspect of ALE is the requirement of a challenging environment which renders ALE perfect for enhancing robustness to chemical inhibitors. While ALE methodology is not in itself a novel methodology, ALE can be significantly optimized when performed on a microorganism that natively exhibits elevated phenotypic performance (i.e., non-model organisms) to generate an exceptionally novel phenotype.

After an enhanced strain is obtained or generated, mechanisms of robustness are identified using comparative OMICs. Specifically, DNA-sequencing for genotype variances, RNA-sequencing for differentially expressed or co-expressed genes, proteomics for differentially abundant proteins, metabolomics for affected pathways, and lipidomics for changes in lipid composition. Together, comparative OMICs provides a

global overview of cellular changes but rarely provides clear genetic candidates for reverse engineering or interpretable mechanisms underlying robustness.

Nonetheless, a common approach for reverse engineering complex phenotypes starts by analyzing transcriptomes or proteomes to identify candidate genes/proteins conferring the desired phenotype (14-16). Traditionally, samples are generated from case (e.g., strain grown in stressful environment) and control (e.g., strain grown without stressful environment) and transcripts are analyzed to identify differentially expressed genes (17). While most transcriptomics studies involve only static samples (i.e., single time point), dynamic (i.e., multiple time points) transcriptomics data is more meaningful since most biological processes are also dynamic (18). To date, dynamic transcriptomic studies are expensive (e.g., numerous time points, replicates and conditions), computationally challenging and generate multi-facetted results that are difficult to interpret and understand (19). Additionally, the results of transcriptomics studies will produce a long list (e.g., potentially hundreds) of candidate genes that are not experimentally feasible. Additionally, since validation of candidate genes rely on the deletion or over-expression of individual genes, logical reduction of this candidate gene pool is critical to overcome experimental limitations (e.g., vector construction, transformation efficiencies, screening capacity) (20). Therefore, novel tools to produce a shortened, feasible list of genetic targets would facilitate an experimentally feasible reverse-engineering experience.

Outline of Dissertation

Presented in this dissertation are my efforts to understand critical processes and genes underlying robust phenotypes in the oleaginous yeast *Yarrowia lipolytica* for reverse engineering. All chapters begin by obtaining or engineering a strain demonstrating the desired robust phenotype which is used for comparative OMICs to enable elucidation of mechanisms and processes underlying robustness and/or genotypes for reverse engineering.

Chapter 1 This chapter describes how sterols are critical for IL-tolerance in *Y. lipolytica*. Using adaptive laboratory evolution, we generated a superior IL-tolerant strain.

By comparing this strain with wildtype *Y. lipolytica*, we elucidated sterols as a robust mechanism underlying IL-tolerance and successfully reverse engineered this robust phenotype by increasing sterol biosynthesis via overexpression of the sterol transcription factor.

Chapter 2 This chapter describes the elimination of thiamine auxotrophy in *Y. lipolytica*. Using comparative genomics with thiamine prototroph *S. cerevisiae* we identified the missing gene for thiamine synthesis. By discovering a novel thiamine-responsive promoter, we were able to control the expression of this missing gene to engineer the first thiamine prototrophic *Y. lipolytica* strain reported to date.

Chapter 3 This chapter describes a new methodology to predict genetic targets for reverse engineering complex phenotypes from temporal RNA-sequencing data. We demonstrated that genetic targets chosen by co-expression connectivity predicted the most IL-tolerant genes in comparison to traditional differential expression approaches. We used these genetic targets to universally reverse engineer high IL-tolerance in 3 *Y. lipolytica* isolates.

Chapter 4 This chapter describes key processes affecting xylose metabolism and controlling lipid accumulation or degradation in *Y. lipolytica*. By investigating the genotypic and phenotypic diversity of non-conventional *Y. lipolytica* isolates, we revealed metabolic processes and regulatory proteins correlated with superior xylose utilization and lipid accumulation or degradation from biomass hydrolysate.

Chapter 5 This chapter describes upcycling of plastic waste into higher-value chemicals using *Y. lipolytica*. Using short-term adaptation on catalytically depolymerized plastic waste as the sole carbon source, we demonstrated improved hydrophobic substrate utilization and bioconversion of plastic waste into citric acid and lipids. Additionally, we highlighted unforeseen technical bottlenecks including depolymerized plastic stability and unknown aspects of hydrophobic substrate metabolism.

References

1. A. Matsushika, H. Inoue, T. Kodaki, S. Sawayama, Ethanol production from xylose in engineered Saccharomyces cerevisiae strains: current state and perspectives. *Applied Microbiology and Biotechnology* **84**, 37-53 (2009).
2. J. R. Almeida *et al.*, Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. *Journal of Chemical Technology & Biotechnology* **82**, 340-349 (2007).
3. M. P. Almario, L. H. Reyes, K. C. Kao, Evolutionary engineering of Saccharomyces cerevisiae for enhanced tolerance to hydrolysates of lignocellulosic biomass. *Biotechnology and Bioengineering* **110**, 2616-2623 (2013).
4. S. A. Nicolaou, S. M. Gaida, E. T. Papoutsakis, A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metabolic Engineering* **12**, 307-331 (2010).
5. N. R. Sandoval, E. T. Papoutsakis, Engineering membrane and cell-wall programs for tolerance to toxic chemicals: beyond solo genes. *Current opinion in microbiology* **33**, 56-66 (2016).
6. H. J. Heipieper, G. Neumann, S. Cornelissen, F. Meinhardt, Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems. *Applied Microbiology and Biotechnology* **74**, 961-973 (2007).
7. L. R. Jarboe, P. Liu, L. A. Royce, Engineering inhibitor tolerance for the production of biorenewable fuels and chemicals. *Current Opinion in Chemical Engineering* **1**, 38-42 (2011).
8. Z. Tan, J. M. Yoon, D. R. Nielsen, J. V. Shanks, L. R. Jarboe, Membrane engineering via trans unsaturated fatty acids production improves *Escherichia coli* robustness and production of biorenewables. *Metabolic Engineering* **35**, 105-113 (2016).
9. Z. Tan *et al.*, Engineering *Escherichia coli* membrane phospholipid head distribution improves tolerance and production of biorenewables. *Metabolic Engineering* **44**, 1-12 (2017).

10. A. J. A. van Maris, B. Oud, J.-M. Daran, J. T. Pronk, Genome-wide analytical approaches for reverse metabolic engineering of industrially relevant phenotypes in yeast. *FEMS Yeast Research* **12**, 183-196 (2012).
11. F. H. Arnold, Design by Directed Evolution. *Accounts of Chemical Research* **31**, 125-131 (1998).
12. S. Vidan, M. Snyder, Large-scale mutagenesis: yeast genetics in the genome era. *Current Opinion in Biotechnology* **12**, 28-34 (2001).
13. J. D. Winkler, K. C. Kao, Recent advances in the evolutionary engineering of industrial biocatalysts. *Genomics* **104**, 406-411 (2014).
14. J. Anfelt, B. Hallström, J. Nielsen, M. Uhlén, E. P. Hudson, Using Transcriptomics To Improve Butanol Tolerance of Synechocystis sp. Strain PCC 6803. *Applied and Environmental Microbiology* **79**, 7419-7427 (2013).
15. A. Topaloğlu, C. Holyavkin, H. İ. Kısakesen, Z. P. Çakar, B. Hacısalihoğlu, Genomic and transcriptomic analysis of a coniferyl aldehyde-resistant Saccharomyces cerevisiae strain obtained by evolutionary engineering. 10.1093/femsyr/foz021 (2019).
16. D. Wang, D. Wu, X. Yang, J. Hong, Transcriptomic analysis of thermotolerant yeast Kluyveromyces marxianus in multiple inhibitors tolerance. *RSC Advances* **8**, 14177-14192 (2018).
17. R. J. Ihry *et al.*, p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine* **24**, 939-946 (2018).
18. Z. Bar-Joseph, A. Gitter, I. Simon, Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics* **13**, 552 (2012).
19. D. Spies, C. Ciaudo, Dynamics in Transcriptomics: Advancements in RNA-seq Time Course and Downstream Analysis. *Computational and Structural Biotechnology Journal* **13**, 469-477 (2015).

20. H. Alper, G. Stephanopoulos, Global transcription machinery engineering: A new approach for improving cellular phenotype. *Metabolic Engineering* **9**, 258-267 (2007).

CHAPTER I

Exceptional Solvent Tolerance in Yarrowia lipolytica Is Enhanced by Sterols

A version of this chapter was originally published by Caleb M. Walker:
Walker, Caleb, Seunghyun Ryu, and Cong T. Trinh. "Exceptional solvent tolerance in Yarrowia lipolytica is enhanced by sterols." Metabolic engineering 54 (2019): 83-95.

CTT conceived and supervised the study. CW, SR, and CTT designed the experiments, and analyzed the data. CW and SR performed the experiments. CW, SR, and CTT wrote and approved the manuscript.

Abstract

Green organic solvents such as ionic liquids (ILs) have versatile use but are inhibitory to microbes even at low concentrations of 0.5−1.0% (v/v) ILs. We discovered the oleaginous yeast Yarrowia lipolytica can grow in 10% (v/v) of 1-ethyl-3-methylimidazolium acetate ([EMIM][OAc]), which makes it more tolerant than any engineered microorganisms and naturally screened isolates. However, the underlying mechanism of IL tolerance in Y. lipolytica is not understood. Through adaptive laboratory evolution, in combination with physiological characterization and omics analysis, we shed light on the underlying mechanism of how Y. lipolytica restructures its membrane to tolerate different types of ILs at high levels up to 18% ILs. Specifically, we discovered that sterols play a key role for exceptional IL tolerance in Y. lipolytica.

Introduction

Robustness is an important beneficial phenotype for any organism to acquire for improved health or efficient biosynthesis of desirable molecules.[1,2] Recently, microbial biocatalysis in green organic solvents such as ionic liquids (ILs) has become attractive to produce high-value chemicals, biofuels, and bioproducts.[3-5] One key advantage is that ILs can effectively process complex and recalcitrant substrates such as lignocellulosic biomass for microbial fermentation.[6-8] Furthermore, these ILs can function as extractants for *in situ* separation of desirable molecules.[9,10] It is thus highly desirable to harness novel microbes for biocatalysis in organic solvents. However, ILs are inhibitory to microbes even at low IL concentrations of 0.5−1.0 % (v/v).[11,12] These solvents are inhibitory because they severely disrupt cell membranes and intracellular processes.[13-15]

Yarrowia lipolytica is a GRAS (generally-regarded-as-safe) oleaginous yeast of the phylum Ascomycota with established use in fundamental study of biological processes and in the biotechnology industry for production of biochemicals, biofuels, and enzymes.[16-18] Recently, *Y. lipolytica* has emerged as a potential bioenergy microbe for conversion of lignocellulosic biomass to biofuels because it is capable of accumulating high levels of neutral lipids [19-21] and functioning in harsh fermentation conditions such as wide pH 2−11 range [22] and inhibitory biomass hydrolysates [23]. Remarkably, wildtype *Y. lipolytica* exhibited robust growth in media containing 10% (v/v) [EMIM][OAc] while producing 92% maximum theoretical yield of alpha-ketoglutarate from IL-pretreated cellulose.[3] This endogenous IL-tolerant phenotype of *Y. lipolytica* surpassed most engineered organisms [24-27] and naturally screened isolates.[28,29] Currently, mechanisms of IL toxicity and superior tolerance in *Y. lipolytica* are not fully understood.

In this study, we aimed to illuminate the underlying processes of how *Y. lipolytica* exhibits robust growth in high (18% v/v) concentrations of ILs. By using the adaptive laboratory evolution in combination with physiological characterization and omics analysis, we elucidated how *Y. lipolytica* restructures its membrane to resist IL disruption and modulates sterol levels to improve IL tolerance. By disrupting and overexpressing the sterol biosynthesis pathway, we further validated that sterols play a key role for exceptional IL tolerance in *Y. lipolytica*.

Results

Generate robust Y. lipolytica mutants via adaptive laboratory evolution as a basis for elucidating IL-tolerant mechanism

***Y. lipolytica* has a novel endogenous metabolism conferring exceptional IL tolerance.** Since wildtype *Y. lipolytica* can grow in at least 10% (v/v) [EMIM][OAc],[3] we hypothesized that it has a novel endogenous metabolism conferring notable IL robustness. To test this, we performed adaptive laboratory evolution (ALE) to generate *Y. lipolytica* mutants with enhanced tolerance to high concentrations of the benchmark IL, [EMIM][OAc] (Figure 1-1, Step1). First, wildtype *Y. lipolytica* was grown in a medium

containing 5% (v/v) [EMIM][OAc] and transferred into a medium containing progressively increased concentrations of [EMIM][OAc], 8% and 10% (v/v). Remarkably, *Y. lipolytica* was able to grow in 5%, 8%, and 10% (v/v) [EMIM][OAc] with specific growth rates of 0.063 ± 0.005 1/h, 0.056 ± 0.033 1/h and 0.060 ± 0.004 1/h, respectively, without any significant growth inhibition (Figure 1-2A). When *Y. lipolytica* was transferred into a medium containing 12% (v/v) [EMIM][OAc], it initially exhibited growth inhibition with significantly reduced specific growth rate of 0.034 ± 0.001 1/h. However, after 16 generations in 12% (v/v) [EMIM][OAc], the specific growth rate was improved up to 0.080 ± 0.006 1/h and maintained for another 22 generations. Next, we increased the concentration of [EMIM][OAc] to 15% (v/v) and continued ALE. The first transfer from 12% (v/v) to 15% (v/v) reduced the specific growth rate by ~62% but cells recovered within 5 generations (0.078 ± 0.014 1/h) and the growth was maintained for an additional 23 generations. We further increased the concentration of [EMIM][OAc] to 18% (v/v) and continued serial transfers for another 106 generations. At the end of ALE (200 generations), we isolated the top performing *Y. lipolytica* mutant, YlCW001, growing in up to 18% (v/v) [EMIM][OAc] with a specific growth rate of 0.055 ± 0.006 1/h (Figure 1-1, Step 2).

Exceptional IL-tolerant phenotype of the evolved mutant YlCW001 is stable. To use YlCW001 for downstream characterization, we subjected it to irreversibility and stability tests (Figure 1-1, Step 3). For irreversibility testing, three biological replicates of wildtype and YlCW001 strains were grown in medium containing 8% (v/v) [EMIM][OAc] to lessen abrupt osmotic shock driven by IL. Mid-exponentially growing cells were then transferred into the medium containing glucose and 18% (v/v) [EMIM][OAc], and the cell growth was investigated. While growth of wildtype *Y. lipolytica* was completely inhibited, YlCW001 was able to grow with a specific growth rate of 0.059 ± 0.001 1/h (Figure 1-2B), comparable to that measured from ALE in 18% (v/v) [EMIM][OAc]. These results confirmed that the improved IL-tolerance for YlCW001 is irreversible.

Further, we investigated stability of YlCW001 by reviving the frozen glycerol stock in the liquid medium containing glucose and [EMIM][OAc] (Figure 1-1, Step 3). Like the irreversibility test, we evaluated cell growth of three biological replicates of YlCW001 in

18% (v/v) [EMIM][OAc] using glucose as a carbon source. The measured specific growth rate was similar to that from ALE and the irreversibility test (data not shown), proving that YlCW001 is a stable strain.

YlCW001 exhibits broad tolerance to a wide range of hydrophilic ILs. To test whether the IL-evolved strain YlCW001 exhibits broad IL-tolerant phenotypes, we investigated the following hydrophilic ILs: 1-ethyl-3-methylimidazolium acetate ([EMIM][OAc]), 1-ethyl-3-methylimidazolium chloride ([EMIM][Cl]), 1-ethyl-3-methylimidazolium bromide ([EMIM][Br]), 1-allyl-3-methylimidazolium chloride ([AMIM][Cl]), 1-butyl-3-methylimidazolium acetate ([BMIM][OAc]), 1-butyl-3-methylimidazolium chloride ([BMIM][Cl])), and 1-butyl-3-methylimidazolium bromide ([BMIM][Br]). We selected these ILs for testing because they can effectively solubilize various types of recalcitrant lignocellulosic biomass and are known to be very inhibitory to microbial growth.[30] Since wildtype growth was inhibited in 10% (v/v) and completely inhibited in 18% (v/v) [EMIM][OAc] (Figure 1-2C), we characterized these strains in two different concentrations of ILs, 0.6 M and 1.09 M (equivalent to 10% and 18% (v/v) of [EMIM][OAc], respectively).

Growth characterization in 0.6 M ILs shows that YlCW001 outperformed wildtype *Y. lipolytica* for all ILs except [EMIM][Cl], for which similar specific growth rates were observed (Figure 1-2D). Remarkably, YlCW001 growth was not affected by all tested ILs regardless of imidazolium alkyl chain length and conjoined anion, excluding [BMIM][OAc], which at 0.6 M was lethal to both wildtype and YlCW001 (Figure 1-2D). In 1.09 M ILs, wildtype *Y. lipolytica* exhibited substantial growth inhibition in [EMIM][Cl] and [EMIM][Br] and no growth in [AMIM][Cl], [BMIM][Cl], [EMIM][OAc], [BMIM][Br], and [BMIM][OAc] (Figure 1-2E). Strikingly, the evolved strain YlCW001 displayed robustness in all ILs except [BMIM][OAc] confirming broad tolerance to ILs (Figure 1-2E).

Inhibition of YlCW001 growth was detected in the following order: [BMIM][OAc] $\gg$ [BMIM][Br] $\approx$ [BMIM][Cl] $\approx$ [EMIM][OAc] $>$ [AMIM][Cl] $>$ [EMIM][Br] $\approx$ [EMIM][Cl]. While 0.6 M [BMIM][OAc] entirely inactivated growth of both wildtype and YlCW001, the mutant tolerated 0.3 M [BMIM][OAc] with a specific growth rate of 0.07 ± 0.02 1/h which remained lethal to wildtype (Figure 1-7).

Our result demonstrated that *Y. lipolytica* possesses novel endogenous metabolism conferring high IL tolerance by generating a novel, exceptionally robust *Y. lipolytica* mutant via ALE. The evolved strain YlCW001 is stable and exhibits broad tolerance towards the hydrophilic ILs in this study. The ability of YlCW001 to exhibit robust growth in up to 18% IL makes it the most IL-tolerant microorganism reported to date. This result provides a strong basis for elucidating the underlying mechanism of solvent toxicity and tolerance in *Y. lipolytica*.

Elucidate IL-responsive physiology and metabolism in Y. lipolytica strains

Cell membrane and morphology of the evolved mutant resist IL interference. Since ILs are known to disrupt the cell membrane,[31] we hypothesize that the mutant YlCW001 might have adapted its membrane structure to cope with inhibitory ILs. To test this hypothesis, we used scanning electron microscopy (SEM) to examine cell membranes and morphologies of both the wildtype and mutant responsive to 0% and 18% (v/v) [EMIM][OAc], 0.3M [BMIM][OAc], and 0.6M [BMIM][OAc].

As a positive control, we observed healthy morphologies for both the wildtype and mutant in non-IL media (Figure 1-3A, 1-3B). However, when exposed to 18% (v/v) [EMIM][OAc], the wildtype developed cavities, dents, and wrinkles along the cell surface, clearly demonstrating that cell membrane and/or cell wall components were severely damaged by the IL (Figure 1-3C). This phenotype is consistent with the complete growth inhibition of the wildtype observed at this high IL concentration (Figure 1-2B, 1-2C). In contrast, the evolved strain YlCW001 displayed barely any signs of membrane damage in 18% (v/v) [EMIM][OAc] (Figure 1-3D).

Likewise, when being exposed to a more toxic IL, [BMIM][OAc], the wildtype exhibited significant morphology deconstruction in both 0.3M and 0.6M [BMIM][OAc] (Figure 1-3E, 1-3G). Strikingly, YlCW001 displayed no significant morphological defects (Figure 1-3F, 1-3H) although growth was marginally inhibited in 0.3M and lethal in 0.6M [BMIM][OAc] (Figure 1-7 and Figure 1-2D). Additionally, the consequences of IL-membrane damage are not likely caused by cell death because we observed similar

morphological defects for the wildtype strain even in 8% (v/v) [EMIM][OAc] (sub-lethal concentration) after only 6 h of incubation (Figure 1-8).

Intracellular metabolism of *Y. lipolytica* was perturbed in presence of ILs. To demonstrate that intracellular metabolism of *Y. lipolytica* is also perturbed in IL, we performed untargeted metabolomics and lipidomics for both the wildtype and YlCW001 growing in media containing 0% and 8% (v/v) [EMIM][OAc]. Our results identified a total of 37 and 40 significantly perturbed pathways in wildtype and YlCW001, respectively, growing in 8% (v/v) [EMIM][OAc] as compared to the wildtype growing in no IL (Figure 1-3I and Table 1-S1). Among these pathways, 29 were found perturbed in both wildtype and YlCW001. These pathways are mostly comprised of amino acid synthesis/degradation, nucleosides and nucleotides synthesis/degradation, but also contain vitamin synthesis (coenzyme A, thiamine, folate, and biotin biosynthesis etc.), carbohydrate biosynthesis/degradation (GDP-mannose biosynthesis, gluconeogenesis, galactose degradation, etc.), respiration (TCA cycle, glyoxylate cycle, etc.), sterols biosynthesis, and some of the central metabolic pathways responsible for generating precursor metabolites and energy (glycolysis, pentose phosphate pathway, etc.)

Furthermore, we found that YlCW001 had a total of 19 perturbed metabolic processes in media without IL in comparison to the wildtype without IL (Figure 1-3I and Table 1-S1). Notably, 15 of these pathways were also perturbed in YlCW001 and wildtype growing in 8% (v/v) [EMIM][OAc] mostly enriched for amino acid biosynthesis/degradation but also included sterol biosynthesis. Overall, we identified significant perturbation of intracellular metabolism for both strains subjected to IL. While the bulk of these pathways are enriched for central carbon metabolism, our untargeted metabolomics and lipidomics identified sterols biosynthesis as the only lipid pathway significantly perturbed among all three biological conditions (wildtype 8%, YlCW001 8%, YlCW001 0%) in comparison to the wildtype without IL (Figure 1-3I).

Remodeling of cell membrane enhanced IL-tolerance in *Y. lipolytica*. The outer-surface of the eukaryotic yeast is a multifaceted, permeable barrier composed of a plasma membrane (i.e., glycerophospholipids, sterols, sphingolipids) and cell wall (i.e., chitin, glucan, mannoproteins) that together allow the cell to adapt to a variety of

environmental conditions to maintain cellular homeostasis.[32-35] Based on SEM images, untargeted metabolomics, and potential mechanisms of IL-toxicity,[13,36-39] we hypothesized that remodeling of cellular membrane and/or cell wall is one key IL stress-responsive process in *Y. lipolytica.* We aimed to identify critical membrane and cell wall components conferring exceptional IL-tolerance of both wildtype and YlCW001 strains by investigating IL-responsive glycerophospholipid, fatty acid, sterol, and chitin metabolism in the benchmark IL, [EMIM][OAc].

***Y. lipolytica* reduced chitin in the presence of ILs.** Chitin is one of the most insoluble biopolymers, even for ILs.[36,40] In *S. cerevisiae*, it has been reported that metabolism of chitin, a cell wall component known to influence membrane rigidity and elasticity, is increased upon cell wall integrity stresses.[41] Since *Y. lipolytica* contains a relatively high content of chitin relative to other yeasts (e.g., ~10-15% in *Y. lipolytica*; ~1-3% in *S. cerevisiae*[36]), we investigated how its membrane chitin is responsive to IL exposure. Counter-intuitively, we observed a ~2-fold reduction in chitin content for both strains upon IL-exposure (Figure 1-9). Of note, we were unable to detect any statistically significant differences in chitin levels between the wildtype and YlCW001 strains as they behaved similarly in 0% and 8% [EMIM][OAc]. While chitin may contribute to native *Y. lipolytica* IL-robustness, our results suggest chitin is not responsible for the enhanced IL-tolerance of YlCW001.

***Y. lipolytica* modulated glycerophospholipid composition in response to ILs.** Upon IL exposure, the backbone and headgroups of the *glycerophospholipids* are expected to interact with both cations and anions of ILs. To understand IL toxicity and tolerance, we next investigated IL-responsive glycerophospholipid metabolism by performing targeted lipidomics on individual headgroup species (i.e., phosphatidylcholine, PC; phosphatidylinositol, PI; phosphatidic acid, PA; phosphatidyl glycerol, PG; phosphatidylserine, PS; phosphatidylethanolamine, PE; and cardiolipins, CL) for both strains in 0% and 8% (v/v) [EMIM][OAc]. In non-IL media, we found that YlCW001 contained a larger amount of each glycerophospholipid species than the wildtype (Figure 1-4 A-G). In 8% (v/v) [EMIM][OAc], most components of glycerophospholipids, including PC, PA, PG, and CL, were upregulated in both strains, except that PI, PE and PS

exhibited different trends. The PE content of both strains in IL significantly decreased (relative to wildtype 0%) whereas PS levels remained unchanged (Figure 1-4E, 1-4F). Additionally, the PI content was only statistically increased by YlCW001 subjected to IL (Figure 1-4B). In non-IL media, YlCW001 significantly increased both PS and PE quantities (Figure 1-4E, 1-4F). Taken altogether, we found IL-responsive, upregulated glycerophospholipid production of all headgroup species except PS and PE. Interestingly, we also observed greater basal glycerophospholipid content in YlCW001 over the wildtype in 0% IL. These findings of membrane restructuring likely contribute to IL-toxicity resistance in *Y. lipolytica.*

ILs modulated the fatty acid composition of *Y. lipolytica*. Since fatty acids can modulate membrane fluidity,[42] we next analyzed the effect of ILs on the fatty acid profiles of wildtype and YlCW001 strains. The most striking differences were observed for C16:1 and C18:1 fatty acid moieties (Figure 1-4H). We found that both strains exposed to IL induced production of C16:1 (7 mol%, $p < 0.01$) fatty acids unlike the wildtype strain without IL, which produced none. In contrast, the wildtype without IL contained mostly C18:1 fatty acids (49 mol%) while all other biological conditions shifted to the di-unsaturated C18:2 moiety (29-36 mol%, $p < 0.02$). Interestingly, YlCW001 in media without IL behaved similarly to IL-exposed strains, with significant increases in C16 and C18:2 production. We did not observe a statistically significant difference between total saturated and unsaturated fatty acid moieties in both conditions. Taken together, IL-exposed cells (and YlCW001 0%) produced C16:1 fatty acids (non-existent in wildtype 0%) with statistically significant larger ratios of C16:C18 and (C18:2):(C18:1) moieties in comparison to the wildtype without IL (Figure 1-10). Shorter chain lengths with higher degrees of unsaturation in fatty acid moieties of the cell membrane are likely expected to increase membrane fluidity [35]. These results indicate that fatty acid metabolism is IL-modulated in *Y. lipolytica* and altered in YlCW001, even without IL.

YlCW001 increased sterols in the presence of ILs. We next investigated the functional role of sterols for IL tolerance in *Y. lipolytica* because i) the sterol biosynthesis pathway was perturbed in untargeted omics analysis, ii) sterols (e.g., cholesterol) can impede cation-insertion into the membrane,[37] and iii) sterols greatly influence membrane

fluidity.[43,44] In IL, we observed ~2 fold increase (p = 0.013) in ergosterol content of YlCW001 over the wildtype strain (Figure 1-5D). Counter-intuitively, we found the largest ergosterol concentrations in both strains without IL. We were unable to identify any other sterol pathway intermediates (e.g., squalene, lanosterol, etc.), in agreement with literature concluding ergosterol as the dominant sterol in yeast.[45] These results imply that IL affects sterol biosynthesis, and unlike the wildtype, YlCW001 adapted by enhancing membrane sterols in response to IL exposure.

Sterol biosynthesis is one key IL-responsive process to improve IL-tolerance in Y. lipolytica

We hypothesized that ergosterol content is a critical component of the membrane contributing to the enhanced IL-tolerance of YlCW001 since we observed a greater ergosterol content in YlCW001 than the wildtype upon exposure to IL (Figure 1-5D). To validate the key role of sterols, we investigated genetic and enzymatic details of how *Y. lipolytica* modulates the sterol biosynthesis pathway in response to IL.

YlCW001 highly upregulated the gene expression of the sterol biosynthesis pathway in the presence of ILs. We first aimed to analyze the mRNA expression levels of the sterol biosynthesis genes from the mid-exponentially growing wildtype and YlCW001 cells cultured in 0% and 8% [EMIM][OAc] (Figure 1-5A). Since the sterol transcription factors Upc and Ecm22 are known to play a key role in regulating the sterol biosynthesis in yeast (e.g., *S. cerevisiae* [46-48]), we also analyzed the transcriptional response of a functionally conserved homolog (SterTF, YALI0B15818g) of *Y. lipolytica* when cells were exposed to ILs. We found that 11 of the 14 sterol pathway genes were upregulated > 2-fold in IL-exposed YlCW001 as compared to the wildtype in 0% IL (Figure 1-5B). Significantly, seven of these genes, including Ster5 (YALI0B23298g), Ster6 (YALI0F11297g), Ster8 (YALI0B17644g), Ster10-1 (YALI0E32065g), Ster10-2 (YALI0B17204g), Ster11 (YALI0D20878g) and SterTF (YALI0B15818g) were upregulated > 4-fold in IL-exposed YlCW001 over the wildtype in 0% IL. As for the wildtype in 8% IL, we found only 1 of the 14 sterol pathway genes, Ster10-2 (YALI0B17204g), significantly upregulated against the wildtype without IL. Without IL, the

sterol genes in YlCW001 were relatively constant or marginally downregulated in comparison to the wildtype.

Enzymatic inhibition of the sterol biosynthesis pathway reduced IL-tolerance of Y. lipolytica. To confirm the contribution of sterols in IL-tolerance at the enzymatic level, we next treated the wildtype and YlCW001 strains with fluconazole, a commonly used anti-fungal drug that inhibits cytochrome P450 enzyme 14α-demethylase (Ster4, Figure 1-5A) critical for sterol biosynthesis.[49] We characterized growth of the wildtype and YlCW001 in media containing either 0% or 8% (v/v) [EMIM][OAc] with incremental concentrations of fluconazole (Figure 1-5C). We expected fluconazole to inhibit sterol biosynthesis and hence decrease IL-tolerance, specifically to a greater extent in the wildtype than in YlCW001. In media containing no IL, we found that fluconazole inhibited growth for both the wildtype and YlCW001 (IC_{50} = 25 μg/mL), (Figure 1-5C). The inhibition of fluconazole became more prominent in the presence of 8% (v/v) [EMIM][OAc]. Remarkably, YlCW001 could tolerate up to 25 μg/mL fluconazole while this concentration completely inhibited growth of the wildtype strain (Figure 1-5C).

Overexpression of the sterol pathway validated high solvent tolerance in the parent *Y. lipolytica* strain. To evaluate whether increasing expression of the sterol biosynthesis pathway could reverse engineer high solvent tolerance in the parent *Y. lipolytica* strain, we individually overexpressed the 14 annotated sterol biosynthesis and endogenous sterol transcription factor genes. We screened the individual, sterol-overexpressing constructs on plates containing incremental concentrations of [EMIM][OAc] (Figure 1-5E). We observed that some of the sterol genes showed marginal improvement (i.e., Ster3, Ster4, Ster6 and Ster12) for IL tolerance. Remarkably, overexpression of SterTF dramatically enhanced IL tolerance because it simultaneously activated upregulation of multiple genes governing the sterol biosynthesis to augment ergosterol production in IL (Figure 1-11). Taken altogether, the disruption and overexpression of the sterol biosynthesis pathway validated that sterols are critical membrane components responsible for exceptional IL-tolerance in *Y. lipolytica*.

Discussion

Microbial biocatalysis in ILs is novel for biosynthesis of fuels and chemicals. For instance, the imidazolium-based ILs (e.g., [EMIM][OAc]) are effective for reducing lignocellulosic biomass recalcitrance [50] to be used for downstream fermentation but greatly inhibit microbial growth even at low concentrations.[36,51] This incompatibility presents a significant barrier for novel microbial biocatalysis in ILs. While mechanisms of IL-toxicity have been proposed,[14,25,27] and IL tolerance has been mediated in part by multidrug efflux pumps,[25,26,52] the complete picture is unclear of how cells resist ILs and whether these cells can adapt to achieve IL-tolerance for industrial compatibility. To illuminate the mechanisms of IL toxicity and enhanced tolerance (Figure 1-6), we characterized naturally IL-tolerant *Y. lipolytica* and its superior evolved mutant, YlCW001, generated by ALE (Figure 1-2A). As compared to both naturally screened isolates and engineered strains reported to date, YlCW001 is the most IL-tolerant microorganism capable of growing in a variety of ILs at high levels up to 18% IL.

Imidazolium-based ILs inhibit the cell by inserting their alkyl chains into the hydrophobic core of the plasma membrane (Figure 1-6B, 1-6E, 1-6C, and 1-6F).[31,37] ILs with longer alkyl chains become more lipophilic [13,53] and cause greater disruption of the membrane as supported by growth rates in various ILs ([BMIM] ≫ [AMIM] ~ [EMIM]) (Figure 1-2D, 1-2E). The conjoint anion likely re-associates with the cation imbedded into the membrane causing greater membrane disturbance.[37] The tendency of an anion to interact with the cation intensifies with increasing basicity, which increases IL toxicity as demonstrated by reduced growth rates in various ILs ([OAc] ≫ [Br] ~ [Cl]) (Figure 1-2D, 1-2E).

Consequences of IL membrane disruption increase membrane fluidity (e.g., ionic surfactant) [15,38] and impose lateral pressure on the membrane [39,54] as evidenced by cavities, dents, and wrinkles in SEM images (Figure 1-3C, 1-3E, 1-3G) and dramatic remodeling of lipids in *Y. lipolytica* (Figure 1-4). In addition, harmful interactions between the IL and membrane result in a cascade of detrimental effects on cellular processes including DNA damage, enzyme inactivation, and protein degradation,[50,55-57] as observed

by reduced sterol and chitin contents (Figure 1-5D and Figure 1-9) and perturbation of intracellular metabolism (Figure 1-3I) of IL-exposed wildtype and YlCW001 strains.

Wildtype and YlCW001 strains combatted IL toxicity in part by rewiring membrane compositions to reduce membrane permeability and bilayer buckling pressure (imposed by ILs).[58,59] Both strains overproduced all glycerophospholipid species except PS and significantly reduced PE (Figure 1-4F), which is vulnerable to lateral pressure.[60,61] In contrast to the wildtype, YlCW001 is more robust because it adapts to produce more sterols, e.g., ergosterol upon IL-exposure (Figure 1-5D), that function to maintain membrane fluidity and stability.[43] This novel phenotype is evidenced by a significant upregulation of sterol biosynthesis genes (Figure 1-5B), improved enzymatic-tolerance to steroid-inhibiting drug, fluconazole[49,62] (Figure 1-5C), and overexpression of the sterol biosynthesis and transcriptional factor genes (Figure 1-5E). The result is also consistent with molecular simulations demonstrating sterols impede IL cations from inserting into artificial membranes.[63-65] Interestingly, *S. cerevisiae* strains evolved in chemical (i.e., butanol) and physical (i.e., heat) stresses have shown increased ergosterol levels or altered sterol compositions that likely function to maintain membrane homeostasis.[44,66]

Taken altogether, ILs inhibit cell growth by fluidizing the membrane and inflicting lateral pressures that destroy cellular homeostasis (Figure 1-6B, 1-6C). Our research provides strong evidence of how intracellular processes of *Y. lipolytica* are rewired to remodel cell membranes upon IL-exposure. Comprehensive metabolic, transcriptomic, and enzymatic analyses provide strong evidence that sterols (i.e., ergosterol) are critical membrane components conferring IL-tolerance in *Y. lipolytica* and enhanced IL-robustness in YlCW001, functioning to impede cation insertion and maintain membrane homeostasis [64] (Figure 1-6E, 1-6F). This study provides a better understanding of IL toxicity mechanism and superior tolerance of *Y. lipolytica,* which can be harnessed for improvement of organism health and industrial biocatalysis in organic solvents.

Methods

Strains and plasmids

Strains. *Escherichia coli* TOP10 strain was used for molecular cloning. *Yarrowia lipolytica* (YlSR001, ATCC MYA-2613), a thiamine, leucine, and uracil auxotroph, was used as the parent strain. The evolved strain YlCW001 was isolated after 200 generations in gradually increased concentrations of [EMIM][OAc] up to 18% (v/v). By transforming the constructed plasmids into YlSR001 via electroporation,[67] we created the *Y. lipolytica* strains YlSR172-YlSR202 that overexpress the individual sterol biosynthesis (Ster1-13) and transcription factor (SterTF) genes (Table 1-S4).

Plasmids. A set of 13 sterol biosynthesis genes and 1 sterol transcription factor gene of *Y. lipolytica* are listed in Table 1-S2. The plasmid pSR008 containing a constitutive *Y. lipolytica* TEF promoter ($P_{TEF(406)}$), a CYC1 terminator (cyc1t) and an uracil selection marker was used as the backbone plasmid.[68] The plasmids pSR166, pSR170, pSR171, and pSR172 carrying the genes Ster1, Ster5, Ster6, and Ster7, respectively were constructed by Gibson assembly [69] of two DNA fragments: i) each of the sterol genes amplified from *Y. lipolytica* genomic DNA using the primers listed in Table 1-S3 and ii) the plasmid backbone amplified from pSR008 using the primers cyc1t_Fwd/pTEF_Rev. The plasmids pSR167-169 and pSR173-179 were generated by inserting the genes Ster2-Ster4 and Ster8-Ster13, respectively, into pSR008 through the AvrII and AscI digestion and T4 DNA ligase treatment. Specifically, each sterol gene was amplified from *Y. lipolytica* genomic DNA using the primers listed in Table 1-S3. The backbone pSR008 was amplified using the primers AvrII_cyc1t_Fwd/AscI_pTEF_Rev. Both the sterol genes and pSR008 backbone were digested by AvrII and AscI at 37°C for 3 h, then ligated by T4 DNA ligase at room temperature for 30 min. Likewise, to construct pSR180, SterTF was amplified from *Y. lipolytica* genomic DNA using the primers $SterTF_{YL}$_Fwd/$SterTF_{YL}$_Rev and then ligated with pSR008 through the AvrII and AscI restriction sites. The complete list of constructed plasmids and strains are presented in Table 1-S4.

Medium and culturing conditions

Growth medium. ALE, irreversibility testing, and broad IL tolerance studies were conducted in defined media containing 6.7 g/L of yeast nitrogen base without amino acids (cat# Y0626, Sigma-Aldrich, MO, USA), 10 g/L of glucose, 100 mg/L of ampicillin, 50 mg/L of kanamycin, 30 mg/L of chloramphenicol, and various concentrations of ILs. Leucine (cat# 172130250, Acros Organics, CA, USA) and uracil (cat# 157301000, Acros Organics, CA, USA) were added to the media at concentrations of 190 mg/L and 20 mg/L, respectively.

For all other growth studies, 380 mg/L of leucine and 76 mg/L uracil were used. All ILs, including 1-ethyl-3-methylimidazolium acetate [EMIM][OAc] (>95 % purity), 1-ethyl-3-methylimidazolium chloride [EMIM][Cl] (>98% purity), 1-ethyl-3-methylimidazolium bromide [EMIM][Br] (99% purity), 1-allyl-3-methylimidazolium chloride [AMIM][Cl] (>98% purity), 1-butyl-3-methylimidazolium acetate [BMIM][OAc] (>98% purity), 1-butyl-3-methylimidazolium chloride [BMIM][Cl] (99% purity), and 1-butyl-3-methylimidazolium bromide [BMIM][Br] (99% purity), were purchased from the Ionic Liquids Technologies Inc. (IoLiTec, AL, USA). Medium pH was at 5.5 without IL and at 6-7 with ILs. Unless specifically mentioned, all experiments were performed with biological triplicates.

Adaptive laboratory evolution. ALE experiment was performed by serial dilution of *Y. lipolytica* in sequentially increasing concentrations of [EMIM][OAc] in 6-well plates with 3 mL working volume using an incubating microplate shaker (cat# 02-217-757, Fisher Scientific, PA, USA) at 28°C and 350 rpm with adhesive, breathable seals to prevent cross contamination (cat# 50-550-304, Fisher Scientific, PA, USA). For each serial dilution, the top performing triplicate was transferred during mid-exponential growth phase into fresh medium at an initial optical density (OD) at 600 nm of 0.2. Increasing concentrations of [EMIM][OAc] were selected to achieve a specific growth rate ≥ 0.02 1/h. The maximum specific growth rates of all three technical replicates were calculated for each serial dilution using a minimum of three time points per replicate. After 200 generations of ALE, the top performing replicate culture was spread onto a petri plate containing defined medium with 10 g/L glucose and 20 g/L agar. The plate was incubated at 28^{o}C for 36-48 h. Single colonies were isolated and individually streaked onto fresh

petri plates. This process was repeated once more to ensure isolation of purified colonies. Purified colonies (from three iterations of plate dilutions) were individually tested in the same growth conditions and [EMIM][OAc] concentration at which the evolved strain was originally collected to determine irreversibility. Individual cultures (from purified colonies) were collected and stored in glycerol at -80°C before streaking onto fresh petri plates, repeating three plate isolations, and retesting the irreversibility of the purified colonies to determine stability of YlCW001.

Plate screening for enhanced IL tolerance. To determine whether overexpression of the sterol biosynthesis and transcription factor genes enhance IL-tolerance, the strains YlSR001, YlCW001, and YlSR001 carrying Ster1-13 and SterTF were grown in defined medium without uracil in 15-mL culture tubes until mid-exponential growth phase. Cells were reconstituted in sterile water to achieve starting OD of 2 and further diluted 10x and 100x. Two microliters of each dilution were spotted on fresh petri dishes containing defined medium without uracil, 20 g/L agar and 0%, 4%, 8%, 10%, and 12% (v/v) [EMIM][OAc] and incubated at 28°C. Plates containing 0% and 4% (v/v) [EMIM][OAc] were incubated for 48 h and plates containing 8%, 10% and 12% (v/v) [EMIM][OAc] for 96 h. Plate screening was repeated three times for each construct.

Analytical methods

Metabolomics. Three biological replicates of the wildtype and YlCW001 were grown in media containing 0% or 8% (v/v) [EMIM][OAc] and collected at the late exponential growth phase for metabolomic analysis. Samples were immediately quenched in liquid nitrogen and stored at -80°C prior to extraction. Metabolites were extracted from a minimum of 1 x10^7 cells in 400µL of the extraction solvent by incubating at -20°C for 20 min.[70] The extraction solvent consists of 40:40:20 (vol) methanol:acetonitrile:water containing 0.1M formic acid. The soluble fraction was separated by centrifugation at 13,700 x g, 4°C for 5 minutes. To ensure complete extraction of cellular metabolites, we repeated extraction 3 times per sample. A total of 1.2 mL of solvent-soluble metabolite samples were subjected to drying under a stream of nitrogen at 4°C overnight to evaporate solvent. Lyophilized metabolites were

reconstituted in 300µL of sterile water and analyzed by liquid chromatography mass spectrometer (LC-MS).

Metabolites were analyzed with an Exactive Plus orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) equipped with Synergi 2.5□m Hydro-RP 100 (100 x 2.00 mm, Phenomenex, Torrance, CA, USA) set to 25°C. The LC-MS method analyzed in full scan negative ionization mode with an electrospray ionization source as previously described.[71]

Lipidomics. Late exponentially growing wildtype and YlCW001 cells cultured in 0% and 8% (v/v) [EMIM][OAc] were used for lipidomic study. Samples were immediately quenched in liquid nitrogen and stored at -80°C. After thawing on ice, samples were centrifuged for 3 minutes at max speed, 4°C before removing supernatant. Next, cell pellets were re-suspended in 800uL of 0.1N hydrochloric acid:methanol 1:1 (vol) with 400uL of chloroform and disrupted with glass beads using a mini bead beater for 5-minute intervals until > 95% of cells were visually disrupted. Disrupted cells were vortexed and centrifuged at 4°C for 5 minutes before extracting the organic phase into glass vials. Finally, samples were dried under a stream of nitrogen overnight at 4°C and reconstituted in 300uL of 9:1 (vol) methanol:chloroform before transferring into auto-sampler vials. Lipid extracts were analyzed in positive and negative ionization modes with an Exactive Plus orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an electrospray ionization probe and a Kinetex HILIC column (150 mm x 2.1 mm, 2.5 µm) (Phenomenex, Torrance, CA, USA) as previously described [72] except that lipid features were verified with external standards instead of fragments.

Fatty acid quantification. Three biological replicates of wildtype and YlCW001 cells were grown in 0% and 8% (v/v) [EMIM][OAc] until mid-late exponential phase and cell pellets were stored at -20°C. The equivalent cell mass of 2 OD was washed once with 0.05 M sodium phosphate solution and incubated in 2:1 (v/v) chloroform:methanol solution overnight at 4°C. 200 µL of chloroform was extracted and mild methanolysis was performed as previously described.[73] Briefly, 1.5 mL of methanol, 0.3 mL of 8% methanolic HCL solution and 50 µL of 2 mg/mL pentadecanoic acid was added to ensure complete transesterification and incubated overnight at 55°C. After cooling to room

temperature, 1mL of hexane containing 0.005 mg/mL pentadecanoic acid ethyl ester as internal standard and 1mL of water was added prior to extracting 250 µL of hexane for GCMS detection of fatty acid methyl esters.

Sterol quantification. Three biological replicates of wildtype and YlCW001 cells were grown in 0% and 8% (v/v) [EMIM][OAc] until mid-late exponential phase and cell pellets were stored at -20°C. Sterol quantification was performed with the equivalent cell mass of 5 OD as previously described [74] using a HP-5MS 30 m x 0.25 mm i.d. x 0.25 µm column (Agilent Technologies, USA) for separation of sterols.

Chitin determination. The relative quantity of chitin was analyzed for the wildtype and YlCW001 late-exponentially growing cells in medium containing 0%, 2%, 5%, 8% and 10% (v/v) [EMIM][OAc]. Cell pellets were washed twice with water before resuspending 1 OD of cell mass in 1 mL of water containing 50 µg/mL calcofluor white (CFW) (cat #18909, Sigma-Aldrich), which binds specifically to chitin.[75,76] Samples were incubated at room temperature for 15 minutes at 650 rpms in a thermomixer (cat #5382000023, Eppendorf). Next, stained cells were washed twice with water to remove excess CFW prior to fluorescence (excitation: 360/40nm, emission: 460/40nm) and absorbance (OD) measurements. Results were calculated by normalizing fluorescence intensity by respective sample OD values. Chitin determination experiments were conducted at least twice for each biological condition in technical replicates per experiment (wildtype 0%, n = 18; wildtype 2%, n = 6; wildtype 5%, n = 15; wildtype 8%, n = 12; YlCW001 0%, n = 25; YlCW001 2%, n = 6; YlCW001 5%, n = 15; YlCW001 8%, n = 29; YlCW001 10%, n = 6).

Scanning electron microscope (SEM). Wildtype and YlCW001 cells were inoculated at 1 OD in 0%, and 18% [EMIM][OAc] and 0.3M and 0.6M [BMIM][OAc] in 6-well microtiter plates using an incubating microplate shaker at 28°C and 350 rpm. After 24 hours, cells were collected, immediately washed once with water and incubated in 2% glutaraldehyde containing 10 mM sodium phosphate buffer overnight at 4°C. Fixed samples were washed 3 times with water before post-fixing in 2% osmium tetroxide for 1 hour at room temperature. The cell pellets were placed on silicon chips and dehydrated with successive ethanol baths (50%, 60%, 70%, 80%, 90%, and 100%) for 15 minutes

each. Finally, the dehydrated samples were dried using critical point drying in carbon dioxide at 1100 psi and 32°C before SEM imaging with Zeiss Argula using SEM2 detector.

RNA-sequencing. To investigate the transcriptional responses of the sterol biosynthesis and transcription factor genes to IL exposure in *Y. lipolytica*, three biological replicates of YlSR001 and YlCW001 growing in 0% and 8% (v/v) [EMIM][OAc] media were harvested at the mid-exponential growth phase, immediately quenched in liquid nitrogen, and stored at -80°C for downstream processing. Total RNA was purified using the Qiagen RNeasy mini kit (cat #74104, Qiagen, CA, USA) and submitted for Illumina sequencing at the Joint Genome Institute (JGI).

Fluconazole treatment to investigate importance of sterol biosynthesis for enhanced IL tolerance. To investigate correlation of sterol biosynthesis with enhancement of IL tolerance, we compared growth of the wildtype and YlCW001 in media containing 0% or 8% (v/v) [EMIM][OAc] and various concentrations of fluconazole (cat# TCF0677, VWR) which inhibits fungal cytochrome P450 enzyme 14α-demethylase (Ster4, Figure 1-5A) required for sterol biosynthesis.[49] First, we cultured the wildtype and YlCW001 in media without IL until the mid-exponential phase. Then, cells were washed by water and resuspended in the fresh media containing 0, 10, 25, 50, 100, and 250 µg/mL fluconazole. Finally, the same conditions were tested with addition of 8% (v/v) [EMIM][OAc] to observe the requirement of sterol incorporation into the plasma membrane for IL tolerance. Results were obtained by calculating maximum specific growth rates for each biological condition. Sterol validation by fluconazole treatment experiments were conducted twice for each biological condition with sacrificial, technical replicates using 96-well microtiter plates and Duetz sandwich cover (cat # SMCR1296, Kuhner) incubated at 28°C and 400 rpms.

Bioinformatics and data analysis

Untargeted LCMS analysis. Metabolomic and lipidomic raw files created by Xcalibur were imported into XCMS online and analyzed in pairwise jobs against the control data set, wildtype in 0% IL.[77] XCMS resulting metabolic features were exported and features with intensity fold changes < 2 were removed. These pairwise sets of 'perturbed' metabolic features were analyzed with metaboanalyst [78] using 'MS peaks to

pathways' tool with significant feature *P*-value cutoff = 0.05. This tool uses mummichog which algorithmically utilizes known metabolic pathways and networks to predict metabolites and pathways without prior identification of metabolites.[79] The resulting pathway files from 'perturbed' pairwise feature sets were exported and pathways with less than two significant features (i.e., metabolite features with *P*-values < 0.05) were removed. Next, we filtered the pathways by defining a new parameter, pathway significance factor (psf) (equation 1), to account for i) total number of metabolites in the pathway (p_{size}), ii) total identified metabolite features ($p_{features}$), and iii) total number of significant metabolite features ($p_{significant}$) identified for each pathway.

$$\mathrm{psf} = \frac{1}{\mathrm{psf_{max}}} \times \frac{\mathrm{p_{features}}}{\mathrm{p_{size}}} \times \mathrm{p_{significant}^{0.5}} \quad (1)$$

In our analysis, we chose a psf cutoff value of 0.58 to illustrate the top 15% most significantly perturbed pathways identified from our untargeted LCMS analysis.

Targeted LCMS analysis. Lipidomics raw data files created by Xcalibur were converted to open source mzML format using the ProteoWizard software.[80] MAVEN software (Princeton University) was applied to performed retention time correction for each sample and used to manually select known lipids based on retention time and mass.[81,82] Glycerophospholipid headgroup species were analyzed individually and extracted signal intensities were corrected by cell number that was determined by a Guava flow cytometer (easyCyte 6HT 2L, EMD Millipore Corp., MA, USA) and a Guava ViaCount kit (cat #4000-0041, EMD Millipore Corp., MA, USA). The corrected intensities for each head group class were summed and normalized relative to wildtype in non-IL media to visualize changes for each glycerophospholipid on a macro scale.

RNAseq analysis. Filtered RNA-sequencing reads were imported and analyzed within the CLC genomics workbench version 11.0.1 (https://www.qiagenbioinformatics.com/). To calculate the differential gene expression, wildtype in 0% IL was used as a reference.

Identification of the putative sterol transcription factor (SterTF) in Y. lipolytica. To identify sterol regulatory elements in *Y. lipolytica*, we first performed BlastP using the well characterized sterol transcription factors (TFs), Upc2 and Ecm22 of

Saccharomyces cerevisiae [46] as the references. Upon identifying a closely related *Y. lipolytica* endogenous homolog (SterTF, YALI0B15818g) with respect to Upc2 (Quary cover 37%; E value $4x10^{-83}$; Identity 48%) and Ecm22 (Quary cover 43%; E value $1x10^{-84}$; Identity 48%), we further performed the transcription factor affinity prediction (TRAP) [83] for SterTF against the 1050 nucleotide sequences upstream of each steroid biosynthesis gene to further validate the regulatory elements of SterTF *in silico* (Figure 1-12).

Statistics. Statistical analysis was performed with SigmaPlot v.14 software using one-way analysis of variance (ANOVA) with Holm-Sidak correction or Student's t-test between biologically relevant conditions where noted.

Acknowledgements

We would like to thank Dr. Stephen Dearth (Biological and Small Molecule Mass Spectrometry Core, UTK) for assistance in metabolomics and lipidomics experiments, Dr. John Dunlap (Advanced Microscopy Imaging Center, UTK) for assistance in SEM experiments, and Mr. Piet Jones (Bredesen Center, UTK) for bioinformatic analysis. The sequencing project was supported by The BioEnergy Science Center (BESC) and The Center for Bioenergy Innovation (CBI), the U.S. Department of Energy (DOE) Bioenergy Research Centers funded by the Office of Biological and Environmental Research in the DOE Office of Science (DOE BER). The RNAseq work conducted by the U.S. DOE Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231. The authors would like to acknowledge financial support from the National Science Foundation (NSF #1511881) and the DOE BER Genomic Science Program (DE-SC0019412).

References

1. Nicolaou, S.A., Gaida, S.M. & Papoutsakis, E.T. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metabolic Engineering* **12**, 307-331 (2010).
2. Couling, D.J., Bernot, R.J., Docherty, K.M., Dixon, J.K. & Maginn, E.J. Assessing the factors responsible for ionic liquid toxicity to aquatic organisms via quantitative structure–property relationship modeling. *Green Chemistry* **8**, 82-90 (2006).
3. Ryu, S., Labbé, N. & Trinh, C.T. Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α-ketoglutaric acid by *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology* **99**, 4237-4244 (2015).
4. van Rantwijk, F. & Sheldon, R.A. Biocatalysis in Ionic Liquids. *Chemical Reviews* **107**, 2757-2785 (2007).
5. Bokinsky, G. et al. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proceedings of the National Academy of Sciences* **108**, 19949-19954 (2011).
6. Li, C. et al. Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. *Bioresource Technology* **101**, 4900-4906 (2010).
7. Yoo, C.G., Pu, Y. & Ragauskas, A.J. Ionic liquids: Promising green solvents for lignocellulosic biomass utilization. *Current Opinion in Green and Sustainable Chemistry* **5**, 5-11 (2017).
8. Socha, A.M. et al. Efficient biomass pretreatment using ionic liquids derived from lignin and hemicellulose. *Proc Natl Acad Sci U S A* **111**, E3587-95 (2014).
9. Cull, S.G., Holbrey, J.D., Vargas-Mora, V., Seddon, K.R. & Lye, G.J. Room-temperature ionic liquids as replacements for organic solvents in multiphase bioprocess operations. *Biotechnology and Bioengineering* **69**, 227-233 (2000).

10. Ventura, S.P.M. et al. Ionic-Liquid-Mediated Extraction and Separation Processes for Bioactive Compounds: Past, Present, and Future Trends. *Chemical Reviews* **117**, 6984-7052 (2017).
11. Wu, Y.-W. et al. Ionic Liquids Impact the Bioenergy Feedstock-Degrading Microbiome and Transcription of Enzymes Relevant to Polysaccharide Hydrolysis. *mSystems* **1**(2016).
12. Ouellet, M. et al. Impact of ionic liquid pretreated plant biomass on *Saccharomyces cerevisiae* growth and biofuel production. *Green Chemistry* **13**, 2743-2749 (2011).
13. Yu, M. et al. Effects of the 1-alkyl-3-methylimidazolium bromide ionic liquids on the antioxidant defense system of *Daphnia magna*. *Ecotoxicology and Environmental Safety* **72**, 1798-1804 (2009).
14. Docherty, K.M. & Kulpa, J.C.F. Toxicity and antimicrobial activity of imidazolium and pyridinium ionic liquids. *Green Chemistry* **7**, 185-189 (2005).
15. Liu, L.-P. et al. Mechanistic insights into the effect of imidazolium ionic liquid on lipid production by *Geotrichum fermentans*. *Biotechnology for Biofuels* **9**, 266 (2016).
16. Blazeck, J. et al. Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. *Nature communications* **5**(2014).
17. Xue, Z. et al. Production of omega-3 eicosapentaenoic acid by metabolic engineering of Yarrowia lipolytica. *Nat Biotech* **31**, 734-740 (2013).
18. Nicaud, J.-M. Yarrowia lipolytica. *Yeast* **29**, 409-418 (2012).
19. Dulermo, T. & Nicaud, J.-M. Involvement of the G3P shuttle and β-oxidation pathway in the control of TAG synthesis and lipid accumulation in Yarrowia lipolytica. *Metabolic engineering* **13**, 482-491 (2011).
20. Liu, L., Pan, A., Spofford, C., Zhou, N. & Alper, H.S. An evolutionary metabolic engineering approach for enhancing lipogenesis in Yarrowia lipolytica. *Metabolic Engineering* **29**, 36-45 (2015).
21. Xu, P., Qiao, K., Ahn, W.S. & Stephanopoulos, G. Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proceedings of the National Academy of Sciences* **113**, 10848-10853 (2016).

22. Epova, E.G., Marina; Kovalyov, Leonid; Isakova, Elena; Deryabina, Yulia; Belyakova, Alla; Zylkova, Marina and Alexei Shevelev. Identification of Proteins Involved in pH Adaptation in Extremophile Yeast Yarrowia lipolytica. in *Proteomic Applications in Biology* (InTech, 2012).
23. Tsigie, Y.A., Wang, C.-Y., Truong, C.-T. & Ju, Y.-H. Lipid production from *Yarrowia lipolytica* Po1g grown in sugarcane bagasse hydrolysate. *Bioresource Technology* **102**, 9216-9222 (2011).
24. Frederix, M. et al. Development of an *E. coli* strain for one-pot biofuel production from ionic liquid pretreated cellulose and switchgrass. *Green Chemistry* **18**, 4189-4197 (2016).
25. Ruegg, T.L. et al. An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. *Nature Communications* **5**, 3490 (2014).
26. Mohamed, E.T. et al. Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. *Microbial Cell Factories* **16**, 204 (2017).
27. Dickinson, Q. et al. Mechanism of imidazolium ionic liquids toxicity in *Saccharomyces cerevisiae* and rational engineering of a tolerant, xylose-fermenting strain. *Microbial Cell Factories* **15**, 1-13 (2016).
28. Sitepu, I.R. et al. Yeast tolerance to the ionic liquid 1-ethyl-3-methylimidazolium acetate. *FEMS Yeast Research* **14**, 1286-1294 (2014).
29. Khudyakov, J.I. et al. Global transcriptome response to ionic liquid by a tropical rain forest soil bacterium, *Enterobacter lignolyticus*. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2173-E2182 (2012).
30. Heckenbach, M.E., Romero, F.N., Green, M.D. & Halden, R.U. Meta-analysis of ionic liquid literature and toxicology. *Chemosphere* **150**, 266-274 (2016).
31. Lim, G.S., Jaenicke, S. & Klahn, M. How the spontaneous insertion of amphiphilic imidazolium-based cations changes biological membranes: a molecular simulation study. *Phys Chem Chem Phys* **17**, 29171-83 (2015).
32. Lindberg, L., Santos, A.X.S., Riezman, H., Olsson, L. & Bettiga, M. Lipidomic Profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* Reveals

Critical Changes in Lipid Composition in Response to Acetic Acid Stress. *PLOS ONE* **8**, e73936 (2013).

33. Santos, A.X.S. & Riezman, H. Yeast as a model system for studying lipid homeostasis and function. *FEBS Letters* **586**, 2858-2867 (2012).
34. Lipke, P.N. & Ovalle, R. Cell Wall Architecture in Yeast: New Structure and New Challenges. *Journal of Bacteriology* **180**, 3735-3740 (1998).
35. Ingram, L.O. Adaptation of membrane lipids to alcohols. *Journal of Bacteriology* **125**, 670-678 (1976).
36. Santos, A.G., Ribeiro, B.D., Alviano, D.S. & Coelho, M.A.Z. Toxicity of ionic liquids toward microorganisms interesting to the food industry. *RSC Advances* **4**, 37157-37163 (2014).
37. Yoo, B., Shah, J.K., Zhu, Y. & Maginn, E.J. Amphiphilic interactions of ionic liquids with lipid biomembranes: a molecular simulation study. *Soft Matter* **10**, 8641-8651 (2014).
38. Lee, H. Effects of imidazolium-based ionic surfactants on the size and dynamics of phosphatidylcholine bilayers with saturated and unsaturated chains. *Journal of Molecular Graphics and Modelling* **60**, 162-168 (2015).
39. Yoo, B., Zhu, Y. & Maginn, E.J. Molecular Mechanism of Ionic-Liquid-Induced Membrane Disruption: Morphological Changes to Bilayers, Multilayers, and Vesicles. *Langmuir* **32**, 5403-5411 (2016).
40. Wu, Y., Sasaki, T., Irie, S. & Sakurai, K. A novel biomass-ionic liquid platform for the utilization of native chitin. *Polymer* **49**, 2321-2327 (2008).
41. Lesage, G. & Bussey, H. Cell Wall Assembly in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* **70**, 317-343 (2006).
42. Tan, Z., Yoon, J.M., Nielsen, D.R., Shanks, J.V. & Jarboe, L.R. Membrane engineering via trans unsaturated fatty acids production improves *Escherichia coli* robustness and production of biorenewables. *Metabolic Engineering* **35**, 105-113 (2016).
43. Kohlwein, S.D. Analyzing and Understanding Lipids of Yeast: A Challenging Endeavor. *Cold Spring Harbor Protocols* **2017**, pdb.top078956 (2017).

44. Caspeta, L. et al. Altered sterol composition renders yeast thermotolerant. *Science* **346**, 75-78 (2014).
45. Vanegas, Juan M., Contreras, Maria F., Faller, R. & Longo, Marjorie L. Role of Unsaturated Lipid and Ergosterol in Ethanol Tolerance of Model Yeast Biomembranes. *Biophysical Journal* **102**, 507-516 (2012).
46. Vik, A. & Rine, J. Upc2p and Ecm22p, dual regulators of sterol biosynthesis in Saccharomyces cerevisiae. *Molecular and cellular biology* **21**, 6395-6405 (2001).
47. Yang, H. et al. Structural mechanism of ergosterol regulation by fungal sterol transcription factor Upc2. *Nature Communications* **6**, 6129 (2015).
48. Hoot, S.J., Smith, A.R., Brown, R.P. & White, T.C. An A643V Amino Acid Substitution in Upc2p Contributes to Azole Resistance in Well-Characterized Clinical Isolates of Candida albicans. *Antimicrobial Agents and Chemotherapy* **55**, 940-942 (2011).
49. de Souza, W. & Rodrigues, J.C.F. Sterol Biosynthesis Pathway as Target for Anti-trypanosomatid Drugs. *Interdisciplinary Perspectives on Infectious Diseases* **2009**, 19 (2009).
50. Fan, L.-L., Li, H.-J. & Chen, Q.-H. Applications and Mechanisms of Ionic Liquids in Whole-Cell Biotransformation. *International Journal of Molecular Sciences* **15**, 12196-12216 (2014).
51. Mehmood, N. et al. Impact of two ionic liquids, 1-ethyl-3-methylimidazolium acetate and 1-ethyl-3-methylimidazolium methylphosphonate, on *Saccharomyces cerevisiae*: metabolic, physiologic, and morphological investigations. *Biotechnol Biofuels* **8**(2015).
52. Higgins, D.A. et al. Natural Variation in the Multidrug Efflux Pump SGE1 Underlies Ionic Liquid Tolerance in Yeast. *Genetics* **210**, 219-234 (2018).
53. Galluzzi, M. et al. Interaction of Imidazolium-Based Room-Temperature Ionic Liquids with DOPC Phospholipid Monolayers: Electrochemical Study. *Langmuir* **29**, 6573-6581 (2013).

54. Yoo, B. et al. Molecular mechanisms of ionic liquid cytotoxicity probed by an integrated experimental and computational approach. *Scientific Reports* **6**, 19889 (2016).
55. Stock, F. et al. Effects of ionic liquids on the acetylcholinesterase - a structure-activity relationship consideration. *Green Chemistry* **6**, 286-290 (2004).
56. Auxenfans, T., Husson, E. & Sarazin, C. Simultaneous pretreatment and enzymatic saccharification of (ligno) celluloses in aqueous-ionic liquid media: A compromise. *Biochemical Engineering Journal* **117**, 77-86 (2017).
57. Nargotra, P., Vaid, S. & Bajaj, B. Cellulase Production from *Bacillus subtilis* SV1 and Its Application Potential for Saccharification of Ionic Liquid Pretreated Pine Needle Biomass under One Pot Consolidated Bioprocess. *Fermentation* **2**, 19 (2016).
58. Sandoval, N.R. & Papoutsakis, E.T. Engineering membrane and cell-wall programs for tolerance to toxic chemicals: beyond solo genes. *Current opinion in microbiology* **33**, 56-66 (2016).
59. Tan, Z. et al. Engineering *Escherichia coli* membrane phospholipid head distribution improves tolerance and production of biorenewables. *Metabolic Engineering* **44**, 1-12 (2017).
60. de Kroon, A.I.P.M., Rijken, P.J. & De Smet, C.H. Checks and balances in membrane phospholipid class and acyl chain homeostasis, the yeast perspective. *Progress in Lipid Research* **52**, 374-394 (2013).
61. Renne, M.F. & de Kroon, A.I.P.M. The role of phospholipid molecular species in determining the physical properties of yeast membranes. *FEBS Letters* **592**, 1330-1345 (2018).
62. Yang, Z. et al. Fluconazole inhibits the cellular ergosterol synthesis to confer synergism to berberine against yeast cells. *Journal of Global Antimicrobial Resistance* (2017).
63. Cromie, S.R.T., Del Pópolo, M.G. & Ballone, P. Interaction of Room Temperature Ionic Liquid Solutions with a Cholesterol Bilayer. *The Journal of Physical Chemistry B* **113**, 11642-11648 (2009).

64. Russo, G., Witos, J., Rantamäki, A.H. & Wiedmer, S.K. Cholesterol affects the interaction between an ionic liquid and phospholipid vesicles. A study by differential scanning calorimetry and nanoplasmonic sensing. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1859**, 2361-2372 (2017).
65. Klahn, M. & Zacharias, M. Transformations in plasma membranes of cancerous cells and resulting consequences for cation insertion studied with molecular dynamics. *Phys Chem Chem Phys* **15**, 14427-41 (2013).
66. Ghiaci, P., Norbeck, J. & Larsson, C. Physiological adaptations of Saccharomyces cerevisiae evolved for improved butanol tolerance. *Biotechnology for Biofuels* **6**, 101 (2013).
67. Wang, J.-H., Hung, W. & Tsai, S.-H. High efficiency transformation by electroporation of Yarrowia lipolytica. *The Journal of Microbiology* **49**, 469-472 (2011).
68. Ryu, S. & Trinh, C.T. Understanding Functional Roles of Native Pentose-Specific Transporters for Activating Dormant Pentose Metabolism in Yarrowia lipolytica. *Applied and Environmental Microbiology* **84**(2018).
69. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343 (2009).
70. Rabinowitz, J.D. & Kimball, E. Acidic Acetonitrile for Cellular Metabolome Extraction from *Escherichia coli*. *Analytical Chemistry* **79**, 6167-6173 (2007).
71. Lu, W. et al. Metabolomic Analysis via Reversed-Phase Ion-Pairing Liquid Chromatography Coupled to a Stand Alone Orbitrap Mass Spectrometer. *Analytical Chemistry* **82**, 3212-3221 (2010).
72. Cassilly, C.D. et al. Role of phosphatidylserine synthase in shaping the phospholipidome of *Candida albicans*. *FEMS Yeast Research* **17**, fox007-fox007 (2017).
73. Ichihara, K.i. & Fukubayashi, Y. Preparation of fatty acid methyl esters for gas-liquid chromatography. *Journal of Lipid Research* **51**, 635-640 (2010).

74. Santivañez-Veliz, M. et al. Development, validation and application of a GC–MS method for the simultaneous detection and quantification of neutral lipid species in *Trypanosoma cruzi*. *Journal of Chromatography B* **1061-1062**, 225-232 (2017).
75. Richard, M., Quijano, R.R., Bezzate, S., Bordon-Pallier, F. & Gaillardin, C. Tagging Morphogenetic Genes by Insertional Mutagenesis in the Yeast *Yarrowia lipolytica*. *Journal of Bacteriology* **183**, 3098-3107 (2001).
76. Sheng, W., Yamashita, S., Ohta, A. & Horiuchi, H. Functional Differentiation of Chitin Synthases in *Yarrowia lipolytica*. *Bioscience, Biotechnology, and Biochemistry* **77**, 1275-1281 (2013).
77. Tautenhahn, R., Patti, G.J., Rinehart, D. & Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Analytical Chemistry* **84**, 5035-5039 (2012).
78. Xia, J. & Wishart, D.S. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Current Protocols in Bioinformatics* **55**, 14.10.1-14.10.91 (2016).
79. Li, S. et al. Predicting Network Activity from High Throughput Metabolomics. *PLOS Computational Biology* **9**, e1003123 (2013).
80. Martens, L. et al. mzML—a Community Standard for Mass Spectrometry Data. *Molecular & Cellular Proteomics : MCP* **10**, R110.000133 (2011).
81. Melamud, E., Vastag, L. & Rabinowitz, J.D. Metabolomic Analysis and Visualization Engine for LC−MS Data. *Analytical Chemistry* **82**, 9818-9826 (2010).
82. Clasquin, M.F., Melamud, E. & Rabinowitz, J.D. LC-MS Data Processing with MAVEN: A Metabolomic Analysis and Visualization Engine. in *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc., 2002).
83. Thomas-Chollier, M. et al. Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nature Protocols* **6**, 1860 (2011).

Appendix

Table 1- 1. Untargeted LCMS summary of perturbed pathways with pathway significance factor > 0.58.

Biological Condition	Perturbed Pathway	Pathway size (P_{size})	Total features ($P_{features}$)	Significant features ($P_{significant}$)	Pathway significance factor (psf)
YlCW001 0%	adenine and adenosine salvage IV	8	6	4	0.58
YlCW001 0%	alanine biosynthesis	4	4	4	0.78
YlCW001 0%	alanine degradation III	4	4	4	0.78
YlCW001 0%	arginine biosynthesis	26	17	10	0.80
YlCW001 0%	de novo biosynthesis of pyrimidine ribonucleotides	23	15	7	0.67
YlCW001 0%	gluconeogenesis	21	13	6	0.59
YlCW001 0%	homocysteine and cysteine interconversion	11	8	5	0.63
YlCW001 0%	isoleucine biosynthesis	15	9	7	0.62
YlCW001 0%	leucine biosynthesis	16	9	8	0.62
YlCW001 0%	lysine biosynthesis	22	12	9	0.63
YlCW001 0%	phenylalanine biosynthesis	11	9	5	0.71
YlCW001 0%	phenylalanine degradation	12	9	6	0.71
YlCW001 0%	purine ribonucleosides degradation to ribose-1-phosphate	13	10	5	0.67
YlCW001 0%	superpathway of purine nucleotides de novo biosynthesis	42	21	11	0.64
YlCW001 0%	tryptophan biosynthesis	17	12	7	0.72
YlCW001 0%	tyrosine biosynthesis	11	9	5	0.71
YlCW001 0%	tyrosine degradation	12	9	5	0.65
YlCW001 0%	uridine-5'-phosphate biosynthesis	20	13	6	0.62
YlCW001 0%	sterol biosynthesis (zymosterol)	22	14	9	0.74
YlCW001 8%	4-amino-2-methyl-5-diphosphomethylpyrimidine biosynthesis	11	8	5	0.63
YlCW001 8%	4-aminobenzoate biosynthesis	7	6	4	0.66
YlCW001 8%	alanine biosynthesis	4	4	3	0.67
YlCW001 8%	alanine degradation III	4	4	3	0.67
YlCW001 8%	arginine biosynthesis	26	19	7	0.75
YlCW001 8%	aspartate degradation II	8	7	3	0.59
YlCW001 8%	chitin degradation	15	9	7	0.62
YlCW001 8%	coenzyme A biosynthesis	13	10	5	0.67
YlCW001 8%	dTMP de novo biosynthesis (mitochondrial)	11	8	5	0.63
YlCW001 8%	folate interconversions	22	14	7	0.65
YlCW001 8%	gluconeogenesis	21	16	8	0.84
YlCW001 8%	glycine biosynthesis from glyoxylate	4	4	3	0.67
YlCW001 8%	glycine biosynthesis from threonine	3	3	3	0.67
YlCW001 8%	glyoxylate cycle	13	10	8	0.84
YlCW001 8%	histidine biosynthesis	22	16	5	0.63
YlCW001 8%	homocysteine and cysteine interconversion	11	8	8	0.80

Table 1-1 continued

YlCW001 8%	isoleucine biosynthesis	15	11	9	0.85
YlCW001 8%	isoleucine degradation	10	8	4	0.62
YlCW001 8%	leucine biosynthesis	16	13	7	0.83
YlCW001 8%	leucine degradation	10	8	4	0.62
YlCW001 8%	lysine biosynthesis	22	16	8	0.80
YlCW001 8%	methionine biosynthesis	18	12	9	0.78
YlCW001 8%	methionine salvage pathway	15	9	7	0.62
YlCW001 8%	p-aminobenzoate biosynthesis	7	6	4	0.66
YlCW001 8%	phenylalanine biosynthesis	11	9	8	0.90
YlCW001 8%	phenylalanine degradation	12	11	7	0.94
YlCW001 8%	phosphopantothenate biosynthesis I	16	12	5	0.65
YlCW001 8%	salvage pathways of pyrimidine deoxyribonucleotides	17	11	8	0.71
YlCW001 8%	superpathway of purine nucleotides de novo biosynthesis	42	26	9	0.72
YlCW001 8%	TCA cycle, aerobic respiration	23	15	10	0.80
YlCW001 8%	thiazole biosynthesis III (eukaryotes)	11	9	5	0.71
YlCW001 8%	threonine degradation	29	20	14	1.00
YlCW001 8%	tryptophan biosynthesis	17	13	6	0.73
YlCW001 8%	tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	9	5	0.60
YlCW001 8%	tryptophan degradation VIII (to tryptophol)	12	11	5	0.79
YlCW001 8%	tyrosine biosynthesis	11	10	6	0.86
YlCW001 8%	tyrosine degradation	12	11	5	0.79
YlCW001 8%	UDP-N-acetylglucosamine biosynthesis	12	9	5	0.65
YlCW001 8%	valine biosynthesis	12	8	6	0.63
YlCW001 8%	valine degradation	10	8	4	0.62
YlCW001 8%	sterol biosynthesis (zymosterol)	22	13	8	0.65
wildtype 8%	4-amino-2-methyl-5-diphosphomethylpyrimidine biosynthesis	11	7	6	0.60
wildtype 8%	4-aminobenzoate biosynthesis	7	6	5	0.74
wildtype 8%	alanine biosynthesis	4	4	4	0.78
wildtype 8%	alanine degradation III	4	4	4	0.78
wildtype 8%	arginine biosynthesis	26	18	10	0.85
wildtype 8%	arginine degradation I (arginase pathway)	11	8	5	0.63
wildtype 8%	aspartate degradation II	8	7	4	0.68
wildtype 8%	galactose degradation I (Leloir pathway)	11	8	5	0.63
wildtype 8%	GDP-mannose biosynthesis	8	6	5	0.65
wildtype 8%	gluconeogenesis	21	16	8	0.84
wildtype 8%	glycogen degradation	16	9	8	0.62
wildtype 8%	glycolysis III (glucokinase)	18	13	8	0.79
wildtype 8%	glyoxylate cycle	13	10	6	0.73
wildtype 8%	histidine biosynthesis	22	16	9	0.85
wildtype 8%	homocysteine and cysteine interconversion	11	8	6	0.69

Table 1-1 continued

wildtype 8%	inosine-5'-phosphate biosynthesis	16	10	6	0.59
wildtype 8%	isoleucine biosynthesis	15	10	9	0.78
wildtype 8%	isoleucine degradation	10	8	4	0.62
wildtype 8%	leucine biosynthesis	16	13	7	0.83
wildtype 8%	leucine degradation	10	8	4	0.62
wildtype 8%	lysine biosynthesis	22	15	9	0.79
wildtype 8%	p-aminobenzoate biosynthesis	7	6	5	0.74
wildtype 8%	pentose phosphate pathway (non-oxidative branch)	7	7	3	0.67
wildtype 8%	phenylalanine biosynthesis	11	9	8	0.90
wildtype 8%	phenylalanine degradation	12	11	7	0.94
wildtype 8%	purine ribonucleosides degradation to ribose-1-phosphate	13	9	5	0.60
wildtype 8%	salvage pathways of pyrimidine deoxyribonucleotides	17	11	8	0.71
wildtype 8%	superpathway of purine nucleotides de novo biosynthesis	42	25	12	0.80
wildtype 8%	TCA cycle, aerobic respiration	23	15	8	0.71
wildtype 8%	threonine degradation	29	19	10	0.80
wildtype 8%	tryptophan biosynthesis	17	11	7	0.66
wildtype 8%	tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	13	8	6	0.58
wildtype 8%	tryptophan degradation VIII (to tryptophol)	12	11	7	0.94
wildtype 8%	tyrosine biosynthesis	11	9	8	0.90
wildtype 8%	tyrosine degradation	12	11	7	0.94
wildtype 8%	valine degradation	10	8	5	0.69
wildtype 8%	sterol biosynthesis (zymosterol)	22	12	8	0.60

Table 1- 2. Locus tags of the sterol pathway genes and sterol transcription factor.

Genes	**Locus tag**
ster1	YALI0A10076g
ster2	YALI0E15730g
ster3	YALI0F04378g
ster4	YALI0B05126g
ster5	YALI0B23298g
ster6	YALI0F11297g
ster7	YALI0C22165g
ster8	YALI0B17644g
ster9	YALI0F08701g
ster10-1	YALI0E32065g
ster10-2	YALI0B17204g
ster11	YALI0D20878g
ster12	YALI0A18062g
ster13	YALI0D19206g
sterTF	YALI0B15818g

Table 1- 3. List of primers used in this study. Bold italics are AscI and AvrII sequences.

Genes	Primers	Sequences
Primers used for constructing plasmids		
pSR008$_{Enz}$	AscI_pTEF_Rev	TT***GGCGCGCC***TTTGAATGATTCTTATACTCAGAAGGAAATGC
	AvrII_cyc1t_Fwd	TCCC***CCTAGG***TCATGTAATTAGTTATGTCACGCTTACATTCAC
pSR008$_{Gibson}$	cyc1t_Fwd	TCATGTAATTAGTTATGTCACGCTTAC
	pTEF_Rev	TTTGAATGATTCTTATACTCAGAAG
Yli ster1	Ster1$_{YL}$_Fwd	cgccgagtcaagacgaaactaaggccttattgaaaATGGGAAAACTCATCGAACTGCTC
	Ster1$_{YL}$_Rev	gaatgtaagcgtgacataactaattacatgaCTAATCTCTCAGAGGAAACATCTTAGAG
Yli ster2	Ster2$_{YL}$_Fwd	TT***GGCGCGCC*** atggtcacccaacagtctgcagcag
	Ster2$_{YL}$_Rev	TCCC***CCTAGG***ctaagtcagctcgctccaaatgtaag
Yli ster3	Ster3$_{YL}$_Fwd	TT***GGCGCGCC***atgggaatccacgaaagtgtgtcgaaac
	Ster3$_{YL}$_Rev	TCCC***CCTAGG***tcactggcaatatcccttgaaatacaacccc
Yli ster4	Ster4$_{YL}$_Fwd	TT***GGCGCGCC***atgatcattctcacgacgctcaacaac
	Ster4$_{YL}$_Rev	TCCC***CCTAGG***ctagttacgctctcgcttgctccacac
Yli ster5	Ster5$_{YL}$_Fwd	ccgagtcaagacgaaactaaggccttattgaaaATGACAGCCAAGAGCAAAACGAC
	Ster5$_{YL}$_Rev	gtgaatgtaagcgtgacataactaattacatgaTTAGTAGATGCCTGGGATGATAGC
Yli ster6	Ster6$_{YL}$_Fwd	ccgagtcaagacgaaactaaggccttattgaaaATGTCTTCGATTCTCGAGAGCATC
	Ster6$_{YL}$_Rev	cgtgaatgtaagcgtgacataactaattacatgaTTACTGCTGCTTCTTGGCCTTC
Yli ster7	Ster7$_{YL}$_Fwd	cgagtcaagacgaaactaaggccttattgaaaATGAACACGGTACTGATCGTCGG
	Ster7$_{YL}$_Rev	gaatgtaagcgtgacataactaattacatgaTTACTTCTTCTCCTTGGTCTCGTTG
Yli ster8	Ster8$_{YL}$_Fwd	TT***GGCGCGCC***atgatccacaacagaaaaacgcagacg
	Ster8$_{YL}$_Rev	TCCC***CCTAGG***ttagttaatctgatccttcaacttctctctc
Yli ster9	Ster9$_{YL}$_Fwd	TT***GGCGCGCC***atgtcttccaacatcaaactggctcag
	Ster9$_{YL}$_Rev	TCCC***CCTAGG***tcagttcttgtcctcgggctttcg
Yli ster10-1	Ster10-1$_{YL}$_Fwd	TT***GGCGCGCC***atgaaattcgtcggtctgatttccctg
	Ster10-1$_{YL}$_Rev	TCCC***CCTAGG***ttagatcttaccccaaagagcgtttcg
Yli ster10-2	Ster10-2$_{YL}$_Fwd	TT***GGCGCGCC***atgaaactcgtcggattcatctcgctgc
	Ster10-2$_{YL}$_Rev	TCCC***CCTAGG***tcagatcttaccccataacatgtttc
Yli ster11	Ster11$_{YL}$_Fwd	TT***GGCGCGCC***atggatatcgctctggagaccatcg
	Ster11$_{YL}$_Rev	TCCC***CCTAGG***ttaatcctgcttggtgtttcgcttgac
Yli ster12	Ster12$_{YL}$_Fwd	TT***GGCGCGCC***atgaacgctacccaaccggagtcgcacattgtgc
	Ster12$_{YL}$_Rev	TCCC***CCTAGG***ctactccctcttcttaaactgcaacag
Yli ster13	Ster13$_{YL}$_Fwd	TT***GGCGCGCC***atgtctgctgttcgacaacgtaaaag
	Ster13$_{YL}$_Rev	TCCC***CCTAGG***ttagaaaacataagggatgaagatccaagggac
Yli sterTF	SterTF$_{YL}$_Fwd	TT***GGCGCGCC***atggatatcgctctggagaccatcg
	SterTF$_{YL}$_Rev	TCCC***CCTAGG***ttaatcctgcttggtgtttcgcttgac
Primers used for RT-PCR		
Yli actin	Actin rt_Fwd	TCCAGGCCGTCCTCTCCC
	Actin rt_Rev	GGCCAGCCATATCGAGTCGCA
Yli ster1	ster1 rt_Fwd	GCGAGATCGTGTGGTGCTAC
	ster1 rt_Rev	GTCCACGCCGTTGTTGTTGA
Yli ster2	ster2 rt_Fwd	CGCACGAGACACACCCAATG
	ster2 rt_Rev	GACGACGGTGAGAGGAGCAA
Yli ster3	ster3 rt_Fwd	AGGCCCAATGCGAACTCACT
	ster3 rt_Rev	AAGCGCGTTGCCAAACTTCA

Table 1-3 continued

Yli ster4	ster4 rt_Fwd	CTCGTTTGTCGTGCCCAAGG
	ster4 rt_Rev	GTCTCGGACTCGGCCATCTT
Yli ster5	ster5 rt_Fwd	ACGTCCAGATTGCCACGTCT
	ster5 rt_Rev	AGCTCTCGGCCAATGAACCA
Yli ster6	ster6 rt_Fwd	CTCTACTTTGGCCGGTGCGT
	ster6 rt_Rev	AAGTGCGAGATGAGCACGGA
Yli ster7	ster7 rt_Fwd	CGGACTCGGAAAGGCCATCT
	ster7 rt_Rev	GGAGAACCGTCGGAGACAGC
Yli ster8	ster8 rt_Fwd	ATGCCGTAGCCAAGGAGGTC
	ster8 rt_Rev	GTTGGCTGCGTTGGAGTTGA
Yli ster9	ster9 rt_Fwd	TGAGAAGGGCGACGGTATCC
	ster9 rt_Rev	TCACCGGCCAGAGGGTAGTA
Yli ster10-1	ster10-1 rt_Fwd	GCCTCTCTGCACGACCGATA
	ster10-1 rt_Rev	CCGACAGCAGTGCCAAAGAA
Yli ster10-2	ster10-2 rt_Fwd	CTGTGGGTACCGAGGGTCAC
	ster10-2 rt_Rev	ACCTGGCCTCGAACCAGATG
Yli ster11	ster11 rt_Fwd	AGGTGGCCAAGTTTCTCCGT
	ster11 rt_Rev	TGGTCGTAGGCGTGTTTCCA
Yli ster12	ster12 rt_Fwd	CGCATACGTCAAGCCCTGTG
	ster12 rt_Rev	TGGGCAGGTACATGCTGAGG
Yli ster13	ster13 rt_Fwd	ACCTCCATTGTGCTGGCCTT
	ster13 rt_Rev	GTGCTCGTGGCCCATGTAGA
Yli sterTF	sterTF rt_Fwd	GCGAGATCGTGTGGTGCTAC
	sterTF rt_Rev	GTCCACGCCGTTGTTGTTGA

Table 1- 4. List of plasmids and strains used in this study.

Plasmids/strains	Genotypes	Source
Plasmids		
pSR008	pSL16-P_{TEF}-cyc1t::URA3	[1]
pSR166	pSL16-P_{TEF}-Yli ster1-cyc1t::ura	This study
pSR167	pSL16-P_{TEF}-Yli ster2-cyc1t::ura	This study
pSR168	pSL16-P_{TEF}-Yli ster3-cyc1t::ura	This study
pSR169	pSL16-P_{TEF}-Yli ster4-cyc1t::ura	This study
pSR170	pSL16-P_{TEF}-Yli ster5-cyc1t::ura	This study
pSR171	pSL16-P_{TEF}-Yli ster6-cyc1t::ura	This study
pSR172	pSL16-P_{TEF}-Yli ster7-cyc1t::ura	This study
pSR173	pSL16-P_{TEF}-Yli ster8-cyc1t::ura	This study
pSR174	pSL16-P_{TEF}-Yli ster9-cyc1t::ura	This study
pSR175	pSL16-P_{TEF}-Yli ster10-1-cyc1t::ura	This study
pSR176	pSL16-P_{TEF}-Yli ster10-2-cyc1t::ura	This study
pSR177	pSL16-P_{TEF}-Yli ster11-cyc1t::ura	This study
pSR178	pSL16-P_{TEF}-Yli ster12-cyc1t::ura	This study
pSR179	pSL16-P_{TEF}-Yli ster13-cyc1t::ura	This study
pSR180	pSL16-P_{TEF}-Yli sterTF-cyc1t::ura	This study
Y. lipolytica strains		
YlSR001	MATA ura3-302 leu2-270 xpr2-322 axp2-ΔNU49 XPR2::SUC2	ATCC MYA-2613
YlCW001		This study
YlSR009	YlSR001 + pSR008	This study
YlSR172	YlSR001 + pSR166	This study
YlSR173	YlSR001 + pSR167	This study
YlSR174	YlSR001 + pSR168	This study
YlSR175	YlSR001 + pSR169	This study
YlSR176	YlSR001 + pSR170	This study
YlSR177	YlSR001 + pSR171	This study
YlSR178	YlSR001 + pSR172	This study
YlSR179	YlSR001 + pSR173	This study
YlSR180	YlSR001 + pSR174	This study
YlSR181	YlSR001 + pSR175	This study
YlSR182	YlSR001 + pSR176	This study
YlSR183	YlSR001 + pSR177	This study
YlSR184	YlSR001 + pSR178	This study
YlSR185	YlSR001 + pSR179	This study
YlSR186	YlSR001 + pSR180	This study
YlSR187	YlCW001 + pSR008	This study
YlSR188	YlCW001 + pSR166	This study
YlSR189	YlCW001 + pSR167	This study
YlSR190	YlCW001 + pSR168	This study
YlSR191	YlCW001 + pSR169	This study
YlSR192	YlCW001 + pSR170	This study
YlSR193	YlCW001 + pSR171	This study
YlSR194	YlCW001 + pSR172	This study
YlSR195	YlCW001 + pSR173	This study
YlSR196	YlCW001 + pSR174	This study
YlSR197	YlCW001 + pSR175	This study
YlSR198	YlCW001 + pSR176	This study
YlSR199	YlCW001 + pSR177	This study
YlSR200	YlCW001 + pSR178	This study
YlSR201	YlCW001 + pSR179	This study
YlSR202	YlCW001 + pSR180	This study

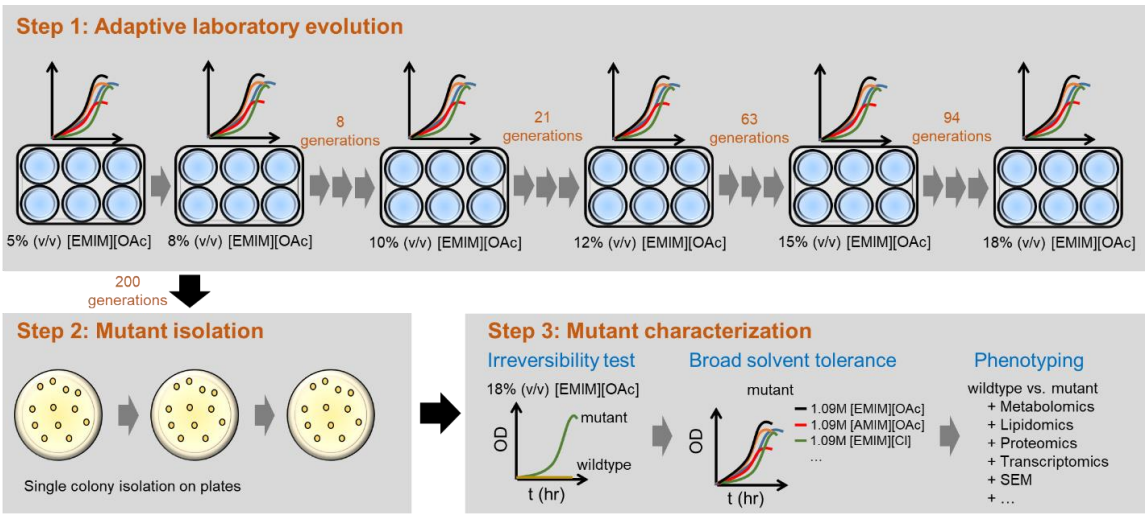


Figure 1- 1. Evolutionary engineering of IL-tolerant *Y. lipolytica* strains. **Step 1**: Adaptive laboratory evolution of *Y. lipolytica* on various concentrations of [EMIM][OAc]. **Step 2**: Single mutant isolation on plates without IL. **Step 3**: Mutant characterization to elucidate the underlying mechanism of solvent tolerance.

Figure 1- 2. (**A**) Adaptive laboratory evolution of *Y. lipolytica* grown with increasing concentrations of [EMIM][OAc]. (**B, C**) Irreversibility testing of the evolved *Y. lipolytica* strain YlCW001 in various concentrations of [EMIM][OAc]. (**D, E**) Broad tolerance test for growth of *Y. lipolytica* wildtype and mutant at various concentrations of 0.6M and 1.09M ILs. Error bars represent the standard deviation of three independent biological replicates repeated twice (n=6). Abbreviations: [EMIM][OAc]: 1-ethyl-3-methylimidizolium acetate; [EMIM][Cl]: 1-ethyl-3-methylimidizolium chloride; [EMIM][Br]: 1-ethyl-3-methylimidizolium bromide acetate; [BMIM][OAc]: 1-butyl-3-methylimidazolium; [BMIM][Cl]: 1-butyl-3-methylimidizollium chloride; [BMIM][Br]: 1-butyl-3-methylimidazolium bromide; and [AMIM][Cl]: 1-allyl-3-methylimidizolium chloride.

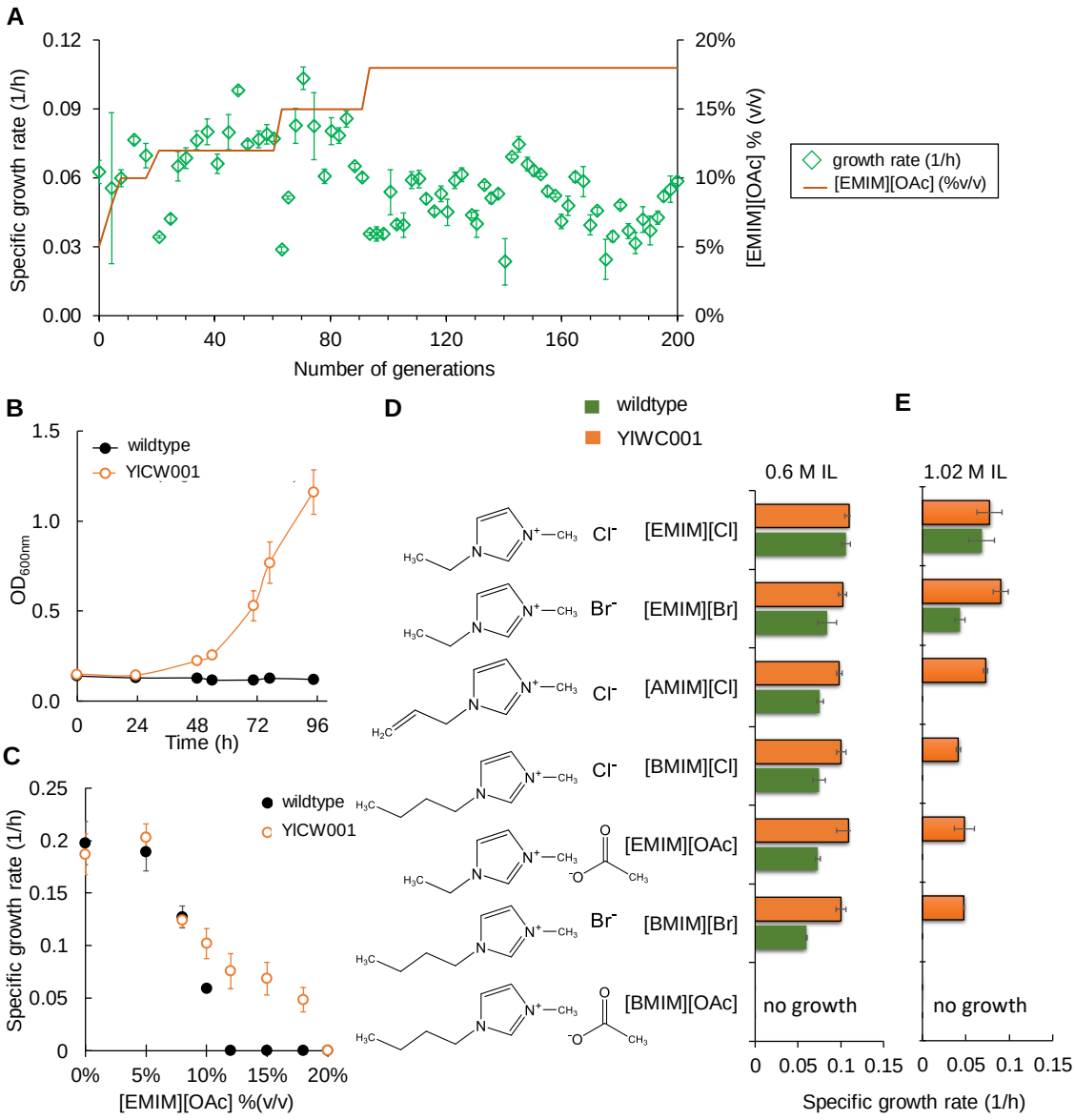
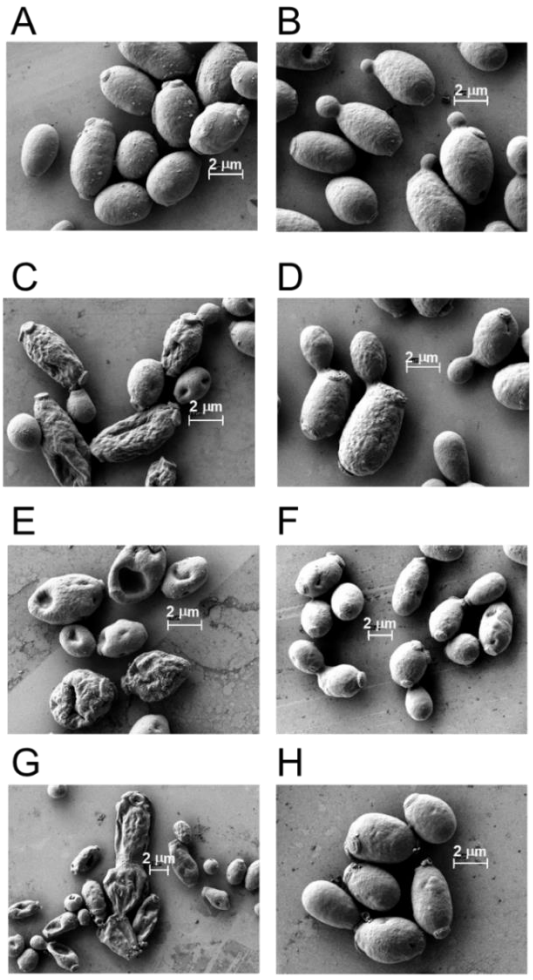

A
Specific growth rate (1/h)
[EMIM][OAc] % (v/v)
Number of generations
growth rate (1/h)
[EMIM][OAc] (%v/v)
B
wildtype
YICW001
OD600nm
Time (h)
C
wildtype
YICW001
Specific growth rate (1/h)
[EMIM][OAc] %(v/v)
D
wildtype
YIWC001
[EMIM][Cl]
[EMIM][Br]
[AMIM][Cl]
[BMIM][Cl]
[EMIM][OAc]
[BMIM][Br]
[BMIM][OAc]
E
0.6 M IL
1.02 M IL
no growth
no growth
Specific growth rate (1/h)

Figure 1- 3. SEM imaging of the wildtype and YlCW001 cells exposed to (**A, B**) no IL, (**C, D**) 18% (v/v) [EMIM][OAc], (**E, F**) 0.3M [BMIM][OAc], and (**G, H**) 0.6M [BMIM][OAc]. (**I**) IL-responsive pathways perturbed with respect to wildtype *Y. lipolytica* growing in medium without IL with a psf score > 0.58. Biological triplicates were performed for all biological conditions and results were combined from hydrophilic metabolomics extraction (negative ionization, n = 3), and hydrophobic lipidomics extractions (negative ionization, n = 3; positive ionization, n = 3).


A
B
C
D
E
F
G
H
2 µm


I
Wildtype 8%
YICW001 8%
YICW001 0%
Threonine degradation
Phenylalanine degradation
Phenylalanine biosynthesis
Tyrosine biosynthesis
Isoleucine biosynthesis
Glyoxylate cycle
Gluconeogenesis
Leucine biosynthesis
TCA cycle, aerobic respiration
Homocysteine and cysteine interconversion
Lysine biosynthesis
Tryptophan degradation VIII (to tryptophol)
Tyrosine degradation
Methionine biosynthesis
Arginine biosynthesis
Tryptophan biosynthesis
Superpathway of purine nucleotides de novo biosynthesis
Salvage pathways of pyrimidine deoxyribonucleotides
Thiazole biosynthesis III
Alanine biosynthesis
Alanine degradation III
Glycine biosynthesis
Coenzyme A biosynthesis
4-aminobenzoate biosynthesis
p-aminobenzoate biosynthesis
Folate interconversions
Phosphopantothenate biosynthesis I
UDP-N-acetylglucosamine biosynthesis
Sterol biosynthesis (zymosterol)
Valine biosynthesis
4-amino-2-methyl-5-diphosphomethylpyrimidine biosynthesis
dTMP de novo biosynthesis (mitochondrial)
Histidine biosynthesis
Isoleucine degradation
Leucine degradation
Valine degradation
Chitin degradation
Methionine salvage pathway
Tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde
Aspartate degradation II
Adenosine and adenosine salvage IV
Arginine degradation I (arginase pathway)
De novo biosynthesis of pyrimidine ribonucleotides
Galactose degradation I (Leloir pathway)
GDP-mannose biosynthesis
Glycogen degradation
Glycolysis III (glucokinase)
Inosine-5'-phosphate biosynthesis
Pentose phosphate pathway (non-oxidative branch)
Purine ribonucleosides degradation to ribose-1-phosphate
Uridine-5'-phosphate biosynthesis
Pathway Significance Factor
< 0.5
0.7
1

Figure 1- 4. Glycerophospholipid and fatty acid reorganization in wildtype *Y. lipolytica* and YlCW001 in 0% and 8% (v/v) [EMIM][OAc]. (**A**) PC: Phosphatidylcholine; (**B**) PI: phosphatidylinositol; (**C**) PA: phosphatidic acid; (**D**) PG: phosphatidyl glycerol; (**E**) PS: phosphatidylserine; (**F**) PE: phosphatidylethanolamine; (**G**) CL: cardiolipin. (**H**) Fatty acid distributions. (**I**) Chemical structures of glycerophospholipid backbone and headgroup species. All error bars represent standard deviation of biological triplicates (n=3) and statistical significance was calculated using one-way analysis of variance (ANOVA) with Holm-Sidak correction against control group, wildtype in 0% IL (degrees of freedom, 3). Symbols: "*": *p*-value < 0.05; "**": *p*-value < 0.01; "***", *p*-values < 0.001.

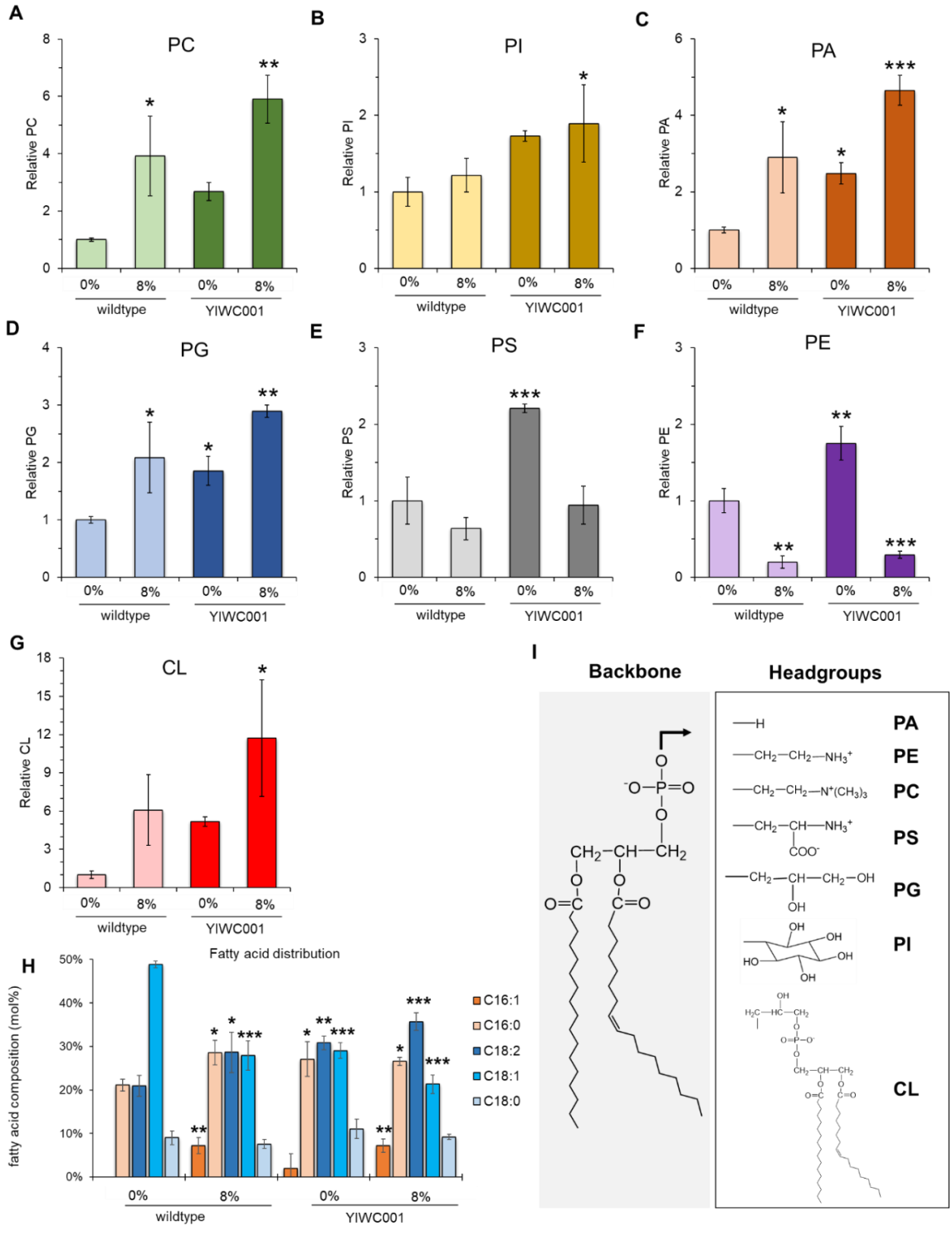
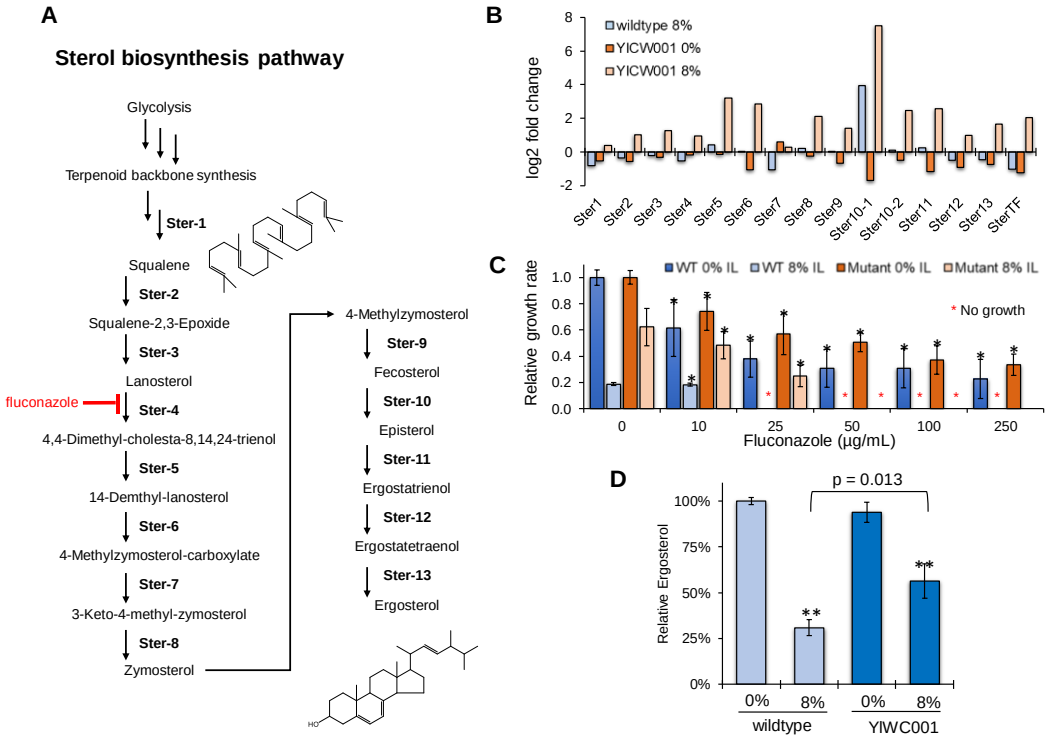

A
PC
Relative PC
B
PI
Relative PI
C
PA
Relative PA
D
PG
Relative PG
E
PS
Relative PS
F
PE
Relative PE
0%
8%
wildtype
YIWC001
G
CL
Relative CL
H
Fatty acid distribution
fatty acid composition (mol%)
C16:1
C16:0
C18:2
C18:1
C18:0
I
Backbone
Headgroups
PA
PE
PC
PS
PG
PI
CL

Figure 1- 5. (**A**) Sterol biosynthesis pathway in *Y. lipolytica*. Fluconazole is an antifungal drug that inhibits sterol 14-demethylase (Ster4) (**B**) Differential gene expression of the steroid pathway as compared to the wildtype strain in medium without IL (Ster1, YALI0A10076g; Ster2, YALI0E15730g; Ster3, YALI0F04378g; Ster4, YALI0B05126g; Ster5, YALI0B23298g; Ster6, YALI0F11297g; Ster7, YALI0C22165g; Ster8, YALI0B17644g; Ster9, YALI0F08701g; Ster10-1, YALI0E32065g; Ster10-2, YALI0B17204g; Ster11, YALI0D20878g; Ster12, YALI0A18062g; Ster13, YALI0D19206g). (**C**) Effect of fluconazole inhibiting the steroid pathway on cell growth relative to wildtype in non-IL, non-fluconazole media. Error bars represent standard deviation from three technical replicates repeated twice per biological condition (n = 6) (**D**) Relative ergosterol content of wildtype and YlCW001 strains in 0% and 8% (v/v) [EMIM][OAc]. (**E**) Toxicity plate assay of *Y. lipolytica* strains expressing the sterol biosynthesis and transcription factor genes in presence of various concentrations of [EMIM][OAc]. Error bars represent standard deviation of biological triplicates (n =3) and statistical significance was calculated between the wildtype and YlCW001 in 8% IL using the Student's t-test (t = -4.244 with 4 degrees of freedom). Abbreviation: "**", *p*-value = 0.013.



E

		0% (v/v) IL			4% (v/v) IL			8% (v/v) IL			10% (v/v) IL			12% (v/v) IL		
No. cells: 1.2x		10^5	10^4	10^3	10^5	10^4	10^3	10^5	10^4	10^3	10^5	10^4	10^3	10^5	10^4	10^3
Strain	wildtype															
	Ster1															
	Ster2															
	Ster3															
	Ster4															
	Ster5															
	Ster6															
	Ster7															
	Ster8															
	Ster9															
	Ster10-1															
	Ster10-2															
	Ster11															
	Ster12															
	Ster13															
	SterTF															
	YICW001															

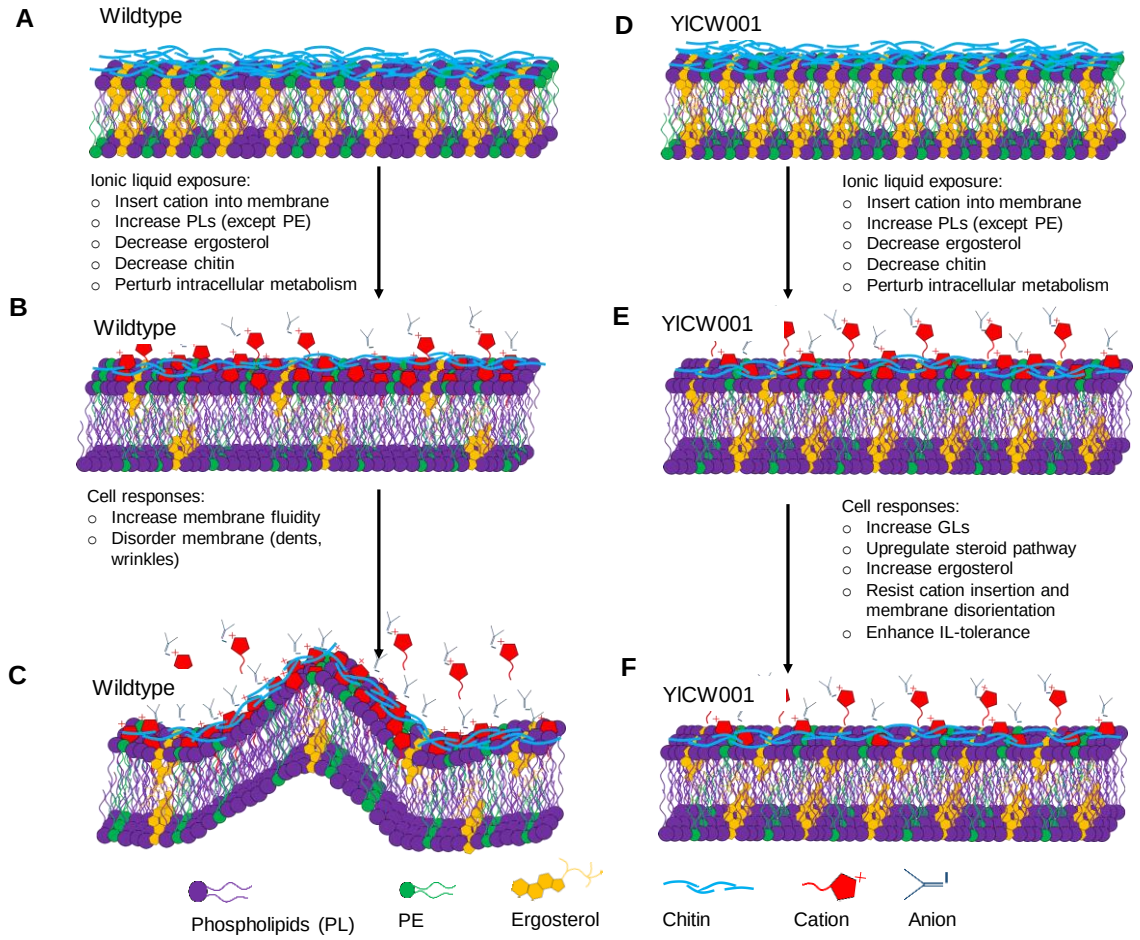


Figure 1- 6. Mechanisms of IL interference and tolerance in wildtype *Y. lipolytica* and YlCW001. (**A, D**) No IL exposure. (**B, E**) IL exposure and interference. (**C, D**) Cellular membrane response.

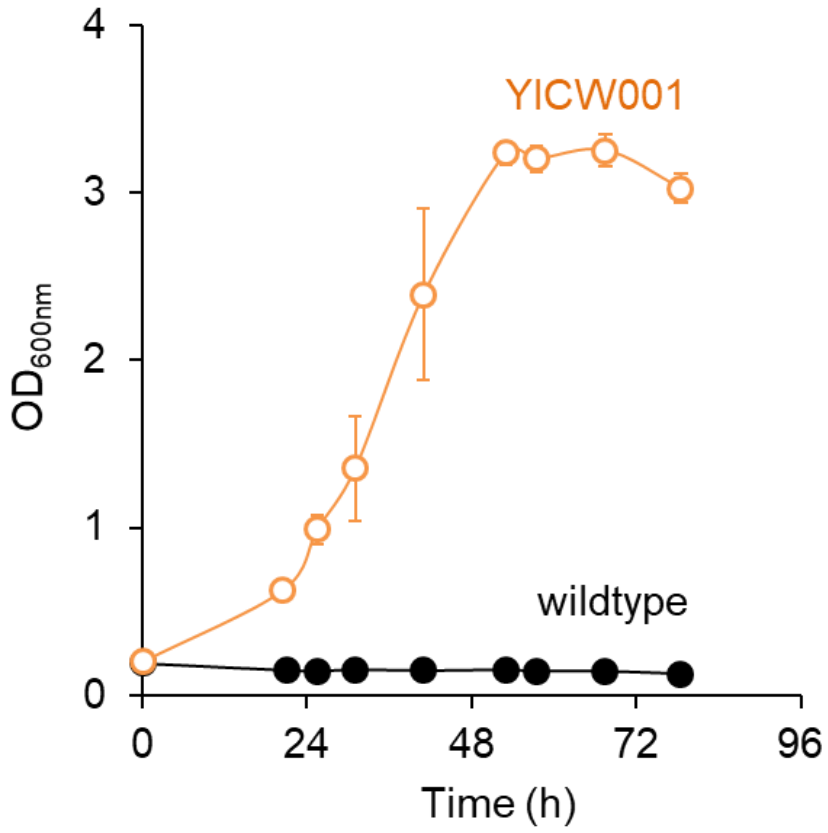


Figure 1- 7**.** Effect of 0.3M [BMIM][OAc] on growth of *Y. lipolytica* wildtype and YICW001. Each data point is an average ± standard deviation of three independent biological replicates.

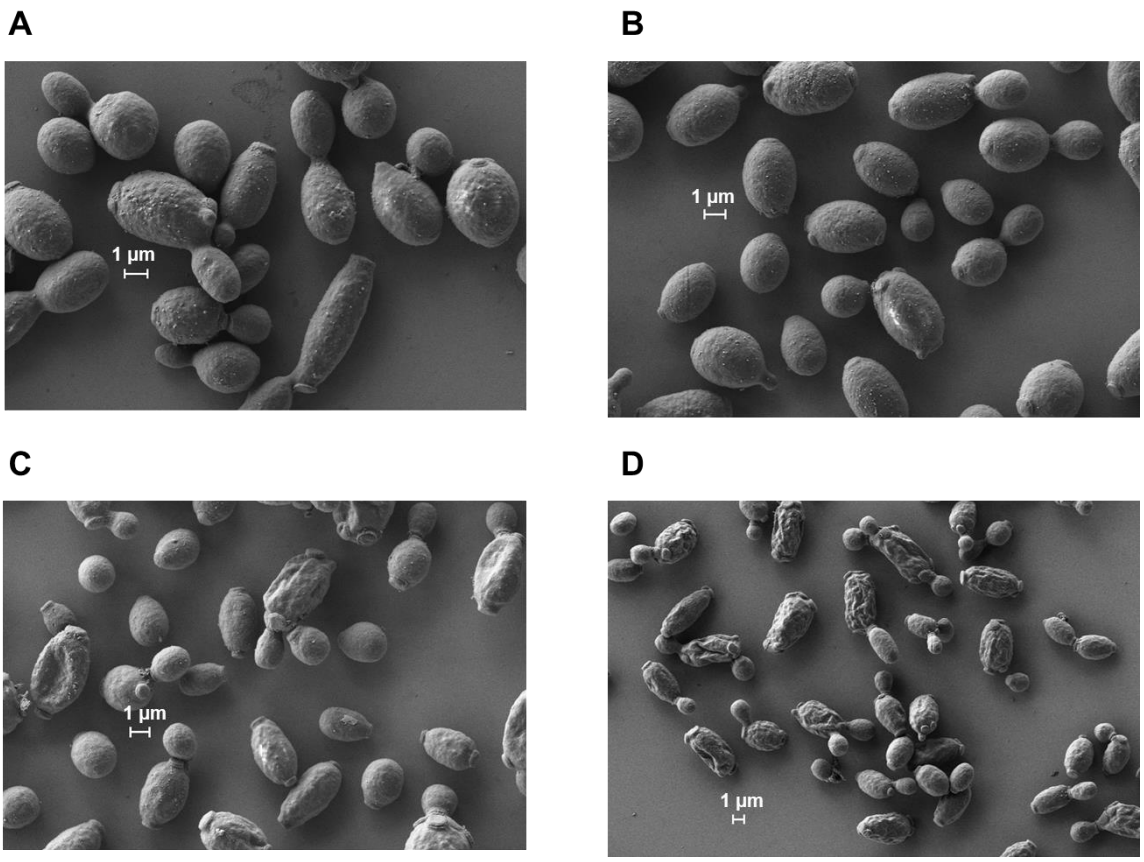


Figure 1- 8. SEM images of wildtype Y. lipolytica cells incubated for 6 hours in (**A**) no IL, (**B**) 2% (v/v) [EMIM][OAc], (**C**) 8% (v/v) [EMIM][OAc] and (**D**) 12% (v/v) [EMIM][OAc].

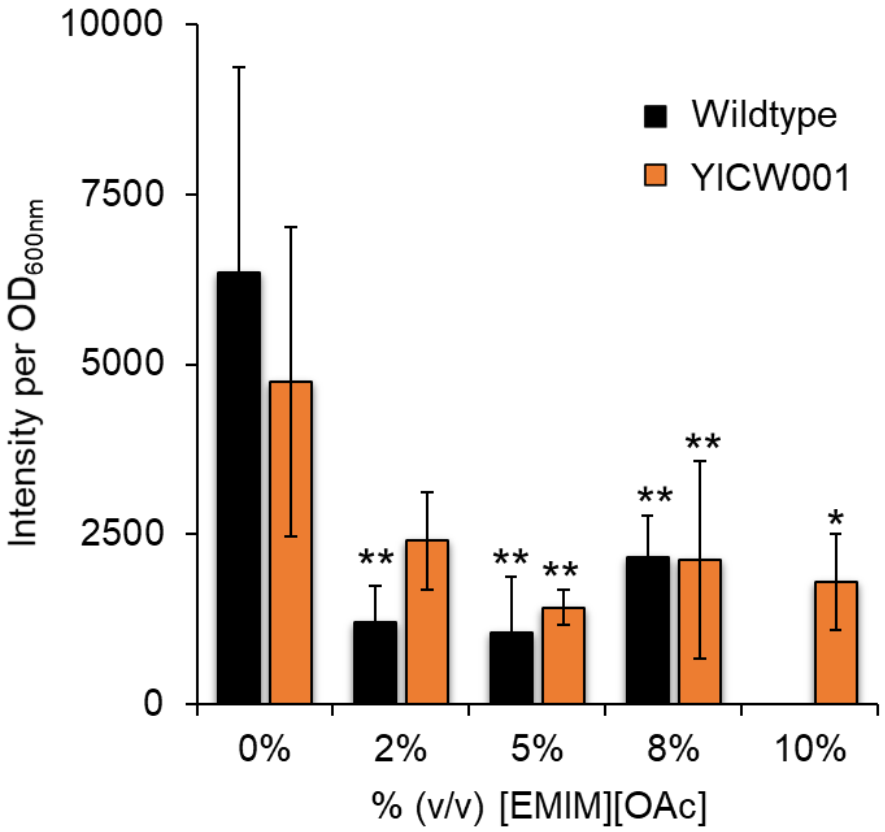


Figure 1- 9. Effect of various concentrations [EMIM][OAc] on the chitin content in the membranes of the wildtype and YlCW001. Each experiment was conducted at least twice for each biological condition in technical replicates, i.e., wildtype 0%, n = 18; wildtype 2%, n = 6; wildtype 5%, n = 15; wildtype 8%, n = 12; YlCW001 0%, n = 25; YlCW001 2%, n = 6; YlCW001 5%, n = 15; YlCW001 8%, n = 29; YlCW001 10%, n = 6. Each error bar represents standard deviation of technical replicates and statistical significance was calculated using one-way analysis of variance (ANOVA) with Holm-Sidak correction against all pairwise groups (degree of freedom=8, F=18.53). Asterisks represent statistical significance against respective strain in 0% IL where *, p-values < 0.01, **p-values < 0.001.

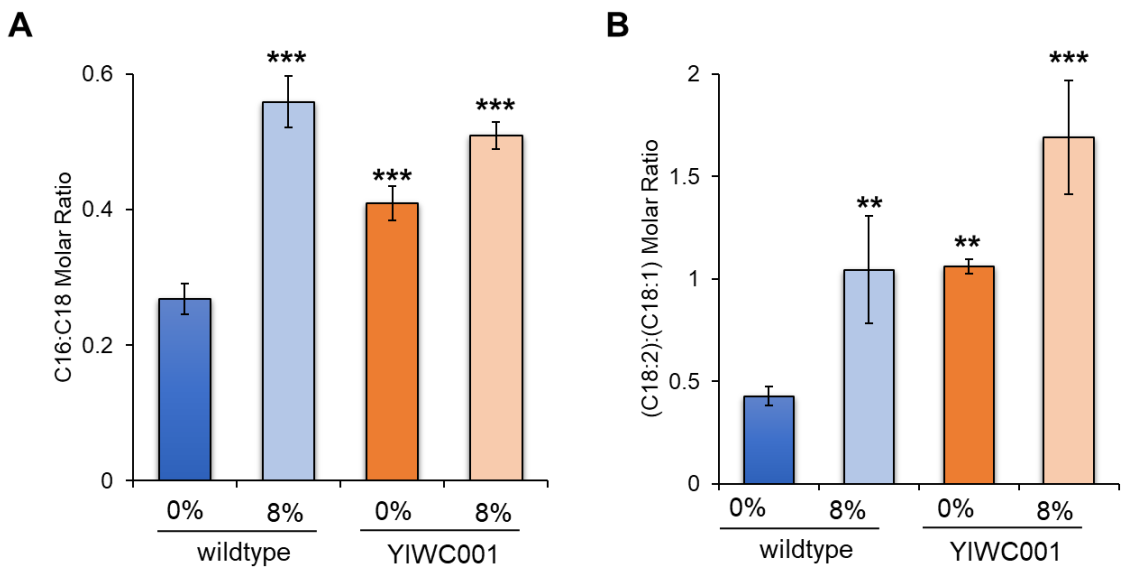


Figure 1- 10. Fatty acid molar ratios of wildtype and YlCW001 in 0% and 8% (v/v) [EMIM][OAc]. (**A**) Molar ratio of C16:C18 fatty acid moieties. (**B**) Molar ratio of (C18:2):(C18:1) fatty acid moieties. Each error bar represents standard deviation of 3 biological replicates and statistical significance was calculated using one-way analysis of variance (ANOVA) with Holm-Sidak correction against control group, wildtype in 0% IL (degree of freedom = 3). *P values < 0.05, **P values < 0.01, ***P values < 0.001.

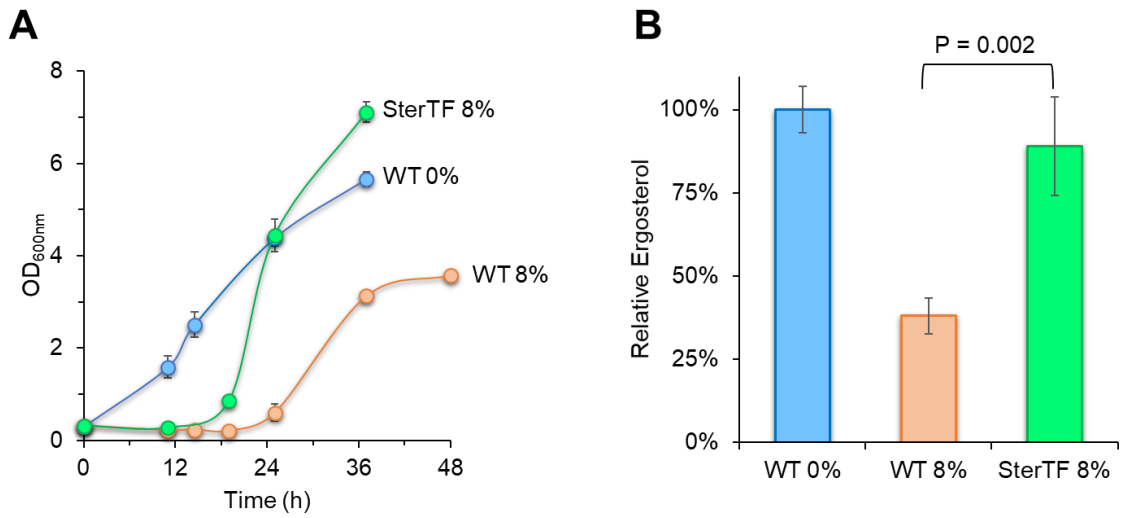


Figure 1- 11. (**A**) Growth kinetics and (**B**) ergosterol content of SterTF overexpression in 8% (v/v) [EMIM][OAc] relative to wildtype in 0% and 8% (v/v) [EMIM][OAc] at 36 h. Error bars represent the average and standard deviation of biological triplicates. Statistical significance was calculated between the wildtype and SterTF-expressing strains in 8% (v/v) [EMIM][OAc] using one-way analysis of variance (ANOVA) with Holm-Sidak correction.

Gene Promoter	Predicted TF	P-value	Sequence Motif (3' to 5')
Ster1	SterTF (UPC2)	<0.00843	
Ster2	SterTF (ECM22)	0.0268	
Ster3	SterTF (ECM22)	0.0469	
Ster4	SterTF (ECM22)	0.046	
Ster5	N/A	N/A	N/A
Ster6	SterTF (UPC2)	<0.000571	
Ster7	N/A	N/A	N/A
Ster8	N/A	N/A	N/A
Ster9	SterTF (UPC2)	1.09E-05	
Ster10-1	SterTF (UPC2)	<0.00449	
Ster10-2	SterTF (ECM22)	0.0235	
Ster11	N/A	N/A	N/A
Ster12	N/A	N/A	N/A
Ster13	SterTF (UPC2)	<0.00527	
pTEF	N/A	N/A	N/A

Figure 1- 12. Transcription factor affinity prediction (TRAP) for sterol biosynthesis genes. TRAP was performed for 1050bp nucleotide sequences upstream of each sterol biosynthesis gene and pTEF promoter as negative control.

CHAPTER II

Understanding and eliminating the detrimental effect of thiamine deficiency on the oleaginous yeast *Yarrowia lipolytica.*

A version of this chapter was originally published by Caleb M. Walker:
Walker, C.**, Ryu, S.**, Giannone, R.S., Garcia, S., Trinh, C.T.*, 2020. Understanding and eliminating the detrimental effect of endogenous thiamine auxotrophy on metabolism of the oleaginous yeast Yarrowia lipolytica, Appl Environ Microbiol, 86(3): e02299-19.

CTT conceived and supervised the study. CW, SR, RSG and CTT designed the experiments, and analyzed the data. CW and SR performed the experiments. CW, SR, and CTT wrote and approved the manuscript.

Abstract

Thiamine is a vitamin that functions as a cofactor for key enzymes in carbon and energy metabolism for all living cells. While most plants, fungi and bacteria can synthesize thiamine *de novo*, the oleaginous yeast, *Yarrowia lipolytica*, cannot. In this study, we used proteomics together with physiological characterization to understand key metabolic processes influenced and regulated by thiamine availability and identified the genetic basis of thiamine auxotrophy in *Y. lipolytica*. Specifically, we found thiamine depletion results in decreased protein abundance of the lipid biosynthesis pathways and energy metabolism (i.e., ATP synthase), attributing to the negligible growth and poor sugar assimilation observed in our study. Using comparative genomics, we identified the missing 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase (THI13) in the *de novo* thiamine biosynthesis of *Y. lipolytica,* and discovered an exceptional promoter, P3, that exhibits strong activation or tight repression by low and high thiamine concentrations, respectively. Capitalizing on the strength of our thiamine-regulated promoter (P3) to express the missing gene from *Saccharomyces cerevisiae* (scTHI13), we engineered the thiamine-prototrophic *Y. lipolytica.* By comparing this engineered strain to the wildtype, we unveiled the tight relationship linking thiamine availability to lipid biosynthesis and demonstrated enhanced lipid production with thiamine supplementation in the engineered thiamine-prototrophic *Y. lipolytica*.

Introduction

Thiamine, or vitamin B1, was the first B vitamin discovered. Its activated form, thiamine pyrophosphate (TPP), functions as a cofactor for key enzymes in carbon metabolism

including citrate (Krebs) cycle, pentose phosphate and branched chain amino acid pathways (Figure 2-1) (1). TPP-dependent enzymes, including pyruvate dehydrogenase (PDH), α-ketoglutarate dehydrogenase (KGDH), transketolase (TKL), branched chain α-ketoacid dehydrogenase (BCKDC) and acetolactate synthase (AHAS), are essential for maintaining cell growth and preventing metabolic stress (2-4). Specifically, PDH links the glycolysis and citrate cycle by catalyzing the conversion of pyruvate into acetyl-CoA (5, 6). KGDH, a key enzyme of the citrate cycle, converts α-ketoglutarate (KGA) to succinyl-CoA (7, 8). TKL participates in the pentose phosphate pathway by interconverting pentose sugars to glycolysis intermediates (9). The pentose phosphate pathway is critical for production of ribose (i.e., RNA), precursor metabolites for aromatic amino acid pathways, and reducing equivalents (i.e., NADPH) necessary for maintaining redox balance and lipid synthesis. Hence, TKL activity is critical for RNA, protein, and lipid production while preventing oxidative stress (10, 11). Meanwhile, AHAS and BCKDC are responsible for the synthesis and degradation of branched chain amino acids (i.e., valine, leucine and isoleucine), respectively (12, 13).

In mammals, thiamine deficiency affects the cardiovascular and nervous systems resulting in tremors, muscle weakness, paralysis and even death (14, 15). Thiamine deficiency can occur from inadequate intake, increased requirement, or impaired absorption of thiamine (16). Biochemical consequences of thiamine deficiency result in failure to produce adenosine triphosphate (ATP), increased acid production (i.e., lactic acid), decreased production of acetylated compounds (i.e., acetylcholine), neurotransmitters (i.e., glutamate, aspartate, aminobutyric acid), reduced nicotinamide adenosine dinucleotide phosphate (NADH), defective ribonucleic acid (RNA) ribose synthesis, and failure to break down branched chain carboxylic acids (i.e., leucine, valine, isoleucine) (17-19).

While mammals require nutritional supplementation of thiamine from dietary sources, most bacteria, fungi and plants can synthesize thiamine endogenously (20). One of these exceptions is the thiamine-auxotrophic oleaginous yeast, *Yarrowia lipolytica,* which has recently emerged as an important industrial microbe with broad biotechnological applications due to its generally regarded as safe (GRAS) status (21),

metabolic capability (22-26), and robustness (27-29). Hence, *Y. lipolytica* is an ideal eukaryotic host to study the fundamental effects of thiamine deficiency on cellular health. Currently, it is not well understood what causes failure of thiamine biosynthesis in *Y. lipolytica* nor how thiamine deficiency affects other processes (e.g., lipid biosynthesis, energy metabolism, etc.)

Interestingly, *Y. lipolytica*'s auxotrophy has been exploited for enhanced production of organic acids (i.e., pyruvate and KGA) by reducing activities of PDH and KGDH under thiamine-limited conditions (30, 31). Consequently, cell growth is negatively affected by thiamine limitation and completely prevented when thiamine is depleted from the media. Not surprisingly, the genes in thiamine metabolism are tightly regulated by thiamine concentrations (32, 33). To this end, numerous thiamine-regulated promoters have been discovered in yeast that enable control of genetic expression by adjusting thiamine concentration in the culture media (34-37). However, endogenous thiamine-regulated promoters have not yet been discovered in *Y. lipolytica*.

In this study, we shed light on the effect of thiamine deficiency on cellular metabolism in the thiamine auxotroph *Y. lipolytica*. We identified the missing gene encoding 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase (THI13) in the *de novo* thiamine biosynthesis pathway of *Y. lipolytica* and discovered a thiamine-regulated promoter, P3, that increases expression in low thiamine concentrations. By employing P3 to control the expression of scTHI13 derived from *Saccharomyces cerevisiae*, we engineered thiamine prototrophy in *Y. lipolytica*. This enabled us to elucidate the relationship between thiamine availability and ATP and lipid biosynthesis and revealed enhanced lipid production in the engineered thiamine-prototrophic *Y. lipolytica*.

RESULTS

The effects of thiamine deficiency in Y. lipolytica

Cell growth and organic acid production are influenced by thiamine limitation and depletion. To demonstrate the effects of thiamine limitation, we

characterized the thiamine-auxotrophic *Y. lipolytica* (YlSR001) in 0, 0.5 and 400 µg/L thiamine (Figure 2-2A). Cell growth was inhibited by limited (0.5 µg/L thiamine) and depleted (0 µg/L thiamine) concentrations of thiamine but restored in high thiamine (400 µg/L thiamine) containing media (Figure 2B). Glucose consumption profiles were closely coordinated with cell growth (Figure 2-2C). Within the first 24 hr, only ~2 g/L glucose was assimilated without thiamine while thiamine-limited cultures consumed ~3 g/L glucose (Figure 2-2C). After 24 h, growth and glucose consumption were completely stalled under the no thiamine condition. In contrast, cells continued to utilize glucose even though growth was significantly inhibited.

Next, we characterized pyruvate and KGA production since the enzymes converting these organic acids, PDH and KGDH respectively, are TPP-dependent. In high thiamine media, *Y. lipolytica* demonstrated slight KGA production (~0.5 g/L KGA) and no pyruvate accumulation (Figure 2-2D, 2-2E). We observed substantial pyruvate titers in thiamine limited (~4 g/L pyruvate) and depleted (~2 g/L pyruvate) media (Figure 2-2E), indicating that PDH is unable to efficiently convert pyruvate to acetyl-CoA. Similarly, KGA accumulation was enhanced in thiamine-limited media (2 g/L KGA) but KGA titers in thiamine-depleted media were comparable to high thiamine media likely due to decreased flux from glycolysis to the TCA cycle through reduced acetyl-CoA production via PDH (Figure 2-2D). Taken together, *Y. lipolytica* requires thiamine supplementation for cell growth and carbon assimilation but produces organic acids under thiamine-limited conditions.

Proteome of thiamine depletion reveals perturbation of critical metabolic pathways related to cellular growth. Next, we investigated the proteome of the thiamine-auxotrophic *Y. lipolytica* growing in 0 and 400 µg/L thiamine (Figure 2-3A). Across 2 exponential time points, we identified 535 upregulated and 515 downregulated proteins (i.e., log2 fold change > |1|) in response to thiamine deficiency (Figure 3B). First, we looked at metabolic enzymes that require TPP as a cofactor including PDH, KGDH, transketolase (TKL), branched chain α-ketoacid dehydrogenase (BCKDC) and acetolactate synthase (AHAS) (Supplementary Table S1). Interestingly, all proteins encoding subunits (E1-E3) of BCKDC were upregulated in thiamine depletion

(Supplementary Table S1). However, none of the other TPP-requiring proteins were upregulated in media lacking thiamine except for dihydrolipoamide dehydrogenase (E3), which serves as a subunit for BCKDC, PDH and KGDH (Table 2-S1).

We then mapped the 535 upregulated- and 515 downregulated proteins to their respective KEGG pathways (Figure 2-3C). Thiamine deficiency resulted in downregulation of over 55% of proteins involved in nucleic acid metabolism (i.e., pyrimidine and purine metabolism) and genetic processes (i.e., ribosome and DNA replication). Without thiamine, cells also exhibited decreased protein abundance in lipid metabolism which includes the synthesis of terpenoids (100% of proteins), steroids (75% of proteins) and glycerophospholipids (80% of proteins). In contrast, thiamine deficiency resulted in upregulation of over 65% of proteins contained in carbohydrate (i.e., glycolysis and citrate cycle), tetrapyrrole, biotin, aromatic (i.e., tryptophan, phenylalanine), branched chain (i.e., leucine, isoleucine, valine), and other amino acid (i.e., glycine, serine, threonine) metabolic pathways. Thiamine deficiency also affected proteins associated with the electron transport chain (ETC), amino acid (i.e., arginine, proline, glutamate, glutamine, methionine, cysteine) biosynthesis and thiamine metabolic pathways.

A closer look at proteins involved in the ETC and thiamine metabolism revealed interesting features. First, thiamine-deficient cells exhibited increased protein abundances for all four complexes of the ETC but decreased protein abundance for ATP synthase that is important for ATP generation (Figure 2-3D). The phenomenon correlates with the stalled assimilation of glucose that is ATP-dependent when cells grow in the absence of thiamine. Second, all but one of the proteins in thiamine metabolism were differentially regulated between thiamine-sufficient and thiamine-depleted cultures (Figure 2-3E). Without thiamine, we observed increased abundance of proteins in the upper branch of thiamine synthesis but decreased abundance in proteins converting thiamine monophosphate into thiamine and TPP into thiamine triphosphate. The only detected protein in thiamine metabolism that was not changed by thiamine concentrations was thiamine kinase (Thi90p), which is responsible for the conversion of thiamine into its activated diphosphate form, TPP. Taken together, thiamine concentration influences regulation of thiamine metabolism, whereby thiamine deficiency severely affects central

carbon metabolism and elicits increased protein abundance for most carbohydrate, amino acid and energy pathways but decreased protein abundance for lipid metabolism, nucleotide metabolism and specifically, ATP synthase.

Restoring thiamine prototrophy in Y. lipolytica

Thiamine metabolism in *Y. lipolytica* is incomplete. Next, we investigated the native thiamine metabolism of *Y. lipolytica* to elucidate underlying genetic deficiency causing thiamine-auxotrophic behavior. We compared the thiamine biosynthesis pathways between the well-characterized thiamine-prototrophic *Saccharomyces cerevisiae* and our thiamine-auxotrophic *Y. lipolytica*. Between the two organisms, we identified that the 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase corresponding to THI13 in *S. cerevisiae* (scTHI13) is missing in *Y. lipolytica*. scTHI13 converts histidine and pyrodixine into hydroxymethylpyrimidine pyrophosphate (HMP-PP), which is likely required for the *de novo* thiamine synthesis in *Y. lipolytica* (Figure 2-3E).

Constitutive expression of the missing thiamine gene does not effectively restore prototrophy. To enable the *de novo* biosynthesis of thiamine in *Y. lipolytica*, we constructed a vector to express scTHI13 under the constitutive promoter, TEF, frequently used for genetic overexpression in *Y. lipolytica*. Unexpectedly, expression of scTHI13 with the TEF promoter only restored the *de novo* TPP biosynthesis after days of adaptation in thiamine-depleted media (Figure 2-S1). Additionally, cell growth was highly deviated between replicates for both times when this experiment was conducted (Figure 2-S1). Though the results were somewhat promising, we endeavored to engineer a true thiamine-prototrophic *Y. lipolytica*.

Bioinformatics of thiamine-responsive genes reveals highly regulated thiamine promoter. We hypothesized that expression of scTHI13 was weak since the constitutive TEF promoter (38, 39) is dependent on cell growth (40, 41) and thiamine deficiency prevents cell growth. To this end, we aimed to find a promoter responsive to thiamine deficiency. Through BlastP and orthology analysis using the well-studied thiamine-regulated genes from *Pichia pastoris*, *Saccharomyces cerevisiae* and

Schizosaccharomyces pombe, we identified three candidate genes that are putatively regulated by thiamine in *Y. lipolytica* (Table 2-S2). The P1 gene (YALI0E04224g) encodes a putative thiaminase that might exhibit hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase activity for conversion of thiamine into 4-amino-5-hydroxymethyl-2-methylpyrimidine diphosphate. The P3 gene (YALI0A09768g) encodes a putative cysteine-dependent adenosine diphosphate thiazole synthase that is involved in the biosynthesis of thiazole, a thiamine precursor. Thiazole synthase converts NAD^+ and glycine into ADP-5-ethyl-4-methylthiazole-2-carboxylate, a thiazole intermediate. Though the function of P2 (YALI0C14652g) has yet to be characterized, it harbors a NMT1 domain (Pfam:PF09084) which is required for biosynthesis of the pyrimidine moiety of thiamine (42); hence, P2 might be thiamine-regulated.

To determine if these promoters (i.e., P1, P2, and P3) are thiamine-regulated, real-time PCR was conducted to quantify mRNA levels of P1, P2, and P3 genes from YlSR001 grown in low (0.5 µg/L) and high (500 µg/L) concentrations of thiamine (Figure 2-4A). The expression levels of P1 and P3 genes were activated in low thiamine but inactivated in high thiamine. Remarkably, the P3 gene expression level was substantially higher in low thiamine than high thiamine. We did not observe transcriptional expression of P2 in either low or high thiamine concentrations.

We next investigated plasmid expression of these 3 putative thiamine-regulated promoters from the *Y. lipolytica* strains that harbor a humanized renilla green fluorescent protein (hrGFP)-expressing gene under the control of either P1, P2, or P3 promoter (strains YlSR1002, YlSR1003 and YlSR1004, respectively). For controls, we also built the native, constitutive TEF promoter and the heterologous NMT1 promoter from *S. pombe* that has been well-studied and applied as a thiamine-regulated promoter (43). The hrGFP intensity was monitored from cells grown at low (1 µg/L) and high (10 mg/L) concentrations of thiamine (Figure 2-4B). As expected, both P1 and P3 promoters were activated in low thiamine and inactivated in high thiamine. Under low thiamine conditions, the hrGFP intensity under promoter P3 was much higher than P1 and TEF promoters by 10.04 ± 0.46-fold and 2.82 ± 0.13-fold, respectively. Not surprisingly, the activity of TEF promoter was much greater in high thiamine than in low thiamine. The NMT1 promoter

showed no activity in low thiamine while limited activity was exhibited in high thiamine as expected due to the species specificity.

We further investigated the sensitivity of the P3 promoter over a range of low thiamine concentrations (0.5-50 µg/L) (Figure 2-4C). Encouragingly, the P3 promoter exhibited tight regulation across all incremental thiamine concentrations. Of note, the activity of the P3 promoter was not affected by poor cell growth in low thiamine concentrations (Figure 2-4D). Taken together, these data indicate that the endogenous thiamine-regulated promoters, P1 and P3, can be applied to strain engineering efforts and are responsive to low thiamine concentrations.

Thiamine prototrophy is restored with thiamine regulated promoter. To restore the *de novo* thiamine synthesis in *Y. lipolytica*, we constructed the vector pSR075 to express the missing thiamine gene, scTHI13, with the thiamine-responsive promoter, P3 (i.e., YlSR1006). Remarkably, this construct demonstrated robust and reproducible growth in media lacking thiamine (Figure 5). Cell growth and glucose uptake were similar despite with low to no added thiamine (Figure 2-5A, 2-5B), but organic acid production was affected. Organic acid production in low (0.5 µg/L) and depleted (0 µg/L) thiamine conditions was diminished during the stationary phase (Figure 2-5C, 2-5D). Meanwhile, KGA accumulation only manifested in high (400 µg/L) thiamine (Figure 2-5C). Taken together, the thiamine prototroph strain created here, YlSR1006, grows reproducibly irrespective of thiamine concentrations but produces no organic acids in thiamine-limited media.

Lipid accumulation is influenced by thiamine concentrations. Finally, we investigated the relationship between thiamine availability and neutral lipid accumulation. We cultured both thiamine-auxotrophic (wildtype) and our engineered thiamine-prototrophic (YlSR1006) strain with 0 µg/L and 400 µg/L thiamine in MpA (no nitrogen limitation, Figure 6A) and lipid production (nitrogen limitation with C:N = 100, Figure 2-6B) media. In non-nitrogen limited media, YlSR1006 produced more lipid than the wildtype grown with 400 µg/L thiamine supplementation (Figure 2-6A). Interestingly, YlSR1006 grown without thiamine accumulated lipid similar to the wildtype with 400 µg/L thiamine. However, under nitrogen-limitation, both the wildtype and YlSR1006 strains

showed similar lipid production profiles when 400 □g/L thiamine was supplemented in growth medium (Figure 2-6B). In thiamine-lacking medium, while no lipid accumulation was expected for the wildtype, YlSR1006 was able to accumulate 3.68 ± 0.22 lipid %DCW, which was ~50% less than YlSR1006 supplemented with 400 µg/L thiamine. Taken together, thiamine supplementation increased lipid production even for thiamine-prototroph YlSR1006, indicating thiamine plays a critical role for lipid biosynthesis in *Y. lipolytica*.

DISCUSSION

In this study, we observed the effects of thiamine deficiency on growth, sugar consumption, organic acid production, and proteome of the thiamine-auxotrophic *Y. lipolytica*. The activated form of thiamine, TPP, is an important cofactor for enzymes involved in vital cellular functions including energy metabolism (44), reducing oxidative and osmotic stresses (45), and catabolism of sugars (46). Hence, the consequences of thiamine deficiency are caused by the reduced activity of TPP-dependent enzymes (i.e., PDH, KGDH, TKL, AHAS and BCKDC), which causes growth cessation and ultimately leads to cell death. A comprehensive model that depicts the detrimental effect of thiamine deficiency on metabolism leading to growth cessation in *Y. lipolytica* is summarized in Figure 2-7.

Loss of PDH and KGDH activities result in poor growth, limited carbon assimilation and accumulation of pyruvate and KGA (Figure 2-2). Failure to metabolize pyruvate via PDH also inhibits production of acetyl-CoA, the precursor metabolite of the TCA cycle. The TCA cycle is further inhibited by reduced KGDH activity, preventing synthesis of NADH required for oxidative phosphorylation (i.e., ETC). Notably, in the model yeast, *Saccharomyces cerevisiae*, deletion of KGDH is known to prevent respiratory growth (47), but the exact mechanism is not well established. In our study, proteomic analysis shows that thiamine-deficient cells increased protein abundance for the first four complexes of oxidative phosphorylation (i.e., NADH dehydrogenase, succinate dehydrogenase, and cytochrome c reductase/oxidase) but decreased protein abundance for ATP synthase (Figure 2-3D).These new findings indicate that thiamine deficiency negatively affects

respiratory energy metabolism caused by a malfunctioning TCA cycle via inhibition of PDH and KGDH as observed by inhibited growth and reduced sugar uptake in thiamine-depleted cells.

Y. lipolytica also decreased protein abundances in glycerophospholipid, terpenoid backbone and sterol biosynthesis pathways in thiamine-deficient cells (Figure 2-3C). These phenomena are likely consequences of reduced PDH activity (i.e., reduced pools of acetyl-CoA) and likely affected by NADPH production by the pentose phosphate pathway. In the pentose phosphate pathway, TKL interconverts pentose sugars and hexose sugars, the later which serve as glycolytic intermediates (e.g., fructose-6P, glyceraldehyde-3P). Hence, loss of TKL activity likely effects the production of ribose-5P and NADPH which are required for synthesis of lipids, RNA, DNA, purines, pyrimidines, and antioxidants (48). Interestingly, loss of TKL activity also prevents production of erythrose-4P, impeding synthesis of folate and aromatic amino acids (i.e., phenylalanine, tyrosine, and tryptophan) as previously observed in TKL-deletion mutants of *S. cerevisiae* (49).

In our study, *Y. lipolytica* also responded to thiamine deficiency by increasing protein abundance of enzymes involved in branched chain α-amino acid metabolism (i.e., valine, leucine and isoleucine). Consistent with the previous studies with *S. cerevisiae*, deletion of BCKDC results in branched chain α-amino acid auxotrophic phenotypes (50). Interestingly, BCKDC was the only TPP-requiring enzyme with all 3 subunits upregulated in our thiamine-depleted *Y. lipolytica* cells. However, this finding might be the result of leucine supplementation in the media since our *Y. lipolytica* strain is leucine auxotrophic. Taken together, thiamine deficiency inhibited cell growth severely limiting energy production, amino acids (i.e., aromatic and branched chain) and lipid synthesis, ultimately leading to cell death (Figure 2-7).

Despite thiamine-mediated growth inhibition, *Y. lipolytica* exhibited strong upregulation of proteins involved in thiamine metabolism in response to thiamine depletion (Figure 2-3E). Interestingly, one of the enzymes upregulated in thiamine depletion, cysteine-dependent adenosine diphosphate thiazole synthase (YALI0A09768g), is driven by the thiamine-regulated promoter, P3. While various

constitutive and inducible promoters are available for *Y. lipolytica* (51-54), the tightly regulated, thiamine-responsive P3 promoter is especially useful for strong, inducible expression in low thiamine conditions. Under optimized conditions, the activity of P3 promoter was 2.82 ± 0.13-fold higher than TEF promoter (Figure 2-4B). Hence, gene overexpression using the P3 promoter is highly desirable over the constitutive TEF promoter in low thiamine concentrations. This was demonstrated by restoring thiamine prototrophy in *Y. lipolytica* using P3 promoter (Figure 2-5), while the TEF promoter was not strong enough to accomplish stable thiamine prototrophy (Figure 2-S1). Overall, the P3 promoter can be used as a thiamine biosensor for applications in synthetic biology and metabolic engineering.

Surprisingly, thiamine prototroph was restored by overexpressing 1 single gene, absent from the native genome of *Y. lipolytica*, for the *de novo* synthesis of thiamine (Figure 2-3E, Figure 2-S1). This begs the question, why does *Y. lipolytica* lack this gene? We hypothesize that most *Y. lipolytica* strains are isolated from thiamine-rich sources (e.g., sausage, oats, plants) (55) while most popular yeasts, including *S. cerevisiae*, are isolated from sugar-rich sources (e.g., fruits, molasses, sugarcane) (55). Remarkably, we observed enhanced lipid production by thiamine supplementation in both thiamine-auxotrophic and prototrophic strains, suggesting a relationship between lipid production and thiamine availability (Figure 2-6A, 2-6B). This relationship shows promise for industrial applications of neutral lipid production and establishes a novel direction for increasing lipid production in *Y. lipolytica*.

MATERIALS AND METHODS

Plasmids and strains

The list of plasmids and strains used in this study is available in Table 2-1. Plasmid pSR005 carrying a humanized renilla green fluorescent protein (hrGFP) was constructed by Gibson assembly method (56) with hrGFP and pSL16-CEN1-1-227 (57). The hrGFP gene was amplified using the primers hrGFP_Fwd and hrGFP_Rev from pBABE GFP

(pBABE GFP was a gift from Dr. William Hahn; Addgene plasmid #10668). The backbone pSL16-CEN1-1-227 was amplified with the primers pSL16_Fwd and pSL16_Rev.

Next, various promoters, TEF(404) (53), NMT1 (58), P1(1000), P2(1000), and P3(1000) were inserted into pSR005 by Gibson assembly method. Each TEF(404), P1(1000), P2(1000), and P3(1000) promoter region (size of each promoter is in parenthesis) was amplified using the primers P_{TEF}_Fwd/P_{TEF}_Rev, P_{P1}_Fwd/P_{P1}_Rev, P_{P2}_Fwd/P_{P2}_Rev, and P_{P3}_Fwd/P_{P3}_Rev, respectively from the genomic DNA of *Y. lipolytica* ATCC MYA-2613. The NMT1 promoter region was amplified using the P_{NMT1}_Fwd/P_{NMT1}_Rev primers from the genomic DNA of *S. pombe* (kindly provided by Dr. Paul Dalhaimer, Department of Chemical and Biomolecular Engineering, University of Tennessee Knoxville, TN, USA). The backbone pSR005 was amplified by using the primers pSR005_Fwd/pSR005_Rev. The constructed plasmids are pAT32y, pSR068, pSR071, pSR072, and pSR073 (Table 2-1).

The plasmid pSR074 was constructed by assembly of (i) THI13sce, amplified using the primers $THI13_{sce}$_Fwd^1/$THI13_{sce}$_Rev from the genomic DNA of *S. cerevisiae* and (ii) pSR008 backbone, amplified using the primers pSR008_Fwd/pSR008_Rev. The plasmid pSR075 was constructed by replacing the hrGFP gene with the THI13sce gene from pSR073. The THI13sce gene was amplified by using the primers $THI13_{sce}$_Fwd^2/$THI13_{sce}$_Rev from the genomic DNA of *S. cerevisiae* and assembled with the P_{P3} promoter carrying backbone, amplified by using the primers pSR008_Fwd/pSR073_Rev.

The *Y. lipolytica* ATCC MYA-2613, obtained from ATCC strain collection, was used as a parent strain. YlSR101, YlSR109, YlSR1001 and-YlSR1006 (Table 2-1) strains were generated by transforming the corresponding plasmids via electroporation (59). Each plasmid was transferred into *Y. lipolytica* YlSR001 via electroporation to generate strains used in this study. The YlSR109, YlSR1001-YlSR1004 strains were confirmed by the respective promoter binding forward primer along with hrGFP_Rev. The YlSR1005 and YlSR1006 strains were confirmed by TEF(-100)_Fwd or P3(-80)_Fwd together with the respective gene binding reverse primer. *Escherichia coli* TOP10 was used for molecular cloning. Primers used in this study are listed in Table 2-2.

Media and culturing conditions

Media. For *E. coli* culture, Luria Bertani medium containing 5 g/L yeast extract, 10 g/L tryptone, and 5 g/L NaCl, 100 mg/L ampicillin as a selection was used. For *Y. lipolytica* characterization, MpA defined media was used for all experiments. MpA components are as follows: 5 g/L $(NH_4)_2SO_4$, 2 g/L KH_2PO_4, 0.5 g/L $MgSO_4$, 44 mg/L $ZnSO_4 \cdot 7H_2O$, 79 mg/L $CaCl_2 \cdot 2H_2O$, 0.8 mg/L Biotin, 100 mM HEPES buffer, 90 mM Na_2HPO_4, 10 mM NaH_2PO_4, trace elements (0.4 mg/L $ZnSO_4 \cdot 7H_2O$, 0.04 $CuSO_4 \cdot 5H_2O$, 0.4 $MnSO_4 \cdot 6H_2O$, 0.2 mg/L $Na_2MoO_4 \cdot 2H_2O$, 0.1 mg/L KI, 0.5 mg/L $FeSO_4 \cdot 7H_2O$, 0.5 mg/L H_3BO_3), 380 mg/L leucine, 20 g/L glucose and various concentrations of thiamine hydrochloride. MpA media was adjusted to pH of 5.

Culturing. All experiments were conducted in a Kuhner LT-X incubator set to 28°C and 250 rpm unless otherwise stated. Fresh colonies were inoculated in 2mL of MpA medium containing 400 µg/L thiamine in 15mL culture tubes overnight. Cultures were centrifuged and resuspended in 2mL of water before transferring 1mL of this suspension into 100mL of MpA containing 5 µg/L thiamine to scale-up cultures for 2 days (Figure 2A). Next, cells were washed once with water and resuspended in 100mL of MpA lacking thiamine for 1 day to eliminate thiamine carry-over (Figure 2A). Finally, cells were washed twice with water before characterization experiments. All experiments were conducted in technical triplicates using 500mL baffled flasks unless otherwise stated.

Analytical methods

Real-time quantitative PCR (rt-PCR). *Y. lipolytica*, grown in MpA medium using glucose as a carbon source together with either low (0.5 µg/L) and high (500 µg/L) thiamine, was collected at mid-exponential phase (OD 2 ~ 3). Total RNA was purified by using the Qiagen RNeasy mini kit (Cat # 74104, Qiagen Inc, CA, USA), and cDNA was subsequently synthesized by the QuantiTect Reverse Transcription kit (Cat # 205311, Qiagen Inc, CA, USA). To quantify mRNA expression level of genes (e.g., actin (YALI0D08272g), P1, P2, and P3), rt-PCR run was performed using the QuantiTect SYBR Green PCR kit (Cat # 204143, Qiagen Inc, CA, USA) and StepOnePlus™ Real-Time PCR System (Applied Biosystems, CA, USA). Primers used for rt-PCR are listed in Table 2-2.

The gene expression level was investigated after normalizing by the actin house-keeping gene as described elsewhere (25).

Promoter characterization with hrGFP. Fresh colonies of *Y. lipolytica* promoter constructs were grown in 2 mL of MpA medium containing 400 µg/L thiamine overnight. Cultures were washed once with water and transferred into 25mL of MpA containing 5 µg/L thiamine overnight. Finally, cells were washed twice with water before being inoculated in MpA medium with various concentrations of thiamine. Incubation was performed at 400 rpm and 28° using 96-well plates and Duetz-system covers (Cat# SMCR1296, Kuhner, Switzerland). Sacrificial samples were collected for fluorescence measurements (excitation at 485 nm and emission at 528 nm) using a synergy HT microplate reader.

High performance liquid chromatography. Prior to HPLC run, 1 mL of culture medium was filtered using 0.2 µm filters. Metabolites, substrates and products were quantified by a Shimadzu HPLC system equipped with UV and RID detectors (Shimadzu Scientific Instruments, Inc., MD, USA) and the Aminex 87H column (Biorad, CA, USA) with 10 mN H_2SO_4 mobile phase at 0.6 mL/min flow rate. The column was maintained at 48 °C (25).

Proteomic analysis. *Y. lipolytica* were grown in biological triplicate in 0 and 400 µg/L thiamine. Samples were collected at two time points during the exponential growth phase and processed for LC-MS/MS analysis. Whole-cell lysates were prepared by bead beating in sodium deoxycholate lysis buffer (4% SDC, 100 mM ammonium bicarbonate, pH 8.0) using 0.15 mM zirconium oxide beads and cell debris cleared by centrifugation (21,000 x g for 10 min). Crude protein concentrations were measured by a NanoDrop OneC (ThermoScientific) using absorbance at 205 nm. Samples were then adjusted to 10 mM dithiothreitol and incubated at 85 °C for 10 min to denature and reduce proteins. Cysteines were alkylated/blocked with 30 mM iodoacetamide followed by 20 min incubation at room temperature in the dark. Proteins (300 µg) were then transferred to a 10-kDa MWCO spin filter (Vivaspin 500, Sartorius) and digested *in situ* with proteomics-grade trypsin (Pierce) as previously described (60). The tryptic peptide solution was then filtered through the MWCO membrane by centrifugation (12,000 x g for 15 min), adjusted

to 1% formic acid to precipitate SDC, and SDC precipitate removed from the peptide solution with water-saturated ethyl acetate. Peptide samples were then concentrated to dryness via SpeedVac, resolubilized in solvent A (5% acetonitrile, 95% water, 0.1% formic acid), and measured by NanoDrop OneC A205 to assess tryptic peptide recovery.

Peptide samples were analyzed by automated 1D LC-MS/MS analysis using a Vanquish UHPLC plumbed directly in-line with a Q Exactive Plus mass spectrometer (Thermo Scientific) outfitted with a trapping column coupled to an in-house pulled nanospray emitter. Both the trapping column (100 µm ID) and nanospray emitter (75 µm ID) were packed with 5 µm Kinetex C18 RP resin (Phenomenex) to 10 cm and 30 cm, respectively. For each sample, 3 µg of peptides were loaded, desalted, separated and analyzed across a 210 min organic gradient with the following parameters: sample injection followed by 100% solvent A chase from 0-30 min (load and desalt), linear gradient from 0% to 25% solvent B (70% acetonitrile, 30% water, 0.1% formic acid) from 30-240 min (separation), followed by a ramp to 75% solvent B from 240-250 min (wash), re-equilibration to 100% solvent A from 250-260 min and a hold at 100% solvent A from 260-280 min. Eluting peptides were measured and sequenced by data-dependent acquisition on the Q Exactive MS as previously described (60).

MS/MS spectra were searched against the *Y. lipolytica* proteome concatenated with common protein contaminants using Proteome Discover v.2.2 (ThermoScientific) employing the CharmeRT workflow (61, 62). Peptide spectrum matches (PSM) were required to be fully tryptic with 2 mis-cleavages; a static modification of 57.0214 Da on cysteine (carbamidomethylated) and a dynamic modification of 15.9949 Da on methionine (oxidized) residues. False-discovery rates, as assessed by matches to decoy sequences, were initially controlled at < 1% at both the PSM- and peptide-levels. FDR-controlled peptides were then quantified by chromatographic area-under-the-curve (AUC), mapped to their respective proteins, and areas summed to estimate protein-level abundance. Protein abundance distributions were then normalized across samples using InfernoRDN (63) and missing values imputed to simulate the MS instrument's limit of detection using Perseus (64). Significant differences in protein abundance were calculated separately for each time point according to the following equation:

$$\text{Fold change} = \frac{\text{WT_00}_{t1,2} - \text{WT_400}_{t1,2}}{\sqrt{0.25 + \sum \text{Variance}/n}}$$

Here, WT_00 and WT_400 represent log2 normalized abundance of a protein in 0 and 400 µg/L thiamine, respectively. The denominator was used to account for error between replicates. Variance represents the variance of protein abundance between replicates, n represents the number of replicates, and 0.25 is the pseudo variance term (65). Proteins with fold changes > |1| were classified as upregulated or downregulated. Pathway annotations were performed with ClueGo (66) using KEGG (https://www.genome.jp/kegg/) database on proteins that were upregulated or downregulated at both time points.

All raw and database-searched LC-MS/MS data pertaining to this study have been deposited into the MassIVE proteomic data repository and have been assigned the following accession numbers: MSV000084437 (MassIVE) and PXD015747 (ProteomeXchange). Data files are available at ftp://massive.ucsd.edu/ MSV000084437.

Bioinformatics. Putative native *Y. lipolytica* thiamine-regulated promoters were identified by BlastP (67) and orthologs search through the KEGG sequence similarity database (https://www.kegg.jp/kegg/ssdb/). Reference genes used in this study were (i) *P. pastoris* hydroxymethylpyrimidine phosphate synthase *thi11* (PAS_chr4_0065) (68), (ii) *S. pombe* 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase *nmt1* (NP_588347.1) (58), iii) *S. pombe* thiamine thiazole synthase *nmt2* (NP_596642.1) (69), and iv) *S. cerevisiae* thiamine thiazole synthase *thi4* (NP_011660.1) (70).

Lipid quantification. Thiamine-auxotrophic and thiamine-prototrophic strains were cultured in 0 and 400 µg/L thiamine in MpA media in triplicates as outlined previously. Lipid samples were taken from 100 µL samples of culture broth (i.e., cells and supernatant) and incubated at room temperature for 15 minutes in the dark after addition of 2 µL of 1 µg/mL BODIPY (cat # D3922, Fisher Scientific) (71) which stains neutral lipids (e.g., triacylglycerols). Lipid standards were created by dissolving 100 mg of corn oil in 20mL ethanol and diluted from 1-.1 mg/mL prior to BODIPY staining procedure. Lipids were measured using fluorescence (ex: 485 nm/em: 528 nm) and quantified from corn oil

standards. For dry cell weight (DCW) measurement, 1mL of culture broth was sampled from each replicate at each time point. Samples were centrifuged at maximum speed for 3 minutes and supernatant was discarded prior to drying samples at 55°C overnight. DCW was calculated by subtracting the dried cell pellet and tube weight by the empty tube weight. Finally, % lipid accumulation was calculated by dividing the measured lipid mg/mL by DCW mg/mL. Statistical significance was calculated using SigmaPlot 14 with one-way analysis of variance (ANOVA) with Holm-Sidak correction between 0 and 400 µg/L thiamine for each strain respectively. Symbols: "*": p-value < 0.05.

ACKNOWLEDGEMENTS

We would like to acknowledge financial support from the DOE BER Genomic Science Program (DE-SC0019412) and the National Science Foundation (NSF #1511881). The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the funding agencies.

References

1. Brown GM. 1970. Biogenesis and Metabolism of Thiamine, 3d ed.. ed, vol IV, p 369-378. New York, Academic Press, New York.
2. Vemuganti R, Kalluri H, Yi J-H, Bowen KK, Hazell AS. 2006. Gene expression changes in thalamus and inferior colliculus associated with inflammation, cellular stress, metabolism and structural damage in thiamine deficiency. European Journal of Neuroscience 23:1172-1188.
3. Bunik VI, Tylicki A, Lukashev NV. 2013. Thiamin diphosphate-dependent enzymes: from enzymology to metabolic regulation, drug design and disease models. The FEBS Journal 280:6412-6442.
4. Kowalska E, Kujda M, Wolak N, Kozik A. 2012. Altered expression and activities of enzymes involved in thiamine diphosphate biosynthesis in *Saccharomyces cerevisiae* under oxidative and osmotic stress. FEMS Yeast Research 12:534-546.
5. Linn TC, Pettit FH, Reed LJ. 1969. α-KETO ACID DEHYDROGENASE COMPLEXES, X. REGULATION OF THE ACTIVITY OF THE PYRUVATE DEHYDROGENASE COMPLEX FROM BEEF KIDNEY MITOCHONDRIA BY PHOSPHORYLATION AND DEPHOSPHORYLATION. Proceedings of the National Academy of Sciences 62:234-241.
6. Patel MS, Nemeria NS, Furey W, Jordan F. 2014. The pyruvate dehydrogenase complexes: structure-based function and regulation. The Journal of biological chemistry 289:16615-16623.
7. Gibson GE, Blass JP, Beal MF, Bunik V. 2005. The α-ketoglutarate-dehydrogenase complex. Molecular Neurobiology 31:43-63.
8. Reed LJ. 1974. Multienzyme complexes. Accounts of Chemical Research 7:40-46.
9. Schenk G, Duggleby RG, Nixon PF. 1998. Properties and functions of the thiamin diphosphate dependent enzyme transketolase. The International Journal of Biochemistry & Cell Biology 30:1297-1318.

10. Boros LG, Steinkamp MP, Fleming JC, Lee W-NP, Cascante M, Neufeld EJ. 2003. Defective RNA ribose synthesis in fibroblasts from patients with thiamine-responsive megaloblastic anemia (TRMA). Blood 102:3556-3561.
11. Rapala-Kozik M, Kowalska E, Ostrowska K. 2008. Modulation of thiamine metabolism in *Zea mays* seedlings under conditions of abiotic stress. Journal of Experimental Botany 59:4133-4143.
12. Pettit FH, Yeaman SJ, Reed LJ. 1978. Purification and characterization of branched chain α-keto acid dehydrogenase complex of bovine kidney. Proceedings of the National Academy of Sciences 75:4881-4885.
13. Ciszak EM, Korotchkina LG, Dominiak PM, Sidhu S, Patel MS. 2003. Structural Basis for Flip-Flop Action of Thiamin Pyrophosphate-dependent Enzymes Revealed by Human Pyruvate Dehydrogenase. Journal of Biological Chemistry 278:21240-21246.
14. WILLIAMS RD, MASON HL, SMITH BF, WILDER RM. 1942. INDUCED THIAMINE (VITAMIN B1) DEFICIENCY AND THE THIAMINE REQUIREMENT OF MAN: FURTHER OBSERVATIONS. JAMA Internal Medicine 69:721-738.
15. Organization WH. 1999. Thiamine deficiency and its prevention and control in major emergencies. Department of Nutrition for Health and Development, Organization WH, Geneva, Switzerland.
16. Liu D, Ke Z, Luo J. 2017. Thiamine Deficiency and Neurodegeneration: the Interplay Among Oxidative Stress, Endoplasmic Reticulum Stress, and Autophagy. Molecular Neurobiology 54:5440-5448.
17. Bettendorff L, Goessens G, Sluse F, Wins P, Bureau M, Laschet J, Grisar T. 1995. Thiamine Deficiency in Cultured Neuroblastoma Cells: Effect on Mitochondrial Function and Peripheral Benzodiazepine Receptors. Journal of Neurochemistry 64:2013-2021.
18. Whitfield KC, Bourassa MW, Adamolekun B, Bergeron G, Bettendorff L, Brown KH, Cox L, Fattal-Valevski A, Fischer PR, Frank EL, Hiffler L, Hlaing LM, Jefferds ME, Kapner H, Kounnavong S, Mousavi MPS, Roth DE, Tsaloglou M-N, Wieringa F, Combs Jr. GF. 2018. Thiamine deficiency disorders: diagnosis, prevalence, and a

roadmap for global control programs. Annals of the New York Academy of Sciences 1430:3-43.

19. Du Q, Wang H, Xie J. 2011. Thiamin (vitamin B1) biosynthesis and regulation: a rich source of antimicrobial drug targets? International journal of biological sciences 7:41-52.
20. Koser SA. 1968. Vitamin requirements of bacteria and yeasts. Springfield, Ill., Thomas, Springfield, Ill.
21. Groenewald M, Boekhout T, Neuvéglise C, Gaillardin C, van Dijck PWM, Wyss M. 2014. *Yarrowia lipolytica*: Safety assessment of an oleaginous yeast with a great industrial potential. Critical Reviews in Microbiology 40:187-206.
22. Markham KA, Alper HS. 2018. Synthetic Biology Expands the Industrial Potential of *Yarrowia lipolytica*. Trends in Biotechnology 36:1085-1095.
23. Hussain MS, Rodriguez GM, Gao D, Spagnuolo M, Gambill L, Blenner M. 2016. Recent advances in bioengineering of the oleaginous yeast *Yarrowia lipolytica*. Aims Bioeng 3:493-514.
24. Ryu S, Hipp J, Trinh CT. 2016. Activating and elucidating metabolism of complex sugars in *Yarrowia lipolytica*. Appl Environ Microbiol 82:1334-1345.
25. Ryu S, Trinh CT. 2018. Understanding functional roles of native pentose-specific transporters for activating dormant pentose metabolism in *Yarrowia lipolytica*. Appl Environ Microbiol 84:e02146-17.
26. Schwartz C, Curtis N, Löbs A-K, Wheeldon I. 2018. Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables *Yarrowia Lipolytica* Growth on Cellobiose. Biotechnology Journal 13:1700584.
27. Walker C, Ryu S, Trinh CT. 2019. Exceptional solvent tolerance in *Yarrowia lipolytica* is enhanced by sterols. Metabolic Engineering 54:83-95.
28. Ryu S, Labbé N, Trinh CT. 2015. Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α-ketoglutaric acid by *Yarrowia lipolytica*. Applied Microbiology and Biotechnology 99:4237-4244.

29. Xu P, Qiao K, Stephanopoulos G. 2017. Engineering oxidative stress defense pathways to build a robust lipid production platform in *Yarrowia lipolytica*. Biotechnology and Bioengineering 114:1521-1530.
30. Zhou J, Zhou H, Du G, Liu L, Chen J. 2010. Screening of a thiamine-auxotrophic yeast for α-ketoglutaric acid overproduction. Letters in Applied Microbiology 51:264-271.
31. Morgunov IG, Kamzolova SV, Perevoznikova OA, Shishkanova NV, Finogenova TV. 2004. Pyruvic acid production by a thiamine auxotroph of *Yarrowia lipolytica*. Process Biochemistry 39:1469-1474.
32. Li M, Petteys BJ, McClure JM, Valsakumar V, Bekiranov S, Frank EL, Smith JS. 2010. Thiamine Biosynthesis in *Saccharomyces cerevisiae* Is Regulated by the NAD^+-Dependent Histone Deacetylase Hst1. Molecular and Cellular Biology 30:3329-3341.
33. Hohmann S, Meacock PA. 1998. Thiamin metabolism and thiamin diphosphate-dependent enzymes in the yeast *Saccharomyces cerevisiae*: genetic regulation. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1385:201-219.
34. Landes N, Gasser B, Vorauer-Uhl K, Lhota G, Mattanovich D, Maurer M. 2016. The vitamin-sensitive promoter PTHI11 enables pre-defined autonomous induction of recombinant protein production in *Pichia pastoris.* Biotechnology and Bioengineering 113:2633-2643.
35. Belén Moreno M, Durán A, Carlos Ribas J. 2000. A family of multifunctional thiamine-repressible expression vectors for fission yeast. Yeast 16:861-872.
36. Basi G, Schmid E, Maundrell K. 1993. TATA box mutations in the *Schizosaccharomyces pombe* nmt1 promoter affect transcription efficiency but not the transcription start point or thiamine repressibility. Gene 123:131-136.
37. Maundrell K. 1990. nmt1 of fission yeast. A highly transcribed gene completely repressed by thiamine. Journal of Biological Chemistry 265:10857-64.
38. Müller S, Sandal T, Kamp-Hansen P, Dalbøge H. 1998. Comparison of expression systems in the yeasts *Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe* and *Yarrowia lipolytica.* Cloning of two novel promoters from *Yarrowia lipolytica*. Yeast 14:1267-1283.

39. Damude HG, Gillies PJ, Macool DJ, Picataggio SK, Pollak DMW, Ragghianti JJ, Xue Z, Yadav NS, Zhang H, Zhu QQ. 2011. High eicosapentaenoic acid producing strains of *Yarrowia lipolytica*. Google Patents.
40. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. Science 330:1099-1102.
41. Klumpp S, Zhang Z, Hwa T. 2009. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. Cell 139:1366-1375.
42. Rodríguez-Navarro S, Llorente B, Rodríguez-Manzaneque MT, Ramne A, Uber G, Marchesan D, Dujon B, Herrero E, Sunnerhagen P, Pérez-Ortín JE. 2002. Functional analysis of yeast gene families involved in metabolism of vitamins B1 and B6. Yeast 19:1261-1276.
43. Maundrell K. 1993. Thiamine-repressible expression vectors pREP and pRIP for fission yeast. Gene 123:127-130.
44. Lonsdale D. 2006. A Review of the Biochemistry, Metabolism and Clinical Benefits of Thiamin(e) and Its Derivatives. Evidence-Based Complementary and Alternative Medicine 3:49-59.
45. Wolak N, Kowalska E, Kozik A, Rapala-Kozik M. 2014. Thiamine increases the resistance of baker's yeast *Saccharomyces cerevisiae* against oxidative, osmotic and thermal stress, through mechanisms partly independent of thiamine diphosphate-bound enzymes. FEMS Yeast Research 14:1249-1262.
46. Singleton CK, Martin PR. 2001. Molecular Mechanisms of Thiamine Utilization. Current Molecular Medicine 1:197-207.
47. Dimmer KS, Fritz S, Fuchs F, Messerschmitt M, Weinbach N, Neupert W, Westermann B. 2002. Genetic Basis of Mitochondrial Function and Morphology in *Saccharomyces cerevisiae*. Molecular Biology of the Cell 13:847-853.
48. Fattal-Valevski A. 2011. Thiamine (Vitamin B1). Journal of Evidence-Based Complementary & Alternative Medicine 16:12-20.

49. Sundström M, Lindqvist Y, Schneider G, Hellman U, Ronne H. 1993. Yeast TKL1 gene encodes a transketolase that is required for efficient glycolysis and biosynthesis of aromatic amino acids. Journal of Biological Chemistry 268:24346-24352.
50. Kispal G, Steiner H, Court DA, Rolinski B, Lill R. 1996. Mitochondrial and Cytosolic Branched-chain Amino Acid Transaminases from Yeast, Homologs of the myc Oncogene-regulated Eca39 Protein. Journal of Biological Chemistry 271:24458-24464.
51. Shabbir Hussain M, Gambill L, Smith S, Blenner MA. 2016. Engineering Promoter Architecture in Oleaginous Yeast *Yarrowia lipolytica*. ACS Synthetic Biology 5:213-223.
52. Trassaert M, Vandermies M, Carly F, Denies O, Thomas S, Fickers P, Nicaud J-M. 2017. New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. Microbial Cell Factories 16:141.
53. Blazeck J, Liu L, Redden H, Alper H. 2011. Tuning Gene Expression in *Yarrowia lipolytica* by a Hybrid Promoter Approach. Applied and Environmental Microbiology 77:7905-7914.
54. Nicaud J-M, Madzak C, van den Broek P, Gysler C, Duboc P, Niederberger P, Gaillardin C. 2002. Protein expression and secretion in the yeast *Yarrowia lipolytica*. FEMS Yeast Research 2:371-379.
55. Tikka C, Osuru HP, Atluri N, Raghavulu PCV, Yellapu NK, Mannur IS, Prasad UV, Aluru S, K NV, Bhaskar M. 2013. Isolation and characterization of ethanol tolerant yeast strains. Bioinformation 9:421-425.
56. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth 6:343-345.
57. Yamane T, Sakai H, Nagahama K, Ogawa T, Matsuoka M. 2008. Dissection of centromeric DNA from yeast *Yarrowia lipolytica* and identification of protein-binding site required for plasmid transmission. Journal of Bioscience and Bioengineering 105:571-578.

58. Moreno MB, Duran A, Ribas JC. 2000. A family of multifunctional thiamine-repressible expression vectors for fission yeast. Yeast 16:861-72.
59. Wang JH, Hung WP, Tsai SH. 2011. High efficiency transformation by electroporation of *Yarrowia lipolytica*. J Microbiol 49:469-472.
60. Clarkson SM, Giannone RJ, Kridelbaugh DM, Elkins JG, Guss AM, Michener JK. 2017. Construction and Optimization of a Heterologous Pathway for Protocatechuate Catabolism in *Escherichia coli* Enables Bioconversion of Model Aromatic Compounds. Applied and Environmental Microbiology 83:e01313-17.
61. Villalobos Solis MI, Giannone RJ, Hettich RL, Abraham PE. 2019. Exploiting the Dynamic Relationship between Peptide Separation Quality and Peptide Coisolation in a Multiple-Peptide Matches-per-Spectrum Approach Offers a Strategy To Optimize Bottom-Up Proteomics Throughput and Depth. Analytical Chemistry 91:7273-7279.
62. Dorfer V, Maltsev S, Winkler S, Mechtler K. 2018. CharmeRT: Boosting Peptide Identifications by Chimeric Spectra Identification and Retention Time Prediction. Journal of Proteome Research 17:2581-2589.
63. Taverner T, Karpievitch YV, Polpitiya AD, Brown JN, Dabney AR, Anderson GA, Smith RD. 2012. DanteR: an extensible R-based tool for quantitative analysis of -omics data. Bioinformatics 28:2404-2406.
64. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods 13:731.
65. Mukhopadhyay A, He Z, Alm EJ, Arkin AP, Baidoo EE, Borglin SC, Chen W, Hazen TC, He Q, Holman H-Y, Huang K, Huang R, Joyner DC, Katz N, Keller M, Oeller P, Redding A, Sun J, Wall J, Wei J, Yang Z, Yen H-C, Zhou J, Keasling JD. 2006. Salt Stress in *Desulfovibrio vulgaris*: an Integrated Genomics Approach. Journal of Bacteriology 188:4068-4078.
66. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics (Oxford, England) 25:1091-1093.

67. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402.
68. Delic M, Mattanovich D, Gasser B. 2013. Repressible promoters – A novel tool to generate conditional mutants in *Pichia pastoris*. Microbial Cell Factories 12:6.
69. Manetti AG, Rosetto M, Maundrell KG. 1994. nmt2 of fission yeast: a second thiamine-repressible gene co-ordinately regulated with nmt1. Yeast 10:1075-82.
70. Praekelt UM, Byrne KL, Meacock PA. 1994. Regulation of THI4 (MOL1), a thiamine-biosynthetic gene of *Saccharomyces cerevisiae*. Yeast 10:481-90.
71. Klapper M, Ehmke M, Palgunow D, Böhme M, Matthäus C, Bergner G, Dietzek B, Popp J, Döring F. 2011. Fluorescence-based fixative and vital staining of lipid droplets in *Caenorhabditis elegans* reveal fat stores using microscopy and flow cytometry approaches. Journal of Lipid Research 52:1281-1293.

Appendix

Table 2-1 List of plasmids and strains.

Plasmids/strains	Description	Source
Plasmids		
pSL16-CEN1-1-227	pSL16-CEN1-1-227	(54)
pSR001	pSL16-P_{TEF}-T_{CYC1}::leu2	(26)
pSR008	pSL16-P_{TEF}-T_{CYC1}::ura3	(23)
pSR005	pSL16-hrGFP-T_{CYC1}::leu2	this study
pAT32y	pSL16-P_{TEF}-hrGFP-T_{CYC1}::leu2	this study
pSR068	pSL16-P_{NMT1}-hrGFP::leu2	this study
pSR071	pSL16-P_{P1}-hrGFP::leu2	this study
pSR072	pSL16-P_{P2}-hrGFP::leu2	this study
pSR073	pSL16-P_{P3}-hrGFP::leu2	this study
pSR074	pSL16-P_{TEF}-THI13sce-T_{CYC1}::ura3	this study
pSR075	pSL16-P_{P3}-THI13sce-T_{CYC1}::ura3	this study
Yeast strains		
YlSR001	MATA ura3-302 leu2-270 xpr2-322 axp2-deltaNU49 XPR2::SUC2	ATCC MYA-2613
YlSR101	YlSR001 + pSR001	(26)
YlSR108	YlSR001 + pSR008	(23)
YlSR109	YlSR001 + pAT32y	this study
YlSR1001	YlSR001 + pSR068	this study
YlSR1002	YlSR001 + pSR071	this study
YlSR1003	YlSR001 + pSR072	this study
YlSR1004	YlSR001 + pSR073	this study
YlSR1005	YlSR001 + pSR074	this study
YlSR1006	YlSR001 + pSR075	this study

Table 2-2 List of primers.

Genes	Primers	Sequences
Primers for plasmid construction		
pSL16	pSL16_Fwd	GCCTGCACGAGTGGGTGTAATCATGTAATTAGTTATGTCACGCTTAC
	pSL16_Rev	CAGGATCTGCTTGCTCACCATAGATCTGTTCGGAAATCAACGG
hrGFP	hrGFP_Fwd	AATCGGTTGAGCATCCGTTGATTTCCGAACAGATCTATGGTGAGCAAGCAGATCCTG
	hrGFP_Rev	GTAAGCGTGACATAACTAATTACATGATTACACCCACTCGTGCAGG
P_{TEF}	P_{TEF}_Fwd	CATCCGTTGATTTCCGAACAGATCTAGAGACCGGGTTGGCGGCGTATTTG
	P_{TEF}_Rev	TTCAGGATCTGCTTGCTCACCATTTTGAATGATTCTTATACTCAGAAGGAAATGCTTAAC
P_{NMT1}	P_{NMT1}_Fwd	GCATCCGTTGATTTCCGAACAGATCTTTGTATTTCAAAGGACATAATCTAAAATAATAAC
	P_{NMT1}_Rev	GGTGTTCTTCAGGATCTGCTTGCTCACCATGATTTAACAAAGCGACTATAAGTCAGAAAG
P_{P1}	P_{P1}_Fwd	GGTTGAGCATCCGTTGATTTCCGAACAGATCTTGAAGTGGGTGAGTCGCCAATTATTC
	P_{P1}_Rev	CAGGCCGGTGTTCTTCAGGATCTGCTTGCTCACCATGATCGAATTGAGTCAGCGACG
P_{P2}	P_{P2}_Fwd	CGGTTGAGCATCCGTTGATTTCCGAACAGATCTCAGGTGGTAGCAGCCCAAGACAATG
	P_{P2}_Rev	GGTGTTCTTCAGGATCTGCTTGCTCACCATGAATTGACGAACAGGTGTTTTGATG
P_{P3}	P_{P3}_Fwd	CGGTTGAGCATCCGTTGATTTCCGAACAGATCTGAGGGGTAGTCGTAAGTTTCATC
	P_{P3}_Rev	CGGTGTTCTTCAGGATCTGCTTGCTCACCATGTTAATTGTAGGTGATATAAGGGGAAG
pSR005	pSR005_Fwd	ATGGTGAGCAAGCAGATCCTG
	pSR005_Rev	AGATCTGTTCGGAAATCAACGGATGCTCAAC
$THI13_{sce}$	$THI13_{sce}$_Fwd[1]	CATTTCCTTCTGAGTATAAGAATCATTCAAAATGTCTACAGACAAGATCACATTTTTG

Table 2- 2 continued

	$THI13_{sce}$_Fwd[2]	CACCCTTCCCCTTATATCACCTACAATTAACATGTCTACAGACAAGATCACATTTTTG
	$THI13_{sce}$_Rev	GAATGTAAGCGTGACATAACTAATTACATGATTAAGCTGGAAGAGCCAATCTCTTG
pSR008	pSR008_Fwd	TCATGTAATTAGTTATGTCACGCTTAC
	pSR008_Rev	TTTGAATGATTCTTATACTCAGAAG
pSR073	pSR073_Rev	GTTAATTGTAGGTGATATAAGGGGAAGG
Primers for checking and sequencing		
hrGFP_seq	hrGFP_Rev	CTTGCCGCAGCCCTCCATGG
TEF_seq	TEF(-100)_Fwd	CACCGTCCCCGAATTACCTTTC
P3_seq	P3(-80)_Fwd	CTGCCGTAAATCACATACTGTCGGCTG
Primers for rt-PCR		
Actin	Actin rt_Fwd	TCCAGGCCGTCCTCTCCC
	Actin rt_Ref	GGCCAGCCATATCGAGTCGCA
P1	P1 rt_Fwd	AGGACAAGGAGCCTGCCAAG
	P1 rt_Rev	GGAGGCAATGGCAGAGGCTA
P2	P2 rt_Fwd	TATGCAATCGGCCTCACCGA
	P2 rt_Rev	CTTGCCCTCCAGCTGGTCTT
P3	P3 rt_Fwd	GCTGGCTCCTGTGGTCTCTC
	P3 rt_Rev	GAACTGCTCGGCAGGCTTTC

Table 2- 3. Protein expression of thiamine-dependent enzymes at both exponential-growth time points.

Pathway	Enzyme	Subunit	Locus tag	Fold change (time 1)	Fold change (time 2)
Glycolysis	PDH	E1	YALI0E27005g	-0.21	-0.43
		E1	YALI0F20702g	0.02	-0.33
		E2	YALI0D23683g	-0.20	-0.13
		E3	YALI0D20768g	1.62	2.34
Citrate cycle	KGDH	E1	YALI0E33517g	0.92	-0.64
		E2	YALI0E16929g	0.88	-2.26
		E3	YALI0D20768g	1.62	2.34
PPP	TKL	N/A	YALI0E06479g	0.82	-0.34
Leucine, isoleucine and valine metabolism	BCKDC	E1-alpha	YALI0D08690g	2.84	1.58
		E1-beta	YALI0F05038g	2.81	0.79
		E2	YALI0D23815g	3.45	2.05
		E3	YALI0D20768g	1.62	2.34
	AHAS	small	YALI0C09636g	-0.63	0
		large	YALI0C00253g	-0.31	-0.89

Table 2- 4. List of putative thiamine-regulated genes in *Y. lipolytica*. Abbreviations: PP: *P. pastoris*, SC: *S. cerevisiae*, and SP: *S. pombe*.

Genes	Functions	*Orthologs*
thi11$_{PP}$	4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase	YALI0E04224g (P1)
nmt1$_{SP}$	4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase	YALI0C14652g (P2)
nmt2$_{SP}$	thiamine thiazole synthase	YALI0A09768g (P3)
thi4$_{SC}$	thiamine thiazole synthase	YALI0A09768g (P3)

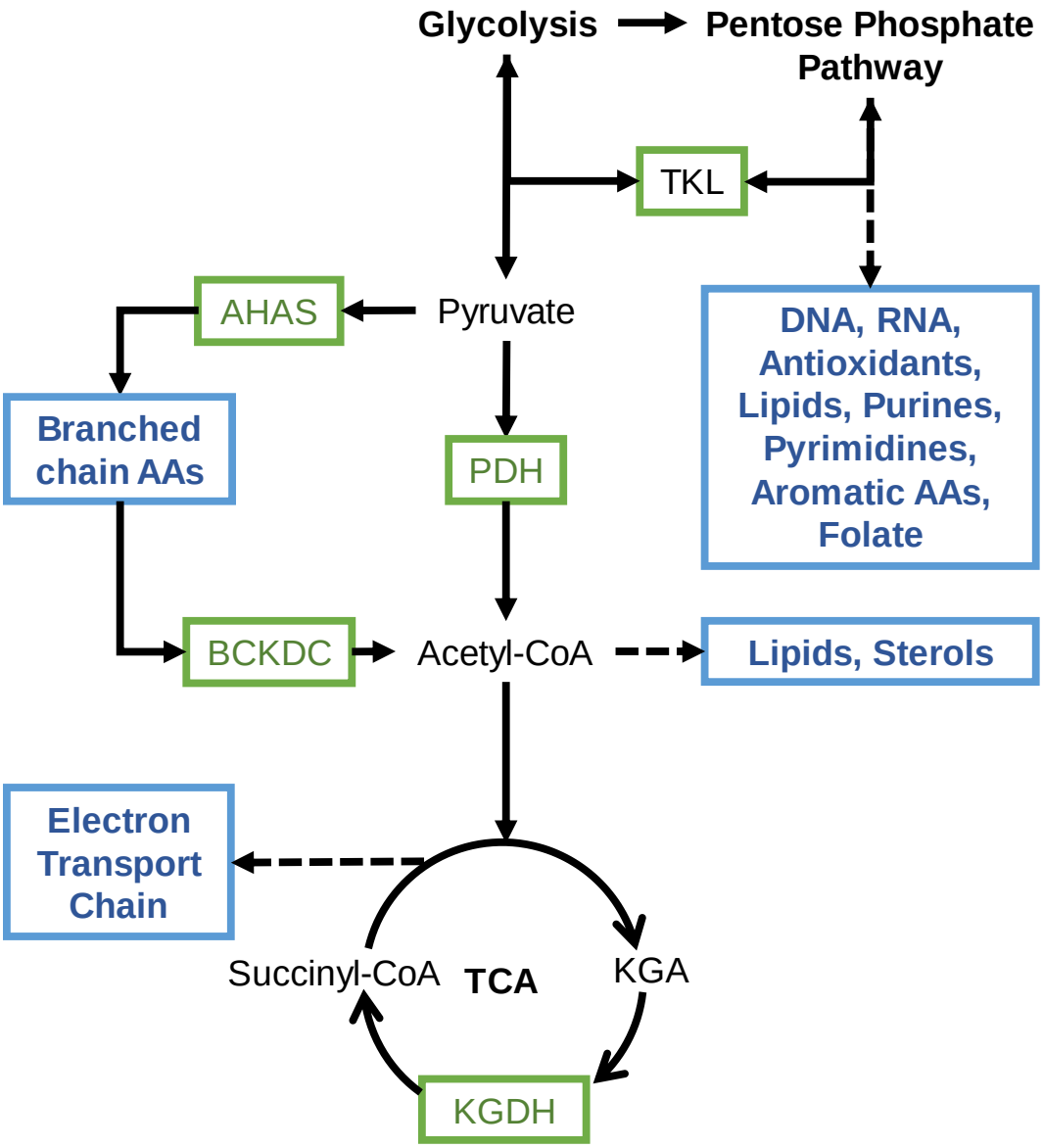


Figure 2- 1. Metabolic map of thiamine-dependent enzymes (green) in relationship to central (black) and peripheral (blue) pathways. Abbreviations: TPP: thiamine pyrophosphate dependent enzymes, PDH: pyruvate dehydrogenase enzyme complex, KGDH: a-ketoglutarate dehydrogenase, TKL: transketolase, AHAS: acetolactate synthase, BCKDC: branched chain α-ketoacid dehydrogenase, TPP: thiamine pyrophosphate, AAs: amino acids, TCA: citrate (Krebs) cycle, KGA: alpha-ketoglutarate.dehydrogenase.

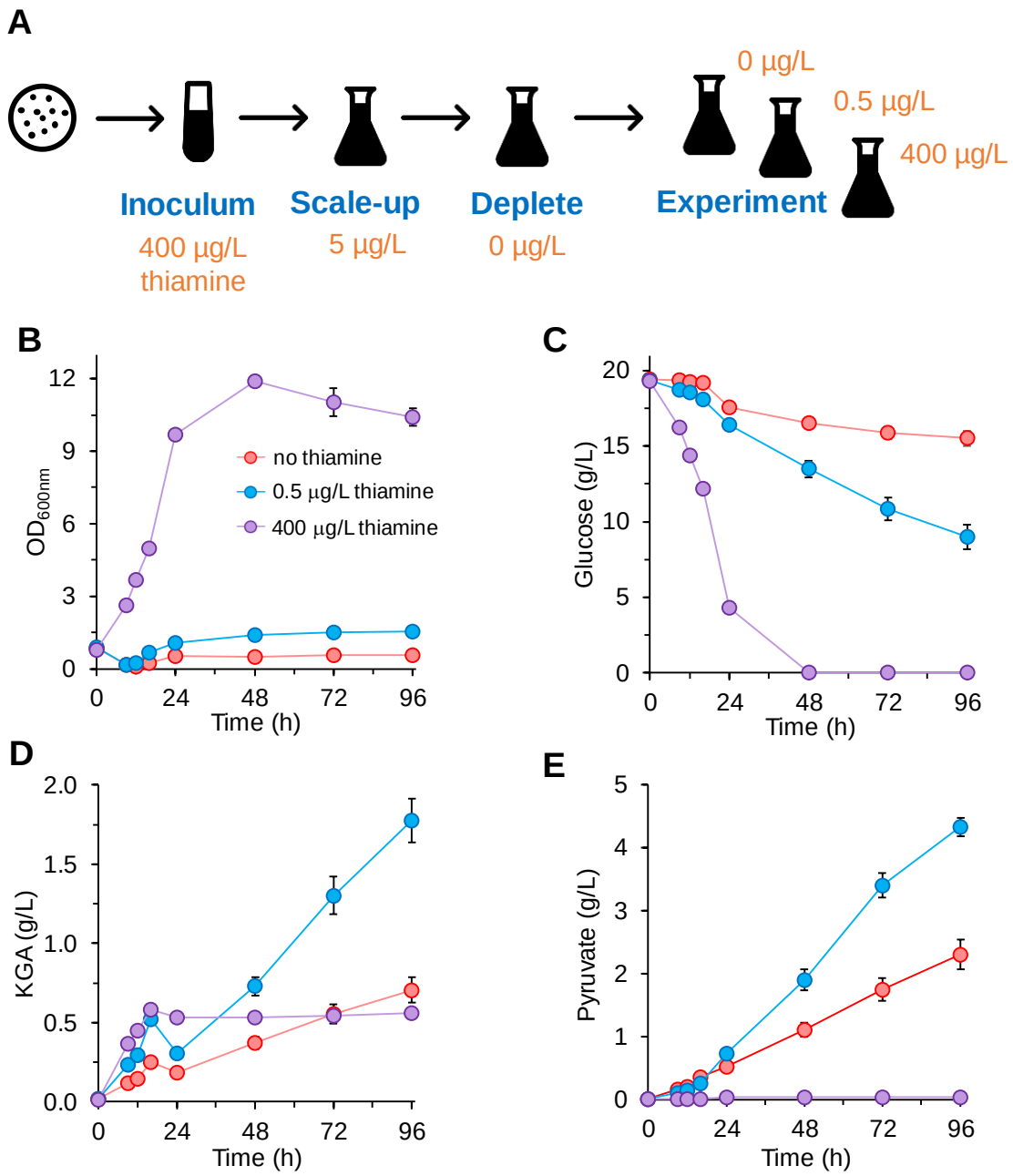
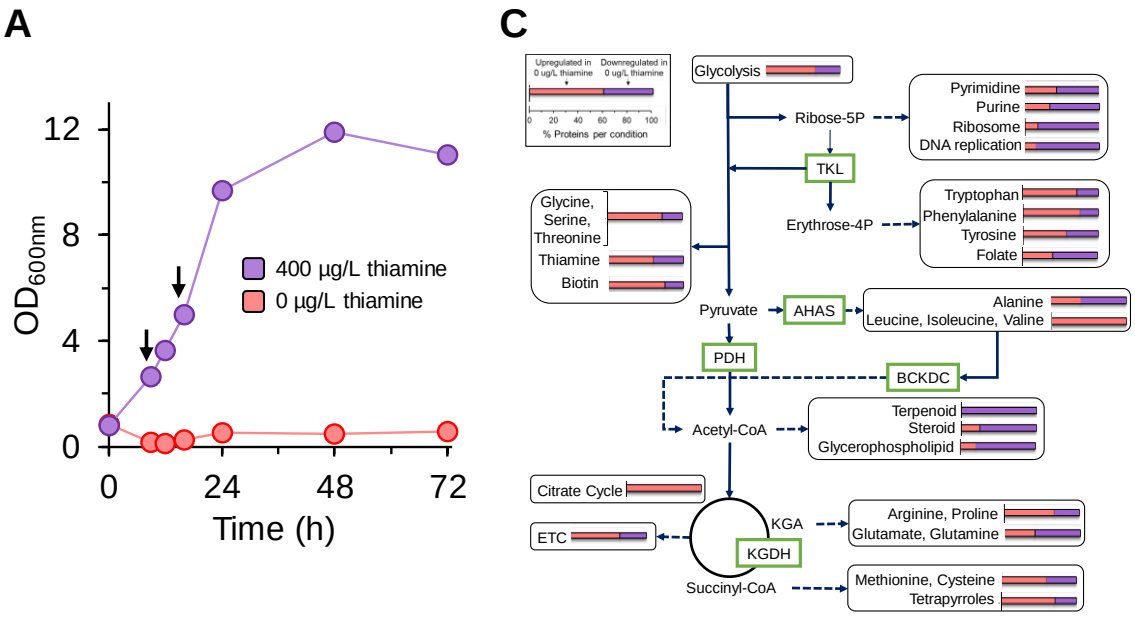
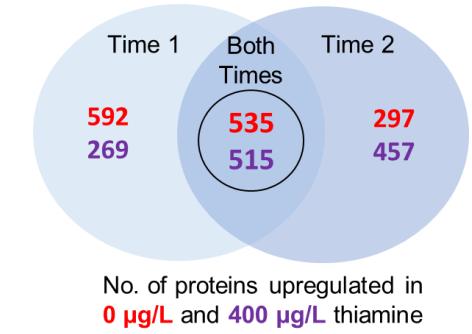
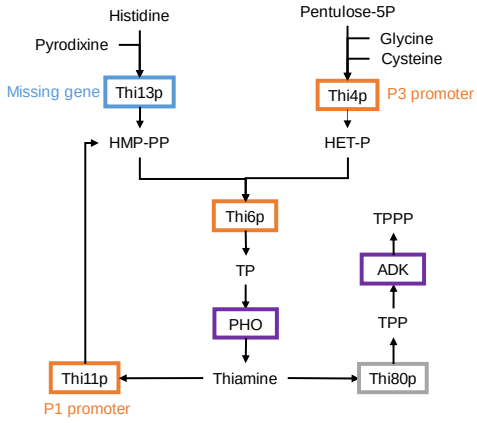


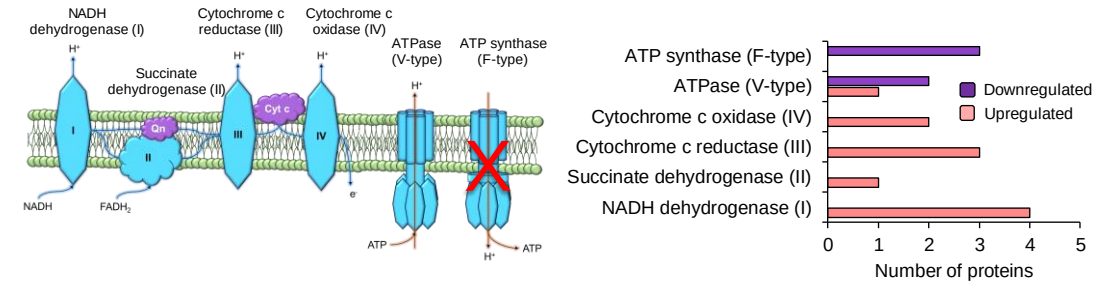
Figure 2- 2. Growth characterization of the thiamine-auxotrophic *Y. lipolytica* YISR001 in 0 (red), 0.5 (blue) and 400 (purple) µg/L thiamine. (A) Scheme of thiamine depletion design experiments. (B) Cell growth profiles. (C) Glucose consumption profiles. (D) KGA production profiles. (E) Pyruvate production profiles.

Figure 2- 3. Proteomics of the thiamine-auxotrophic *Y. lipolytica* YLSR001 grown in 0 (red) and 400 (purple) µg/L thiamine. (**A**) Growth profiles and proteomics samples indicated by arrows. (**B**) Venn diagram representing upregulated and downregulated proteins at both exponential time points. (**C**) Thiamine-responsive proteins on metabolic pathways. Bar graphs represent the percentage of proteins upregulated in 0 (red) or 400 (purple) µg/L thiamine for each pathway (**D**) Electron transport chain and ATPase. (**E**) Thiamine metabolism of *Y. lipolytica*. Enzymes: Thi4P (YALI0A09768p, thiamine thiazole synthase), Thi11p (YALI0E04224p, 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase), Thi6p (YALI0C15554p, bifunctional hydroxyethylthiazole kinase/thiamine-phosphate diphosphorylase), PHO (YALI0A12573p, acid phosphatase), Thi80p (YALI0E21351p, thiamine diphosphokinase), Thi13p (4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase), and ADK (YALI0F26521p, adenylate kinase). Abbreviations: PDH: pyruvate dehydrogenase enzyme complex, KGDH: a-ketoglutarate dehydrogenase, TKL: transketolase, AHAS: acetolactate synthase, BCKDC: branched chain α-ketoacid dehydrogenase, TPP: Thiamine pyrophosphate, HET-P: hydroxyethylthiazole phosphate, HMP-PP: hydroxymethylpyrimidine pyrophosphate, TP: thiamine monophosphate, TPP: thiamine pyrophosphate, TPPP: thiamine triphosphate.


A
12
8
4
0
OD600nm
0
24
48
72
Time (h)
400 µg/L thiamine
0 µg/L thiamine
C
Upregulated in 0 ug/L thiamine
Downregulated in 0 ug/L thiamine
0 20 40 60 80 100
% Proteins per condition
Glycolysis
Ribose-5P
Pyrimidine
Purine
Ribosome
DNA replication
TKL
Erythrose-4P
Tryptophan
Phenylalanine
Tyrosine
Folate
Glycine, Serine, Threonine
Thiamine
Biotin
Pyruvate
AHAS
Alanine
Leucine, Isoleucine, Valine
PDH
BCKDC
Acetyl-CoA
Terpenoid
Steroid
Glycerophospholipid
Citrate Cycle
ETC
KGA
KGDH
Arginine, Proline
Glutamate, Glutamine
Succinyl-CoA
Methionine, Cysteine
Tetrapyrroles


B
Time 1
Both Times
Time 2
592
269
535
515
297
457
No. of proteins upregulated in
0 µg/L and 400 µg/L thiamine


E
Histidine
Pyrodixine
Missing gene
Thi13p
HMP-PP
Pentulose-5P
Glycine
Cysteine
Thi4p
P3 promoter
HET-P
Thi6p
TP
PHO
Thiamine
TPPP
ADK
TPP
Thi80p
Thi11p
P1 promoter


D
NADH dehydrogenase (I)
Succinate dehydrogenase (II)
Cytochrome c reductase (III)
Cytochrome c oxidase (IV)
ATPase (V-type)
ATP synthase (F-type)
H+
Qn
Cyt c
NADH
FADH2
e-
ATP
ATP synthase (F-type)
ATPase (V-type)
Cytochrome c oxidase (IV)
Cytochrome c reductase (III)
Succinate dehydrogenase (II)
NADH dehydrogenase (I)
Downregulated
Upregulated
0 1 2 3 4 5
Number of proteins

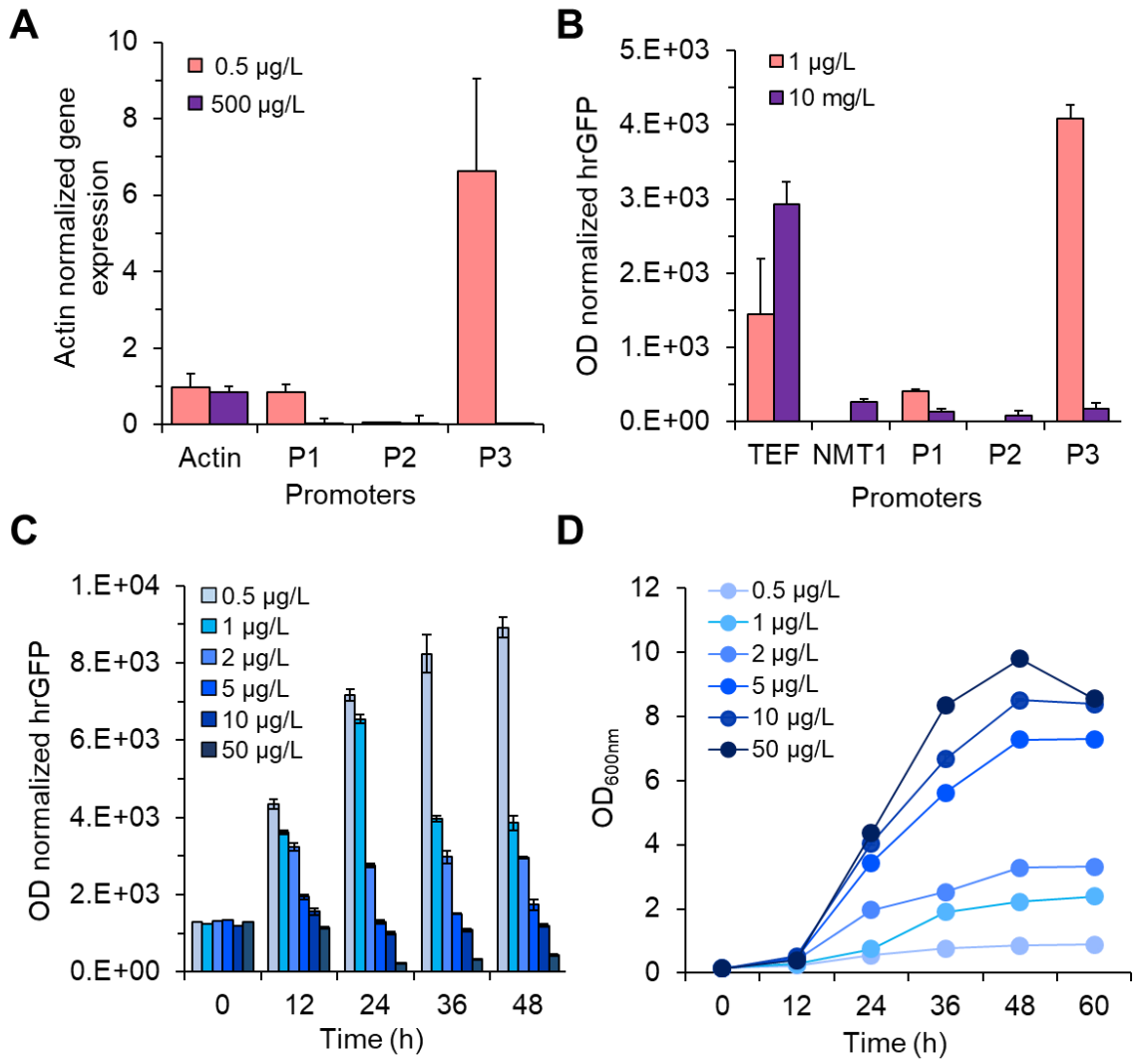


Figure 2- 4. Thiamine-responsive promoter characterization. (**A**) Chromosomal gene expression of thiamine-responsive genes using rt-PCR. (**B**) hrGFP protein expression. (**C, D**) Sensitivity of thiamine-responsive promoters in presence of increasing thiamine concentrations.

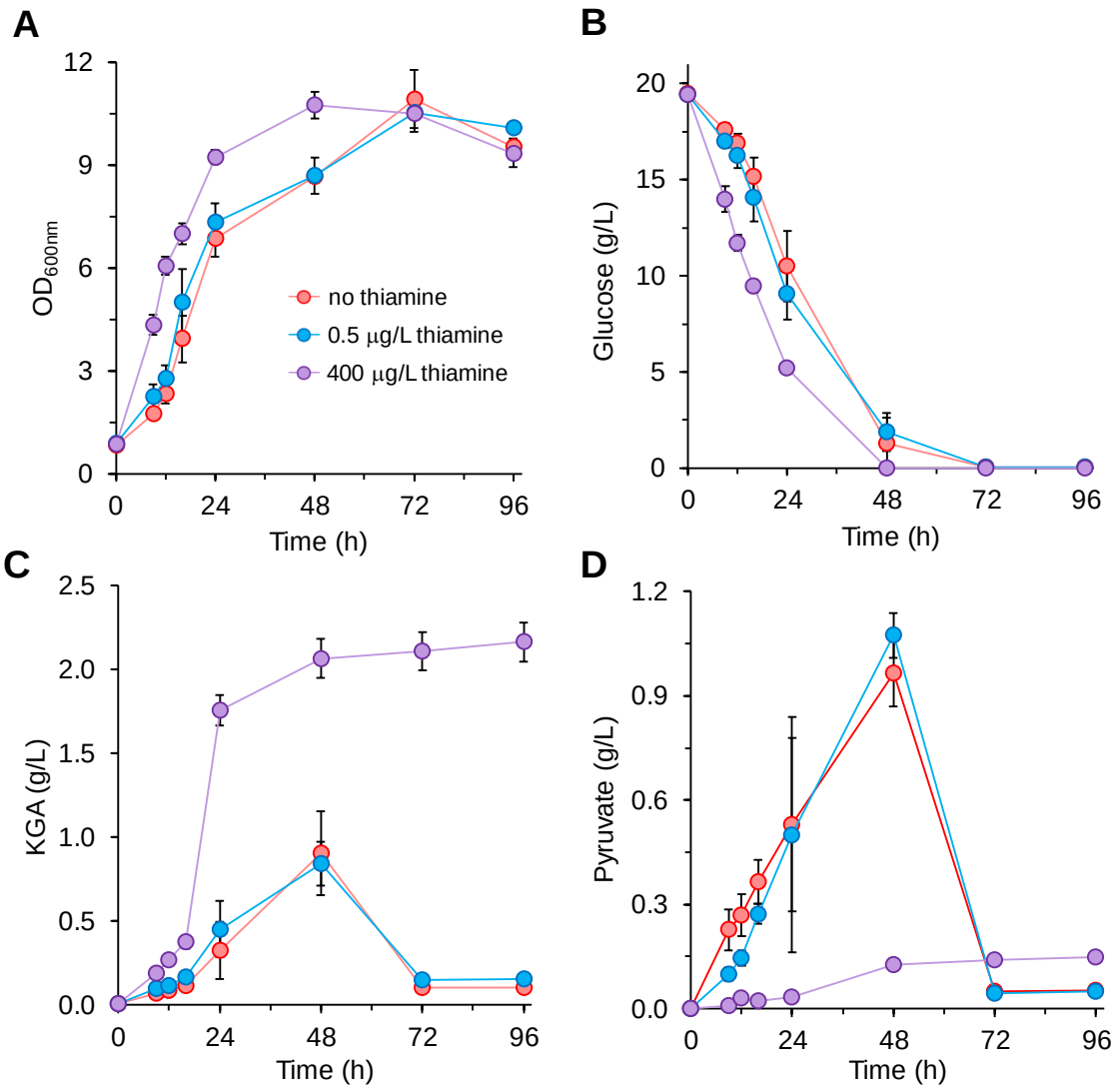


Figure 2- 5. Expression of pP3-scTHI13 in YlSR1006 restores the *de novo* thiamine biosynthesis in *Y. lipolytica*. (A) Cell growth profiles. (B) Glucose consumption profiles. (C) KGA production profiles. (D) Pyruvate production profiles. Growth characterization was conducted in 0 (red), 0.5 (blue) and 400 (purple) µg/L thiamine. Abbreviations: KGA, alpha-ketoglutarate.

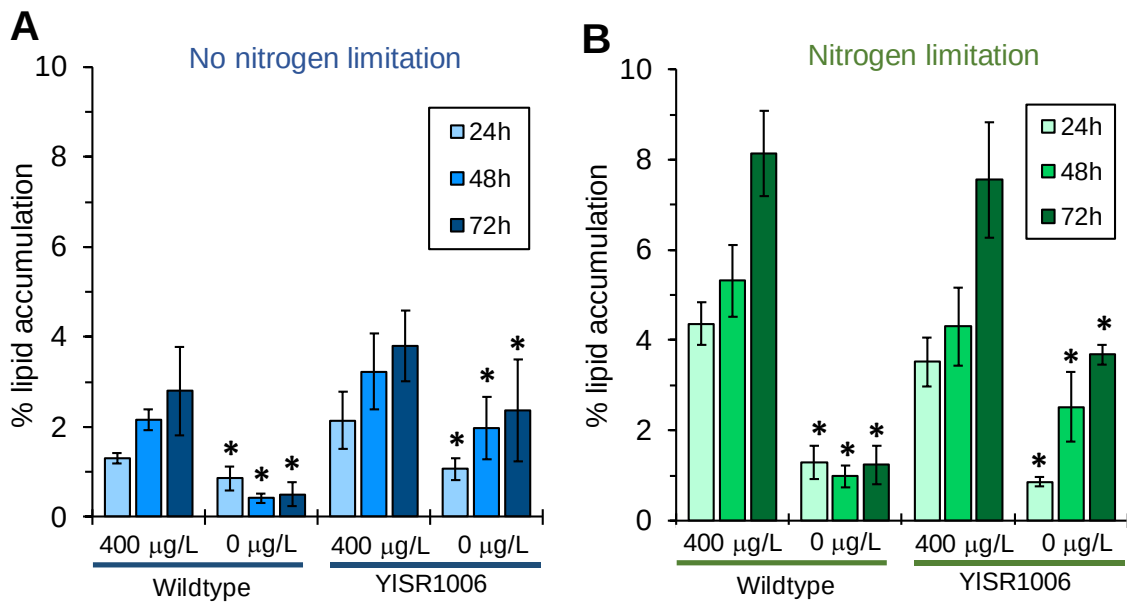


Figure 2- 6. Lipid production is influenced by thiamine availability. (**A**) Lipid accumulation profiles for the thiamine-auxotrophic wildtype YlSR001 and thiamine-prototrophic strain YlSR1006 in 0 and 400 µg/L thiamine. (**B**) Lipid accumulation profiles for thiamine-auxotrophic and prototrophic strains in lipid production (C:N = 100) media with 0 and 400 µg/L thiamine. Statistical significance was calculated with one-way analysis of variance (ANOVA) with Holm-Sidak correction between 0 and 400 µg/L thiamine for each strain respectively. Symbols: "*": p-value < 0.05.

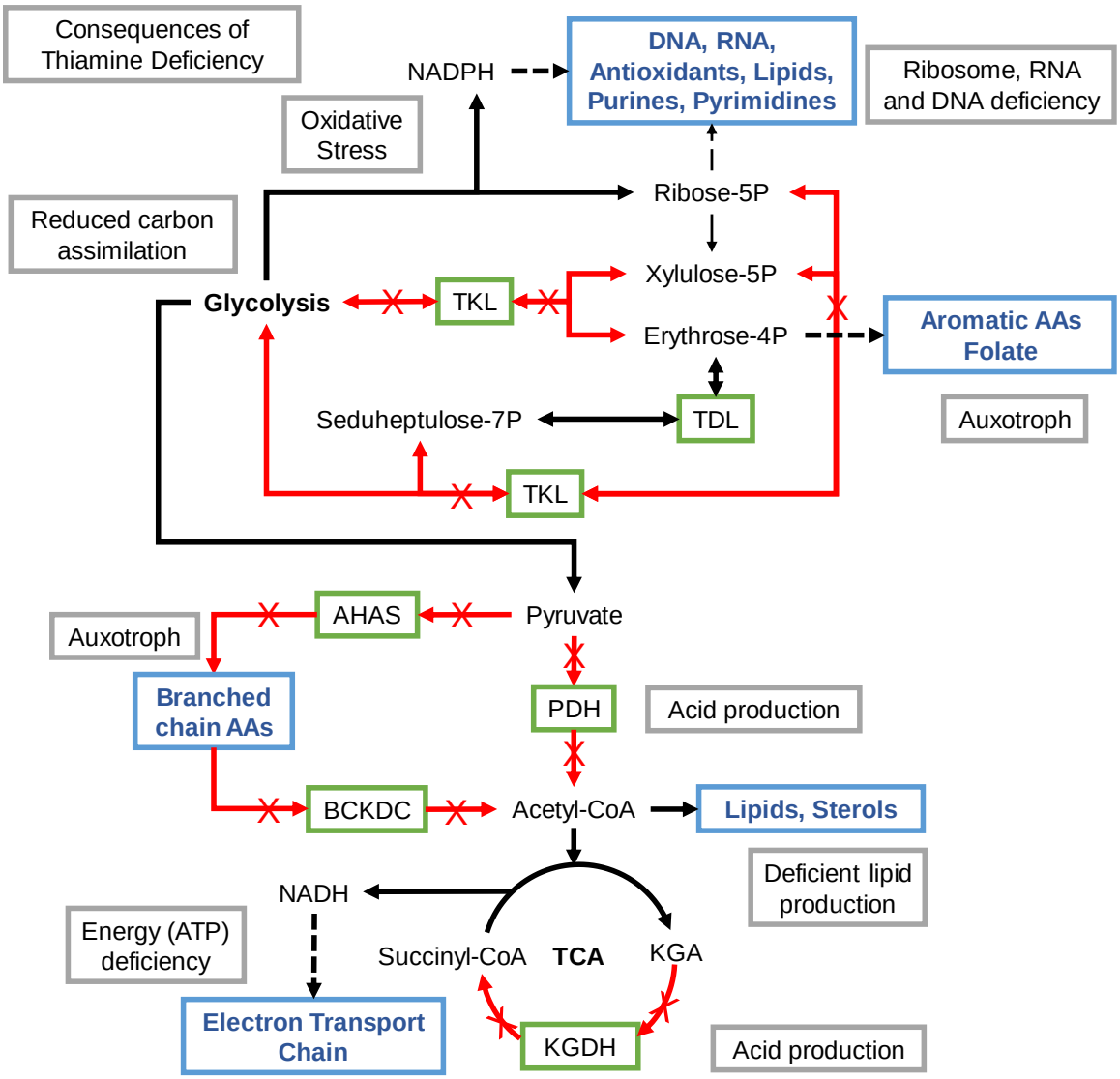


Figure 2- 7. A comprehensive model explains detrimental consequences of thiamine deficiency on metabolism and cellular growth of *Y. lipolytica*.

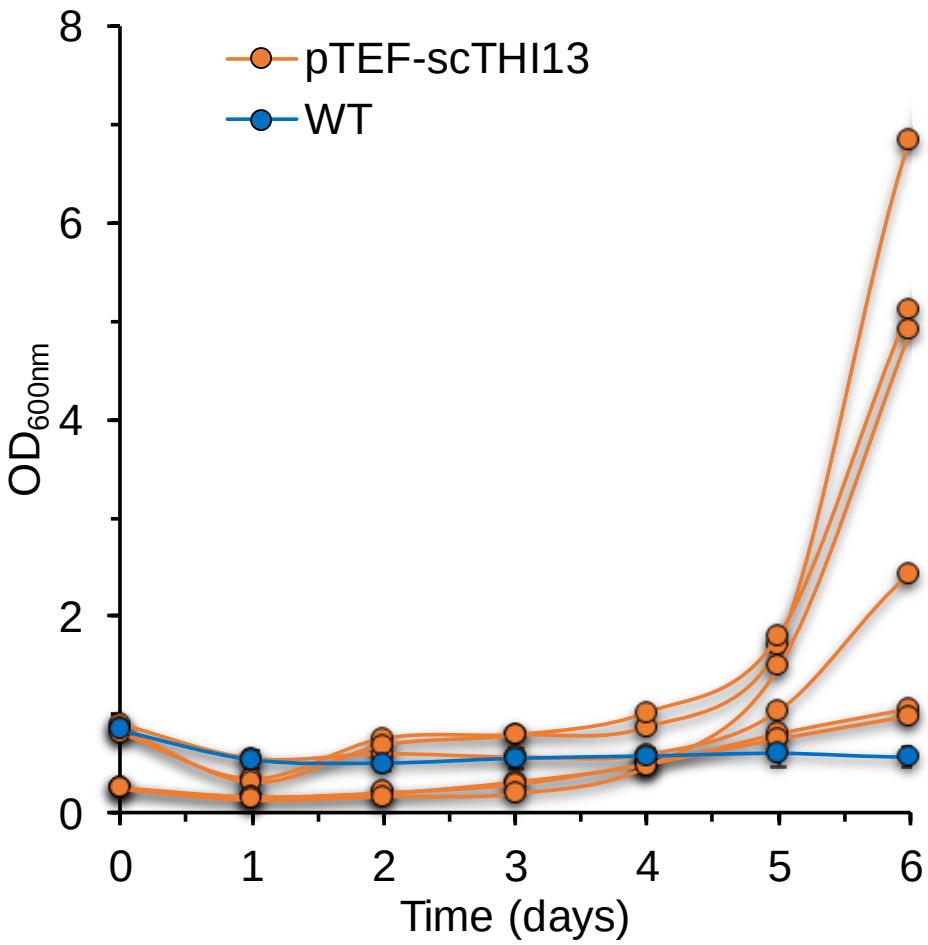


Figure 2- 8. Constitutive expression of missing thiamine gene does not effectively restore thiamine prototrophy in Y. lipolytica. Wildtype is in blue and 6 replicates of YlSR005 growth without thiamine (0 µg/L thiamine) in orange.

CHAPTER III

Co-expression Connectivity Predicts Genetic Targets Underlying Solvent Tolerance in *Yarrowia lipolytica*

CTT conceived and supervised the study. CW and CTT designed the experiments and analyzed the data. CW performed the experiments. CW and CTT wrote and approved the manuscript.

Abstract

Microbial tolerance to solvents such as ionic liquids (ILs) is a robust phenotype beneficial for effective bioconversion of renewable substrates to chemicals and fuels. While most microbes become inhibited in 1%-5% (v/v) IL such as 1-ethyl-3-methylimidazolium acetate ([EMIM][OAc]), we have recently created a robust *Yarrowia lipolytica* (YlCW001) that can tolerate up to 18% (v/v) IL. Currently, genotypes conferring this high IL tolerance in YlCW001 still remain to be discovered. By using dynamic RNA-sequencing data, we formulated a novel co-expression connectivity metric capable of identifying key processes affected by IL and genes conferring high IL-tolerance that might be overlooked by traditional differential expression approaches. Upon exposure to IL, we found that the wildtype increased expression of genes involved in mitochondrion process while the mutant (YlCW001) upregulated genes associated with lipid metabolism. From the genes upregulated by the mutant in IL, genetic targets were selected from genes upregulated by the mutant in IL based on fold change, co-expression connectivity and gene ontology associations. Upon characterizing these genetic targets using single-gene overexpression screening and dual-gene library enrichment selection, we identified the best combination of non-intuitive genotypes conferring IL-tolerance. By validating these genotypes in three *Y. lipolytica* isolates, we demonstrate this strategy is effective for identification of genotypes underlying complex phenotypes.

Keywords: *Yarrowia lipolytica*; ionic liquid; solvent tolerance; resequencing; dynamic RNA-sequencing; reverse engineering; co-expression connectivity; genetic targets.

Introduction

Biocatalysis in organic solvents has been widely studied for decades, offering a plethora of unique strategies for substrate solubilization (1), enhanced enzymatic activity (2), and product recovery (3). More recently, the use of ionic liquids (ILs) has emerged to replace organic solvents in bioprocessing (4). Considered as green solvents, ILs offer a novel

reaction medium for biotransformation with superior results in comparison to conventional organic solvents due to their ability to dissolve a wider range of compounds and their adjustable properties for enzyme stabilization and activation (e.g., lipases, alcohol dehydrogenases, proteases, and oxidoreductases) (5). Importantly, ILs are a promising biomass pretreatment technology that effectively reduces recalcitrance, achieving efficient dissolution of lignocellulosic biomass and permitting enzymatic accessibility to the sugar polymers (i.e., cellulose and hemicellulose) for saccharification (6-8). Fermentation of the sugars released by IL pretreated biomass could be directly assimilated by microbial cell factories if not for the toxic properties ILs impose on microorganisms (9, 10). For these reasons, microbial high solvent tolerance is a desirable phenotype for the conversion of renewable feedstocks to replace petroleum derived chemicals and fuels.

While most studies have focused on isolated enzymes, whole cell biotransformation in ILs offers greater diversity of chemical products and biofuels with improved economics due to the high cost of purified enzymes. Preceding screens for IL-tolerant microorganisms identified the non-conventional oleaginous yeast, *Yarrowia lipolytica*, as one of the top performers in benchmark IL, 1-ethyl-3-methylimidazolium acetate ([EMIM][OAc]) (11). We further demonstrated the use of wildtype *Y. lipolytica* in IL-pretreated cellulose by achieving 92% of the theoretical yield of alpha ketoglutaric acid (KGA) in 10% (v/v) [EMIM][OAc] (12). More recently, we employed adaptive laboratory evolution (ALE) to generate a platform strain with enhanced activity in high concentrations of IL (18% (v/v) [EMIM][OAc]), which also outperformed wildtype *Y. lipolytica* in all imidazolium-based ILs tested (13). We discovered the IL-robust mutant's upregulation of steroid biosynthesis gene expression and demonstrated sterols are critical to this phenotype by overexpressing the steroid biosynthesis transcription factor and enhancing the performance of wildtype *Y. lipolytica* in IL (13).

In this study, we aim to better understand the IL-tolerant phenotype of the mutant strain using dynamic (i.e., time series) RNA-sequencing to identify key genotype(s) conferring IL-tolerance. To effectively utilize dynamic transcriptomics data, we showcase a new metric to facilitate selection of genetic targets and reduce the number of candidate

genes for reverse engineering. Specifically, we investigate the efficacy of identifying genetic targets conferring IL-tolerance by selecting the most connected genes from a co-expression network of upregulated genes in the IL-tolerant MT. By single and dual overexpression of genes selected by fold change, connectivity, and ontology, we show that connectivity is a promising metric to reveal key genotype(s) behind complex phenotype(s) and use these genetic targets to reverse engineer high solvent-tolerance in the *Y. lipolytica* clade.

Results

Genotype of the IL-tolerant evolved strain YlCW001

To connect genetic mutations with enhanced IL-tolerance, we re-sequenced the evolved strain YlCW001 (MT) and analyzed the mutations. Across the entire genome, we detected a total of 648 variants but only 40 genes contained variants that caused an amino acid change (Figure 3-1A). These 40 mutated genes (MGs) are randomly distributed across the 6 chromosomes of *Y. lipolytica* and held a total of 68 mutations (i.e., variants) including 35 single nucleotide variants (SNV), 12 deletions, 11 multiple nucleotide variants (MNV), 8 insertions and 2 replacements (Figure 3-1B). The most MGs were located in chromosome A (13 MGs) followed by chromosome E (9 MGs) and chromosome F (7 MGs) while the remaining chromosomes contained between 5 and 3 MGs (Figure 3-1A). By performing gene ontology (GO) associations for the 40 MGs using cluego, IMG, and blastP against *Saccharomyces cerevisiae*, we were able to assign the functional annotations for 68% of the MGs (Figure 3-1C). Kinases represent the most abundant GO terms containing 3 MGs, followed by transporters, transcription factors, mRNA processing, regulation, nuclear pore complex and membrane components (2 MGs each).

Basal gene expression differences between the WT and MT strains without IL exposure

We first examined the basal changes in gene expression between the WT and MT strains growing in medium without IL (Figure 3-2A). Temporal RNAseq data for both the WT and

MT strains were collected at the early and mid-exponential growth phases and subjected to the gene classification analysis. The results identified 470 upregulated genes and 88 increasing genes in the MT strain relative to the WT strain cultured without IL (Figure 3-2B). These genes with greater expression in the MT strain were annotated for mitochondrion, nucleus, chromosome, membrane protein complex, microbody, respirasome and transmembrane signal receptors (Figure 3-2C). For the WT strain, this resulted in 329 upregulated genes and 167 increasing genes relative to the MT strain cultured without IL (Figure 3-2B). These genes were involved in cell periphery, cell cortex, cell wall organization/biogenesis, endoplasmic reticulum, and the Golgi apparatus (Figure 3-2C). Interestingly, 5 of the MGs were differentially expressed between the MT and WT cultured without IL. The MT upregulated 3 of these MGs involved in ribosomal biogenesis (YALI0B08734g), mRNA binding (YALI0F12375g) and 1 gene of unknown function (YALI0C13002g) while the WT upregulated an expansin-like protein (YALI0E17941g) and a mitogen-activated protein kinase (YALI0A18590g).

IL-responsive gene expression

Wildtype response*.* Next, we aimed to understand gene expression changes for the WT strain exposed to IL (Figure 3-3A). This RNA-seq pairwise set lead to the greatest perturbation of gene expression relative to all other data sets with only ~62% of genes classified as no change. This resulted 365 upregulated- and 240 increasing genes for the WT growing in 8% (v/v) IL (Figure 3-3B). These IL-induced genes were enriched for mitochondrion (including mitochondrial transmembrane transport and respirasome), genetic processing (i.e., nucleus, chromosome, chromatin-binding/regulation, DNA-binding transcription factors, ribosome, and signal transduction), microtubule processing, and microbody (Figure 3-3C). In contrast, 365 upregulated- and 240 increasing genes were identified for the WT cultured without IL (Figure 3-3B). However, only 3 annotated processes were enriched more than 2-fold which included cell cortex, cell wall organization/biogenesis and the endoplasmic reticulum (Figure 3-3C).

Mutant response*.* Likewise, we compared gene expression changes for the MT strain exposed to IL (Figure 3-4A). The gene classification analysis identified 564

upregulated- and 385 increasing genes for the MT strain growing in 8%(v/v) IL relative to the MT strain without IL (Figure 3-4B). These IL-induced genes for the MT strain were enriched for lipid processes (i.e., membrane, cellular lipid metabolic process, membrane traffic protein, Golgi apparatus, endoplasmic reticulum), cell periphery (including cell cortex, cytoskeleton and microtubule process), intracellular vesicles (i.e., vacuole, cytoplasmic vesicle, vesicle-mediated transport), and genetic processes (i.e., chromosome, chromatin/chromatin-binding/regulator, DNA-binding transcription factor, signal transduction) (Figure 3-4C). The MT strain cultured without IL lead to 668 upregulated- and only 50 increasing genes (Figure 3-4B). These gene sets were annotated for anion and mitochondrial transmembrane transport, respirasome and the ribosome (Figure 3-4C). This pairwise set also identified 9 of the MGs among the classified genes. Of these, 5 were upregulated in IL (YALI0A18381g, YALI0B04268g, YALI0B18194g, YALI0F12793g, YALI0E07139g), 2 were increasing in IL (YALI0C16247g and YALI0F16665g), 1 was upregulated without IL (YALI0F16665g) and 1 gene was classified with changed regulation (YALI0F08107g).

Enhanced IL-responsive gene expression

Mutant enhanced IL-response. We reasoned that the best candidates for reverse engineering high solvent tolerance would be found by comparing WT and MT gene expression in IL (Figure 3-5A). Our gene classification of time series RNAseq data revealed 202 over expressed- and 319 increasing- genes for the MT strain in IL relative to the WT strain in IL (Figure 3-5B). These genes used by the MT were associated with lipid processes (i.e., membrane, membrane traffic protein, Golgi apparatus, endoplasmic reticulum), cell periphery (including cell cortex, cytoskeleton, cell wall organization/biogenesis and microtubule process), intracellular vesicles (i.e., vacuole, cytoplasmic vesicle, vesicle-mediated transport) and genetic processes (i.e., chromosome, chromatin/chromatin-binding/regulator and signal transduction) (Figure 3-5C). The WT upregulated 327 genes and increased 223 genes relative to the MT when both strains were cultured in IL for the MT (Figure 3-5B). These genes dominated by the WT were associated with mitochondrion, mitochondrial transmembrane transport,

respirasome and the ribosome (Figure 3-5C). Interestingly, 8 of the MGs were identified in the classified genes. Four of these were classified as changed regulation (YALI0A13233g, YALI0A18381g, YALI0C13002g and YALI0F23287g), 3 were increased by the MT (YALI0A09108g, YALI0C16247g and YALI0F16665g) and 1 was upregulated by the MT (YALI0E07139g).

A new metric to select genetic targets for reverse engineering. Traditionally, genetic targets are selected based on greatest fold change between strains or conditions and/or gene ontology if the system is well understood. However, our analysis identified 202 candidate genetic targets that were upregulated across 2 time points by the MT in IL relative to the WT in IL (Figure 3-5B) that varied in fold change across the two time points. Further, gene ontology associations for the many of these genes are vague and IL-tolerance is not a fully understood mechanism in *Y. lipolytica*. For these reasons, we sought another metric to help narrow the candidate list of genetic targets.

We began exploring the co-expression network of the genes differentially expressed and regulated between the MT and WT strains in IL and found interesting network topology where each group of classified genes showed tight co-expression (Figure 3-6A). In other words, genes that were upregulated or increased by the MT in IL were separated from genes that were upregulated or increased by the WT in IL and these gene clusters were connected by genes classified as changed regulation. Since co-expressed networks of genes is often used in systems biology to infer or explain a relationship, we hypothesized that genes with greater connectivity in the subnetwork of upregulated genes by the MT in IL would be the most influential genes conferring IL-tolerant phenotype of the MT strain. In other words, we hypothesized that gene co-expression connectivity could be a metric to predict the best genetic targets conferring IL-tolerance from these 202 upregulated genes (Figure 3-6B).

Selection of genetic targets. For each of the 202 genes upregulated by the MT in IL, the average fold change between the 2 transcriptomics time points (i.e., X (early) and Y (mid) scores) and the degree of each gene from the co-expression subnetwork (i.e., containing only the 202 genes upregulated by the MT strain in IL relative to the WT in IL) was calculated (Figure 3-6B). These genes were divided into 4 quadrants based

on the 75th percentile of average fold change and on the 75th percentile of degree. Quadrant 1 represents genes that were in the 75th percentile of both average fold change and degree, quadrant 2 represents genes that belong only in the 75th percentile of degree, quadrant 3 represents genes that were not in either 75th percentiles of degree or average fold change, and quadrant 4 represents genes that belong in both 75th percentiles of degree and average fold change (Figure 3-6C). We then selected genetic targets from the following target groups: the top 5 genes based on average fold change (quadrant 4); the top 5 genes based on degree (quadrant 2); the top 5 genes based on average fold change overlapping with degree (i.e., overlap by fold change, quadrant 1); the top 5 genes based on degree overlapping with average fold change (i.e., overlap by degree, quadrant 1); and 9 genes based on gene ontology associations related to membrane, transport, kinase, cell wall and myosin complex (quadrants 1, 2 and 4) (Figure 3-6C).

Single gene overexpression. Each of the 29 selected genetic targets were tested for their efficacy in conferring IL-tolerance by individually overexpressing each gene in the WT strain and phenotype characterization in 11% IL (Figure 3-7). In this extremely harsh concentration of IL, genes chosen from overlap by degree exhibited the best cell growth in IL among all the other target groups except genes chosen by GO term (Figure 3-7A). Of all the target groups, only 6 of genes caused a decrease in cell mass including 3 genes chosen from overlap by fold change (#1, #2 and #4), 2 genes chosen by GO terms (#21 and #27), and 1 gene chosen by degree (#19).

Six genes exhibited a statistical increase in cell mass in this high concentration of IL (#23, #10, #9, #25, #8 and #28). Interestingly, these genetic targets were selected from only 2 target groups of genes chosen from i) overlap by degree and ii) GO association (Figure 3-7B). The 3 genes conferring enhanced IL-tolerance chosen from overlap by degree included a membrane traffic protein (#8, YALI0B22396g), a kinase activator (#9, YALI0B15180g) and a microtubule binding motor protein (#10, YALI0F02673g). The 3 genes exhibiting increased IL-robustness selected by GO terms were annotated as an integral component of the membrane (#23, YALI0B12738g), a mitogen activated kinase (#25, YALI0D19470g), and a heavy chain myosin (#28, YALI0F13343g).

Discussion

ILs inhibit microbial catalytic activity by affecting numerous cellular processes and components which makes IL-tolerance a difficult phenotype to engineer. Currently, reverse engineering complex phenotypes requires extensive characterization to understand key mechanisms and often leads to numerous genetic targets for further experimentation. Here, we aimed to facilitate reverse engineering IL-robustness by exploiting a superior IL-tolerant MT strain and using co-expression to predict genetic targets. By comparing transcriptomes between WT and MT strains in IL, we discovered the best genetic targets were predicted using co-expression connectivity in conjunction with differential expression and found key cellular processes underlying the superior IL-tolerant phenotype of the MT strain.

The drastic differences in gene expression between WT and MT strains even when cultured without IL (Figure 3-2B) indicates that the mutations (Figure 3-1) acquired through 200 generations of ALE irreversibly changed basal regulation of gene expression in the MT strain. However, the significant differences in basal gene expression are likely governed by changes in epigenetics and mutations in non-coding DNA regions (e.g., promoters) since only 5 of the mutated genes were found differentially expressed.

Upon exposure to IL, only the WT increased expression of genes associated with mitochondrion processes (i.e., respirasome and mitochondrial transmembrane transport) (Figure 3-3C) suggesting ILs hinder mitochondrial function and oxidative phosphorylation. Interestingly, ILs interfered with ion transport across the mitochondrial membrane, negatively affecting mitochondria function and membrane potential in *S. cerevisiae* (14). In contrast, the WT repressed genes associated involved in cell wall organization/biogenesis (Figure 3C) which explains the reduced cell wall chitin content when *Y. lipolytica* is exposed to IL (13).

The MT strain, however, induced expression of genes associated with lipid metabolism (e.g., membrane, endoplasmic reticulum, Golgi apparatus, membrane traffic proteins, cytoplasmic vesicle, etc.) (Figure 3-4C) which is supported by our previous discovery that biogenesis of sterols is critical for IL-tolerance in *Y. lipolytica*. These results

postulate other lipid metabolic processes (e.g., vesicle-mediated transport) underlying the superior IL-tolerance of the MT strain as seen by that the drastic remodeling of lipid composition in IL (13). Both WT and MT strains responded to IL by increasing expression of genes associated with genetic processing (i.e., chromatin-binding/regulation, chromosome, transcription factors, and signal transduction) (Figure 3-3C, 3-4C) indicating ILs severely affect gene regulation and signal transduction.

Aiming to understand genotype(s) conferring high solvent tolerance, we reasoned that the best gene candidates would be found by comparing WT and MT expression in IL. Most of the genetic targets increased cell mass in IL indicating our strategy of exploiting genes used by the MT strain in this comparison was successful, though we did not test genes from other pairwise comparisons (e.g., WT 8% IL vs. WT 0% IL). Strikingly, co-expression connectivity (i.e., overlap by degree) had the highest statistical prediction of IL-tolerant genetic targets (Figure 3-7A). Though limited by the number of genes tested in our study, this metric is a promising approach to predict genetic targets for reverse engineering other complex phenotypes. While co-expression networks have been used to infer gene function, gene association and gene regulation, this is the first study where co-expression was used a metric to select genetic targets for reverse engineering to our knowledge.

Unfortunately, the top 6 performing genetic targets have not previously been characterized in *Y. lipolytica* but these genes have orthologs in the model yeast *S. cerevisiae* except for #23. TRS23 (#8, YALI0B22396g) is associated with regulation of traffic from endoplasmic reticulum to Golgi and autophagy. CLB5 (#9, YALI0B15180g) is a non-essential B-type cyclin involved in replication of DNA whereby deletion increases sensitivity to rapamycin, caffeine, hydroxyurea, doxorubicin, and many other chemicals. CIN8 (#10, YALI0F02673g) is a non-essential kinesin motor protein involved in mitotic spindle assembly and chromosome segregation. Cells deleted for CIN8 in *S. cerevisiae* exhibited abnormally large cell size, poor growth, increased mitophagy and increased sensitivity to heat and various chemicals. SPS1 (#25, YALI0D19470g) is required for prospore membrane closure. Finally, MYO1 (#28, YALI0F13343g) is a class II myosin heavy chain that plays a critical role in cytokinesis.

To further validate our methodology, we duplicated the 29 genetic target overexpression plasmids with a different selection marker and combined the 2 sets of 29 overexpression plasmids into a library stock at equimolar concentrations. This library was transformed into *Y. lipolytica* and 3 biological replicate transformants were enriched in 0%, 12% and 13% (v/v) IL for 2 rounds. We are currently sequencing the enriched cultures to identify combinations of genes enriched by IL and hence responsible for increased IL-tolerance. Additionally, we plan to validate the most enriched combinations of genes by overexpressing them in 3 *Y. lipolytica* isolates. We expect these results to successfully and universally reverse engineer high solvent tolerance in the *Y. lipolytica* clade and further establish utilization of co-expression connectivity as a metric for predicting genotypes underlying robust phenotypes.

Materials and Methods.

Strains

The *Y. lipolytica* strain, ATCC MYA-2613, a thiamine, leucine, and uracil auxotroph, was purchased from American Type Culture Collection and used as the parent strain (WT). The evolved strain YlCW001 (MT) was isolated after 200 generations in gradually increased concentrations of [EMIM][OAc] up to 18% (v/v) (13). The TOP10 Escherichia coli strain was used for molecular cloning. IL-gene overexpressing genes were generated by transforming constructed plasmids into the WT *Y. lipolytica* strain via electroporation (15).

Plasmids

The plasmid pSR001 (12) (harboring leucine selection marker, TEF promoter and CYC1 terminator) was used as the backbone plasmid for individual IL-gene overexpression cassettes. pSR001 was linearized by PCR amplification and each genetic target was amplified from *Y. lipolytica* genomic DNA. Plasmids were constructed with Gibson assembly (16) using the linearized pSR001 backbone and each individual amplified IL-gene. The complete list of primers, plasmids and strains are presented in Table 3-1.

Medium and culturing conditions

Growth medium. All cultures were conducted in defined media containing 6.7 g/L of yeast nitrogen base without amino acids (cat# Y0626, Sigma-Aldrich, MO, USA), 20 g/L of glucose, 380 mg/L leucine (cat# 172130250, Acros Organics, CA, USA), 76 mg/L uracil (cat# 157301000, Acros Organics, CA, USA), and various concentrations of 1-ethyl-3-methylimidazolium acetate [EMIM][OAc] (>95 % purity) (IoLiTec, AL, USA).

Analytical methods

Variant detection (mutation analysis). Resequencing quality filtered reads of WT (GCA_009372015.1) and MT (GCA_009194645.1) strains from whole genome resequencing (17) were subjected to variant detection within the CLC genomic workbench software version 11.0.1 (https://www.qiagenbioinformatics.com/) using the reference genome Clib122 (GCA_000002525.1) (18). Types of nucleotide changes are classified in different categories including single nucleotide variances (SNV), multiple nucleotide variants (MNV), insertions, deletions, and replacements.

RNA-sequencing. Samples were collected in biological triplicates for RNA-sequencing from wildtype (WT) and evolved mutant YlCW001 (MT) strains at early- and mid-exponential growth phases in 0% and 8% (v/v) [EMIM][OAc]. Samples were immediately quenched in liquid nitrogen and stored at -80°C until RNA isolation. Total RNA was purified using the Qiagen RNeasy mini kit (cat #74104, Qiagen, CA, USA) prior to submission at the Joint Genome Institute (JGI) using Illumina sequencing. Filtered RNA-seq reads were analyzed within the CLC genomics workbench version 11.0.1 (https://www.qiagenbioinformatics.com/) using RNA-seq analysis tool to produce transcripts per million (TPM) values for each gene/sample which was used for all further transcript-analyses.

Gene expression classification. Gene TPM values were floored to a value of 5 and averaged into $\log_2$-scaled values to compare biological conditions (BC) in the following pairwise sets: i) MT 0% IL versus WT 0% IL, ii) WT 8% IL versus WT 0% IL, iii) MT 8% IL versus MT 0% IL, and iv) MT 8% IL versus WT 8% IL (Figure 3-8 Step 1). For each pairwise set, X (i.e., early exponential fold change), Y (i.e., mid exponential fold

change) and Z (i.e., genes that are regulated differently during exponential growth) scores were calculated for each gene (Figure 3-8 Step 2). For each equation, the numerator calculates fold changes while the denominator accounts for error by dividing the fold change by the square root of pseudo variance (0.25) and the summation of variance (V) divided by respective number of replicates for each sample in the numerator.

Next, genes were categorized based on the values of X, Y and Z into 1 of 6 classes: BC1 upregulated, BC1 increasing, BC2 upregulated, BC2 increasing, changed regulation, and no change (Figure 3-8 Step 3). The upregulated class represents genes with greater expression values at both early- and mid-exponential samples (i.e., X and Y values). The increasing class represents genes without a significant change in expression value at the early-exponential sample (i.e., X) but a greater expression value at mid-exponential sample (i.e., Y). The changed regulation class represents genes that were neither upregulated nor increasing but with a regulation score greater than 1.5 (i.e., Z). The no change class represents genes that failed to meet the requirements of upregulated, increasing or changed regulation classifications.

Gene ontology enrichment. Gene ontology (GO) for genes containing variants identified from resequencing (i.e., mutated genes) were identified using ClueGo (19), integrated microbial genomes (IMG) (20) and blastP against model yeast *S. cerevisiae* with E-value < 1 E^{-10}. Annotations were combined and curated manually to classify mutated gene ontology. GO terms associated with classified genes were manually retrieved from Panther database (21) and mapped to classified genes to calculate the number of genes associated with each annotation term. GO terms were defined as enriched if the sum of upregulated and increasing genes in an annotation term were equal to or greater than double the sum of upregulated and increasing genes between the two biological conditions in each pairwise data set.

Correlation network construction and genetic targets selection. Pearson correlation was conducted on early- and mid-exponential WT 8% IL and MT 8% IL sample-gene TPM values using a stringent cutoff of 0.95. Next, this correlation network was reduced by removing all genes except those classified as upregulated by the MT 8% IL in pairwise set iv (i.e., MT 8%IL vs WT 8%IL) and used to calculate degree. These

genes were ranked according to their degree and their average fold change between the two transcriptomic time points (i.e., average of X and Y scores) and the 75th percentiles were calculated for both criteria. Finally, we chose 29 genetic targets (i.e., IL-genes) from the top 5 genes ranked by fold change, the top 5 genes degree, the top 5 genes overlapping in both 75th percentiles by fold change, the top 5 genes overlapping in both 75th percentiles by degree, and 9 genes in the 75th percentiles of either criterion with relevant gene ontology terms (e.g., membrane, kinase, transport, cell wall and myosin complex).

***Statistics*.** One-way analysis of variance (ANOVA) with Holm-Sidak correction was used for all statistical analyses with SigmaPlot v.14 software.

References

1. P.-Y. Kim, D. J. Pollard, J. M. Woodley, Substrate Supply for Effective Biocatalysis. *Biotechnology Progress* **23**, 74-82 (2007).
2. A. M. Klibanov, Improving enzymes by using them in organic solvents. *Nature* **409**, 241-246 (2001).
3. L. J. Bruce, A. J. Daugulis, Solvent Selection Strategies for Extractive Biocatalysis. *Biotechnology Progress* **7**, 116-124 (1991).
4. T. Itoh, "Chapter 1 - Biotransformation in ionic liquid" in Future Directions in Biocatalysis, T. Matsuda, Ed. (Elsevier Science B.V., Amsterdam, 2007), https://doi.org/10.1016/B978-044453059-2/50001-7, pp. 3-20.
5. M. Moniruzzaman, K. Nakashima, N. Kamiya, M. Goto, Recent advances of enzymatic reactions in ionic liquids. *Biochemical Engineering Journal* **48**, 295-314 (2010).
6. C. Li *et al.*, Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. *Bioresource Technology* **101**, 4900-4906 (2010).
7. C. G. Yoo, Y. Pu, A. J. Ragauskas, Ionic liquids: Promising green solvents for lignocellulosic biomass utilization. *Current Opinion in Green and Sustainable Chemistry* **5**, 5-11 (2017).
8. A. M. Socha *et al.*, Efficient biomass pretreatment using ionic liquids derived from lignin and hemicellulose. *Proceedings of the National Academy of Sciences* **111**, E3587 (2014).
9. E. T. Mohamed *et al.*, Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. *Microbial Cell Factories* **16**, 204 (2017).
10. M. Frederix *et al.*, Development of an E. coli strain for one-pot biofuel production from ionic liquid pretreated cellulose and switchgrass. *Green Chemistry* **18**, 4189-4197 (2016).
11. I. R. Sitepu *et al.*, Yeast tolerance to the ionic liquid 1-ethyl-3-methylimidazolium acetate. *FEMS Yeast Research* **14**, 1286-1294 (2014).

12. S. Ryu, N. Labbé, C. T. Trinh, Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α-ketoglutaric acid by Yarrowia lipolytica. *Applied Microbiology and Biotechnology* **99**, 4237-4244 (2015).
13. C. Walker, S. Ryu, C. T. Trinh, Exceptional solvent tolerance in Yarrowia lipolytica is enhanced by sterols. *Metabolic Engineering* **54**, 83-95 (2019).
14. Q. Dickinson *et al.*, Mechanism of imidazolium ionic liquidstoxicity in Saccharomyces cerevisiae and rationalengineering of a tolerant, xylose-fermenting strain. *Microbial Cell Factories* **15**, 17 (2016).
15. K. A. Markham, S. Vazquez, H. S. Alper, High-efficiency transformation of Yarrowia lipolytica using electroporation. *FEMS Yeast Research* **18** (2018).
16. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343-345 (2009).
17. C. Walker *et al.*, Draft Genome Assemblies of Ionic Liquid-Resistant Yarrowia lipolytica PO1f and Its Superior Evolved Strain, YlCW001. *Microbiology Resource Announcements* **9**, e01356-01319 (2020).
18. B. Dujon *et al.*, Genome evolution in yeasts. *Nature* **430**, 35-44 (2004).
19. G. Bindea *et al.*, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093 (2009).
20. I. M. A. Chen *et al.*, The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Research* **49**, D751-D763 (2021).
21. H. Mi *et al.*, PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research* **49**, D394-D403 (2021).

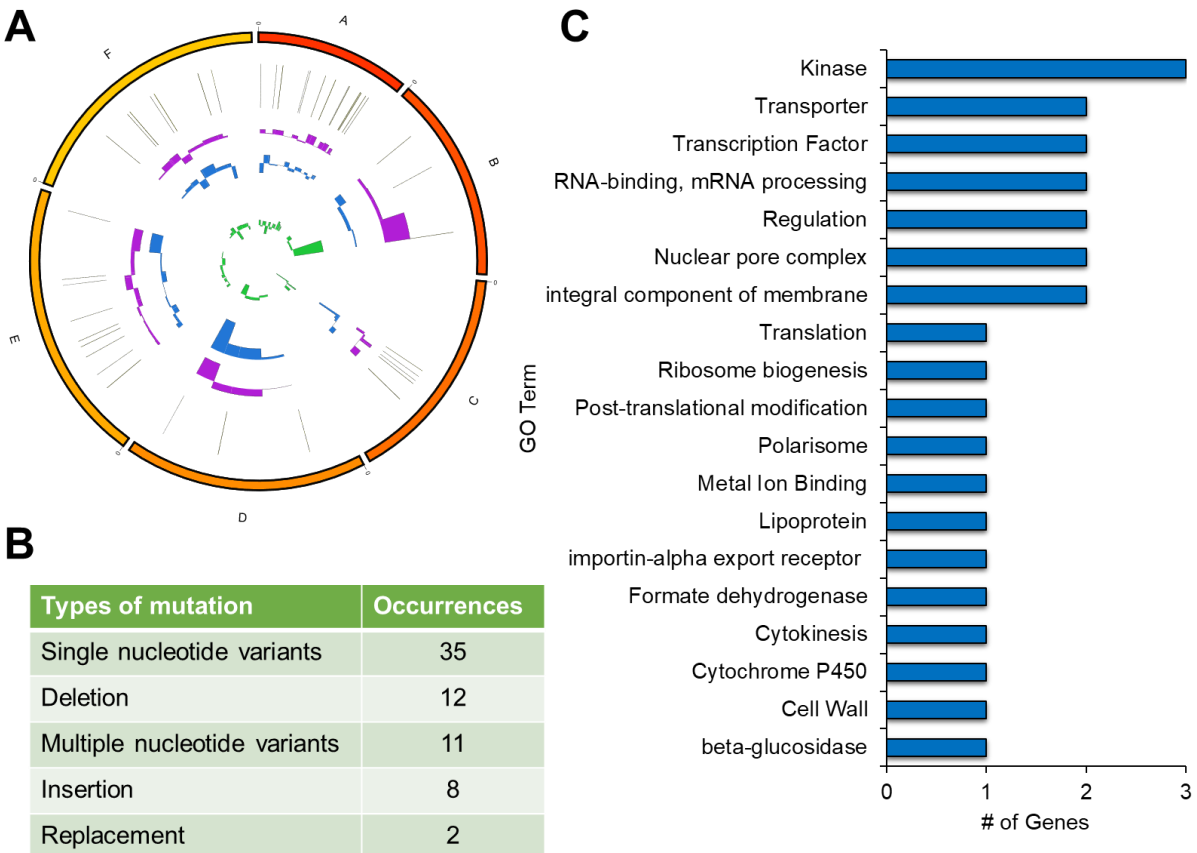
Appendix

Table 3- 1. Primers used for genetic target overexpression.

Gene ID	Locus Tag	Primer	Sequence
	N/A	2	TTTGAATGATTCTTATACTCAGAAG
pSR001	N/A	3	TCATGTAATTAGTTATGTCACGCTTAC
		79	ccttctgagtataagaatcattcaaaATGCGCGCCATTGAGGGCCTGGGC
#1	YALI0E10417g	91	gcgtgacataactaattacatgaTTAGAAACCGGTGAAGCCCGACTGCATGGAG
		80	ccttctgagtataagaatcattcaaaATGTCTCGACAAAAAAAATCGCGAGAAATCCGG
#2	YALI0C21758g	92	gcgtgacataactaattacatgaTTAAGATGCAGAAACTCTTGACCGTCGTCGG
		81	ccttctgagtataagaatcattcaaaATGTCGAACAAGCCTCGAGACCTGCCCGAGCC
#3	YALI0B16324g	93	gcgtgacataactaattacatgaTTACTGTTGGCCAGTGAGCCCCTGGTGTCTTTTGAGG
		82	ccttctgagtataagaatcattcaaaATGCCAATATTTGGGTATTACCGCTTTGGG
#4	YALI0D08118g	94	gcgtgacataactaattacatgaTTACAATCTCTCAACAGGCCCCTCGAAATCAGCCTC
		83	ccttctgagtataagaatcattcaaaATGCTATCCACGTGTGCTTCCAATCTCGAGAAGC
#5	YALI0E00792g	95	gcgtgacataactaattacatgaCTACCGGACCACTGGTTTGGCGATACGCACG
		84	ccttctgagtataagaatcattcaaaATGTCGGAATTCATTCTCAACGAGGTGTACACAG
#6	YALI0F06402g	96	gcgtgacataactaattacatgaTTAGACCTTACCCACCACCTTGAAACCCTCGGC
		85	ccttctgagtataagaatcattcaaaATGGAGTCTTCTATTTCAGTGGCCGTCAGAGTAC
#7	YALI0E07491g	97	gcgtgacataactaattacatgaCTACTGTGTCACCCTCTGGGGCATCCCCAGC
		86	ccttctgagtataagaatcattcaaaATGACCGTGTATGCACTCTACATTCTCAACAAAG
#8	YALI0B22396g	98	gcgtgacataactaattacatgaCTAGCTCACCGACTCGAGGTACTTGTTGAGATGCAGG
		87	ccttctgagtataagaatcattcaaaATGAACTTTCCAGTGAGTACTAATGACCCGGAC
#9	YALI0B15180g	99	gcgtgacataactaattacatgaTTAGATAAATAGGTTAGTTTCGTAAGAATTGACACC
		88	ccttctgagtataagaatcattcaaaATGTCTTCGCGCCACAACAGTTTGGGAATGAGG
#10	YALI0F02673g	100	gcgtgacataactaattacatgaCTACTGCTTTCTATACTTCCTCGGAGGCACCAACCTG
		89	ccttctgagtataagaatcattcaaaATGTCGACAAAATCGCCTGTCACTCCTCGATACG
#11	YALI0C03377g	101	gcgtgacataactaattacatgaCTATCTAATACCCACAGTGGAAATCTCCTTGCTCTGG
		90	ccttctgagtataagaatcattcaaaATGACACAGATTATCCATAACGCCACCATCCCC
#12	YALI0E12419g	102	gcgtgacataactaattacatgaTTACAGCTTGGCCTTGTTGAGACCCAGAACAACATCG
		103	ccttctgagtataagaatcattcaaaATGAAAATCGGCGGTCTGTGGAAAAAATTCCCCG
#13	YALI0E21549g	115	gcgtgacataactaattacatgaTTACTGAAAGTAAGTTCTCACTCTCTGTTGAATGGC
		104	ccttctgagtataagaatcattcaaaATGCTCAAGCAACTGGAAAATATTCTTGAGTCGC
#14	YALI0E17545g	116	gcgtgacataactaattacatgaCTACAACCGCCCTCTCATCTCCTTCTGAGAATGCC

Table 3-1 continued

		105	ccttctgagtataagaatcattcaaaATGGACACTCCGACCCGTCCCCGAGCCGATCTG
#15	YALI0C21296g	117	gcgtgacataactaattacatgaCTACACCTTCAGTTTGGCCTCAAGCTTCATAAAATCC
		106	ccttctgagtataagaatcattcaaaATGTCACTTCCATACGCCCCACTAACCCATATGC
#16	YALI0A18722g	118	gcgtgacataactaattacatgaCTAATCTAGTTCTGCAAAGAAACCCTTGTAAG
		107	ccttctgagtataagaatcattcaaaATGGCAAGACCAGCTCGACGACGGTCTTCCGG
#17	YALI0B03476g	119	gcgtgacataactaattacatgaTTATGCTCCCAGCTGATCCAGGCGAACATTCATATCC
		108	ccttctgagtataagaatcattcaaaATGACACAAACGCACAATCTGTTTTCGCCAATC
#18	YALI0D16247g	120	gcgtgacataactaattacatgaTTACATCTTGTACGCAGGGTAATCAGTGTAGCCC
		109	ccttctgagtataagaatcattcaaaATGAAGTTCCTCGTTTGTCTTCTGGTTCTGGCGC
#19	YALI0E34852g	121	gcgtgacataactaattacatgaTTAAACCAAAGTCTTGACCTCAAAGAATCGC
		110	ccttctgagtataagaatcattcaaaATGTTTGGCATCTTTTCGCGCATTCCCGTGGAGC
#20	YALI0F27423g	122	gcgtgacataactaattacatgaCTAACTGCTCGTAGTGGCCACCTCAACAGGCTCG
		111	ccttctgagtataagaatcattcaaaATGATCATCAGCCTGACTGTGACCGTTAAAACG
#21	YALI0D12232g	123	gcgtgacataactaattacatgaTCAAACAACCTCTTCTCGGACAAAGCCCCGGC
		112	ccttctgagtataagaatcattcaaaATGGAGCACTTCACTGCGCACGATGGGTCGTTCC
#22	YALI0C08096g	124	gcgtgacataactaattacatgaTTACATAATAACACAATCGTCGGCCTCCTTGGTCACC
		113	ccttctgagtataagaatcattcaaaATGGACACGAAACGAAGCCTTTCAGGGGCTCCCC
#23	YALI0B12738g	125	gcgtgacataactaattacatgaCTACCTCGTCAGACCAACTCCTAATCCTACCG
		114	ccttctgagtataagaatcattcaaaATGTTGCCACGAGCCCTGAGTCGCCACGTGG
#24	YALI0C18183g	126	gcgtgacataactaattacatgaCTACTTATAGTCTGCGACTTCTTCAACGTCACCAGGG
		127	ccttctgagtataagaatcattcaaaATGGCATCGGCAAGCTCAGAACTATCACGGCTGG
#25	YALI0D19470g	132	gacataactaattacatgaCTAGTAACGATCGTACTCTTCGAACTCCTCGATCCACCG
		128	ccttctgagtataagaatcattcaaaATGAAGCTCGAAATCAAGGTGAGTAATTGAGCG
#26	YALI0C21802g	133	gcgtgacataactaattacatgaTTACAAGGTGTTGGCAGCCTCATCAACCTCCTCC
		129	ccttctgagtataagaatcattcaaaATGCCTTCCGAATCCACCCCTCTCATTGGCCGAC
#27	YALI0B09273g	134	gcgtgacataactaattacatgaTTACATATCCGAGGGATAGTAGAAGAAGGCAATGGCG
		130	ccttctgagtataagaatcattcaaaATGTGCGTGATGATGTTTTGTTTCAGCTGGATTG
#28	YALI0F13343g	135	gcgtgacataactaattacatgaTTAAACAAAGGAGGGCTCAAAGTTGGAGTTGTTGTTG
		131	ccttctgagtataagaatcattcaaaATGTCGGAACATTCGTTTGCCAACAACTTTTGGG
#29	YALI0E29557g	136	gcgtgacataactaattacatgaCTAAATGTCAGCCAAAAAGTTGCTGGGAGCTAGACC



Types of mutation	Occurrences
Single nucleotide variants	35
Deletion	12
Multiple nucleotide variants	11
Insertion	8
Replacement	2

Figure 3- 1. Mutated genes of IL-tolerant mutant (MT, YlCW001) *Y. lipolytic*a which include 68 variants across 40 genes that change amino acid residues. (**A**) Chromosomal location of mutated genes depicting log2 fold changes at the mid-exponential RNA-seq sample between the MT in 8% [EMIM][OAc] verses WT in 0% IL (purple), WT in 8% IL (blue) and MT in 0% IL (green). (**B**) Types of variants in the mutated genes. (**C**) Gene ontology of the mutated genes.

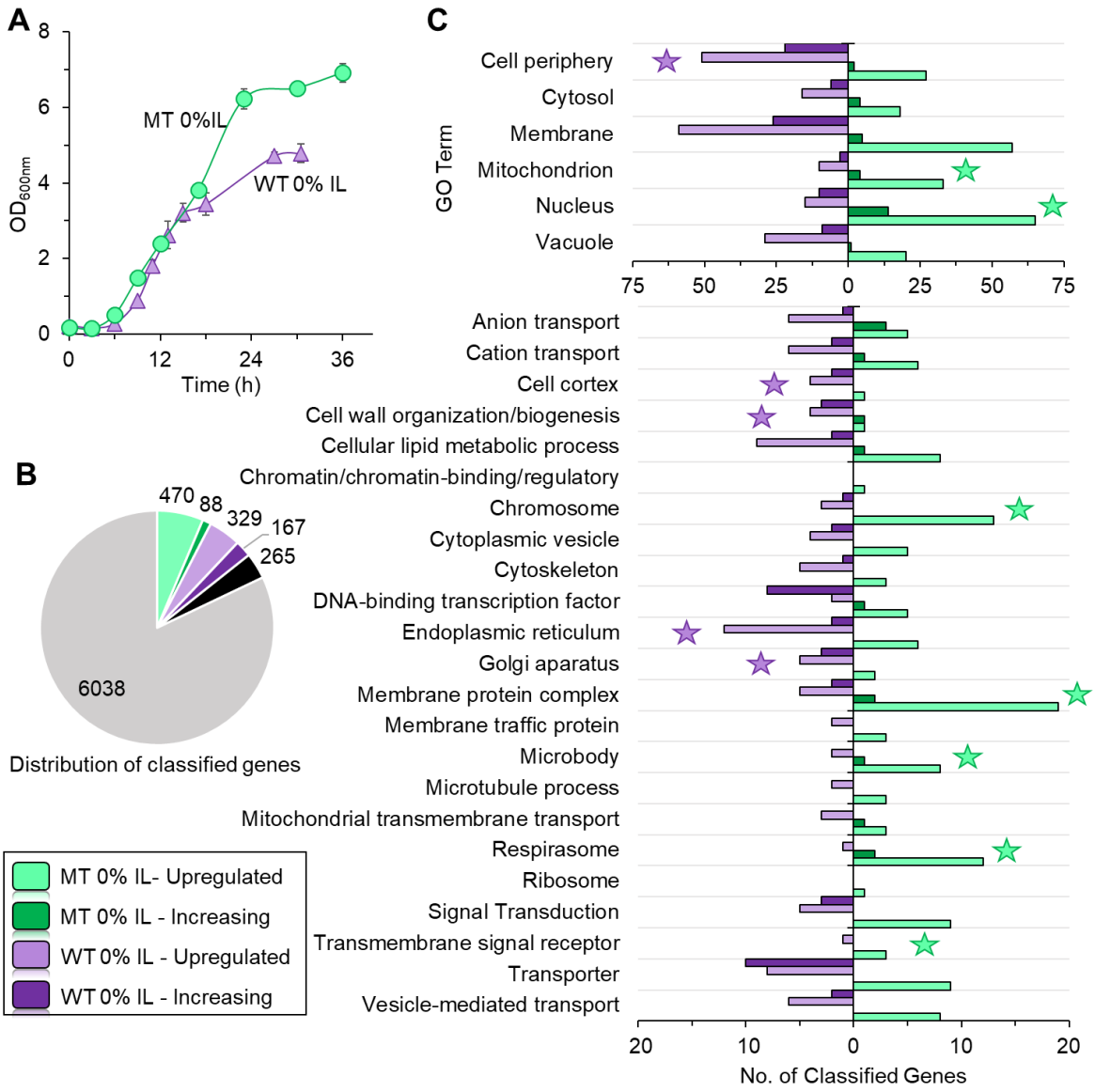


Figure 3- 2. Basal differences in gene expression between the MT and WT in 0% IL. (**A**) Growth of MT and WT in 0% IL. (**B**) Distribution of classified genes. (**C**) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition.

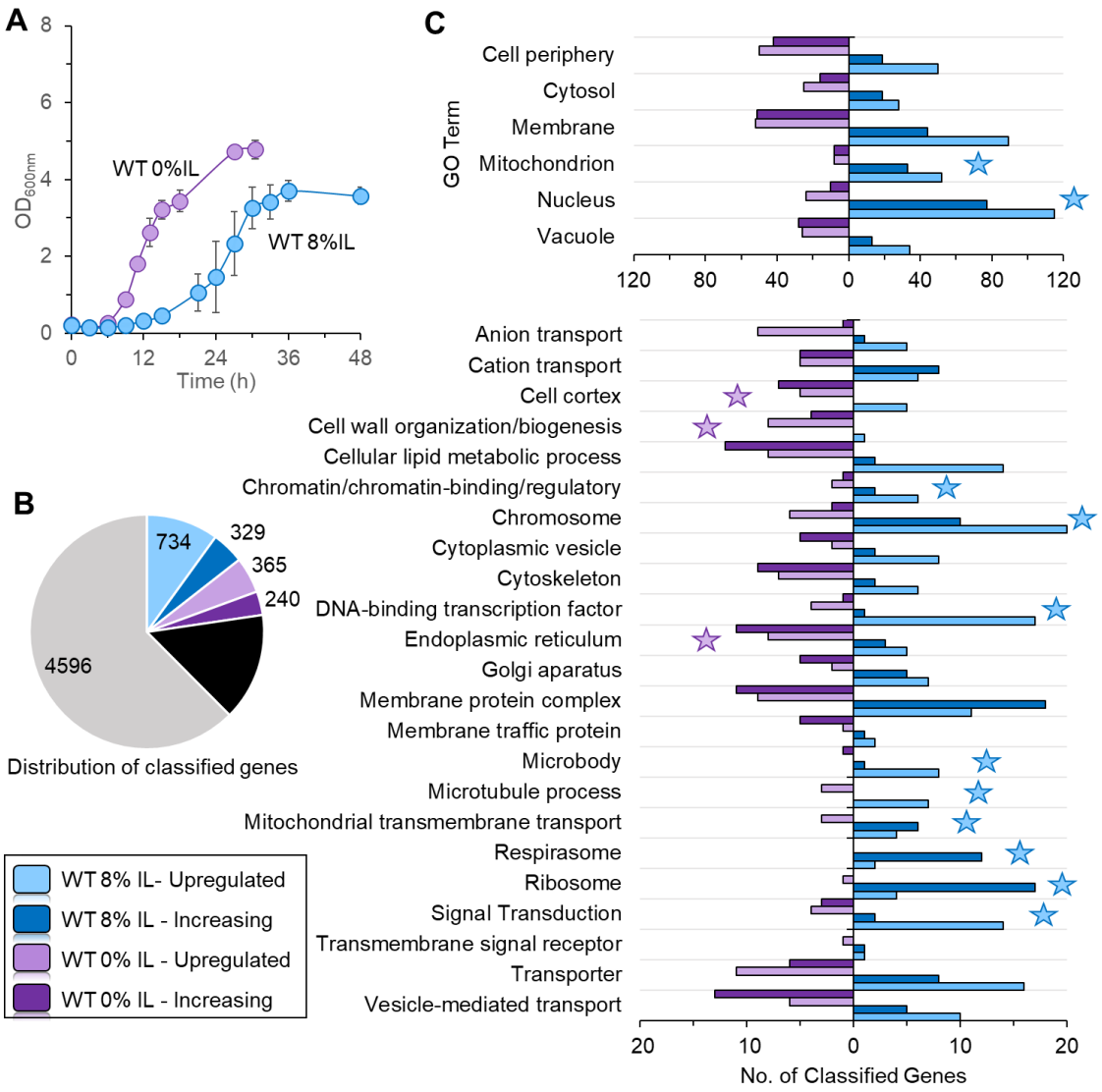


Figure 3- 3. IL-responsive gene expression of the WT strain. (**A**) Growth of WT in 8% and 0% IL. (**B**) Distribution of classified genes. (**C**) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition.

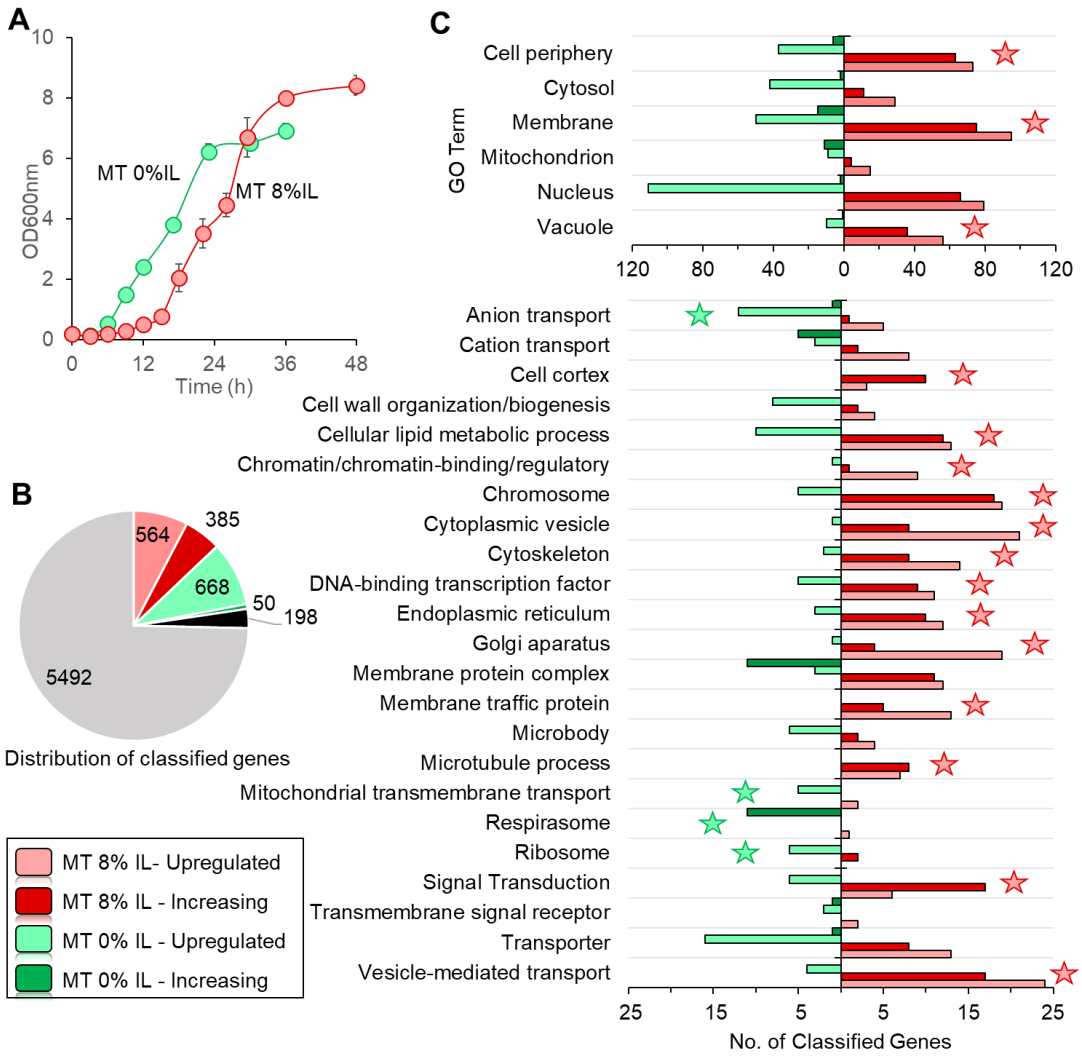


Figure 3- 4. IL-responsive gene expression of the MT strain. (**A**) Growth of MT in 8% and 0% IL. (**B**) Distribution of classified genes. (**C**) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition.

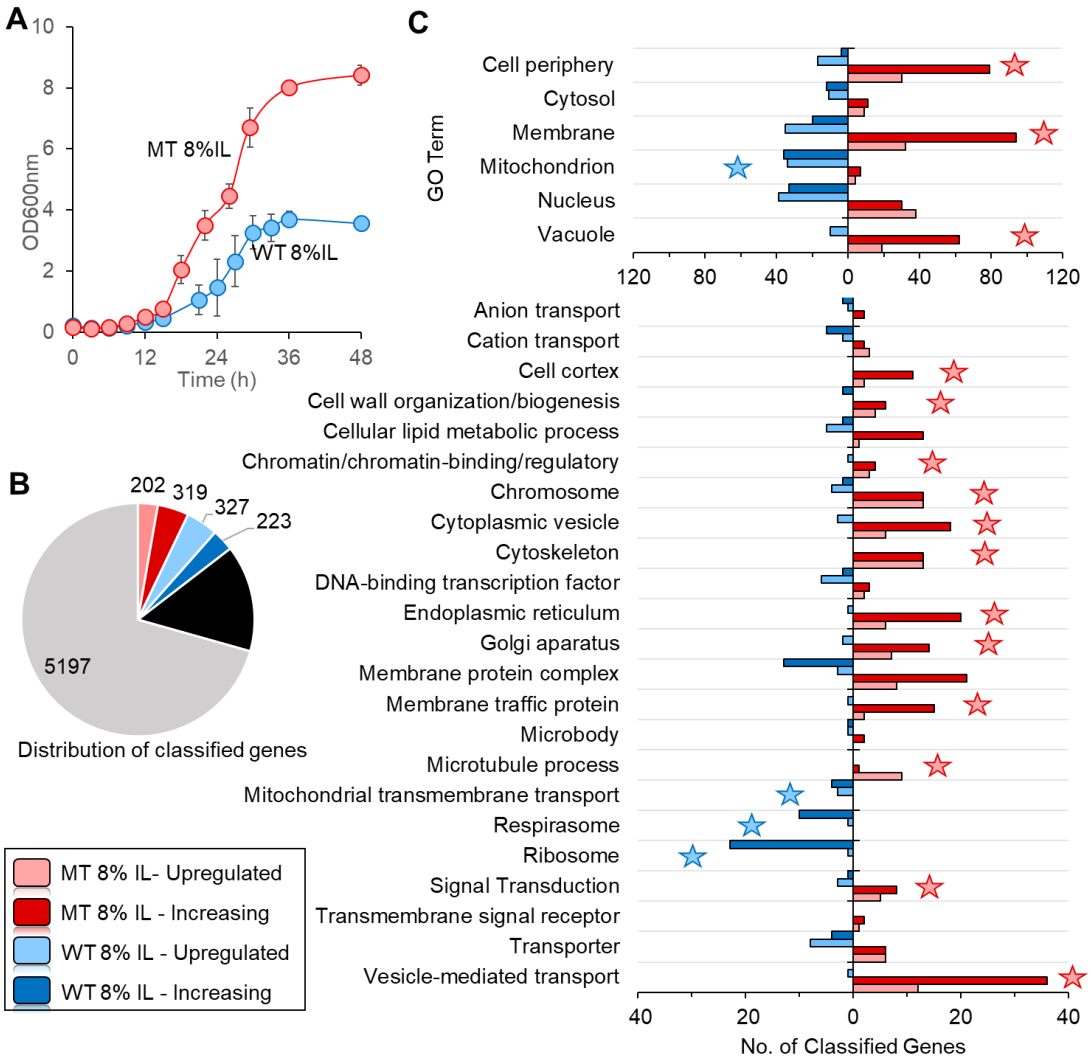
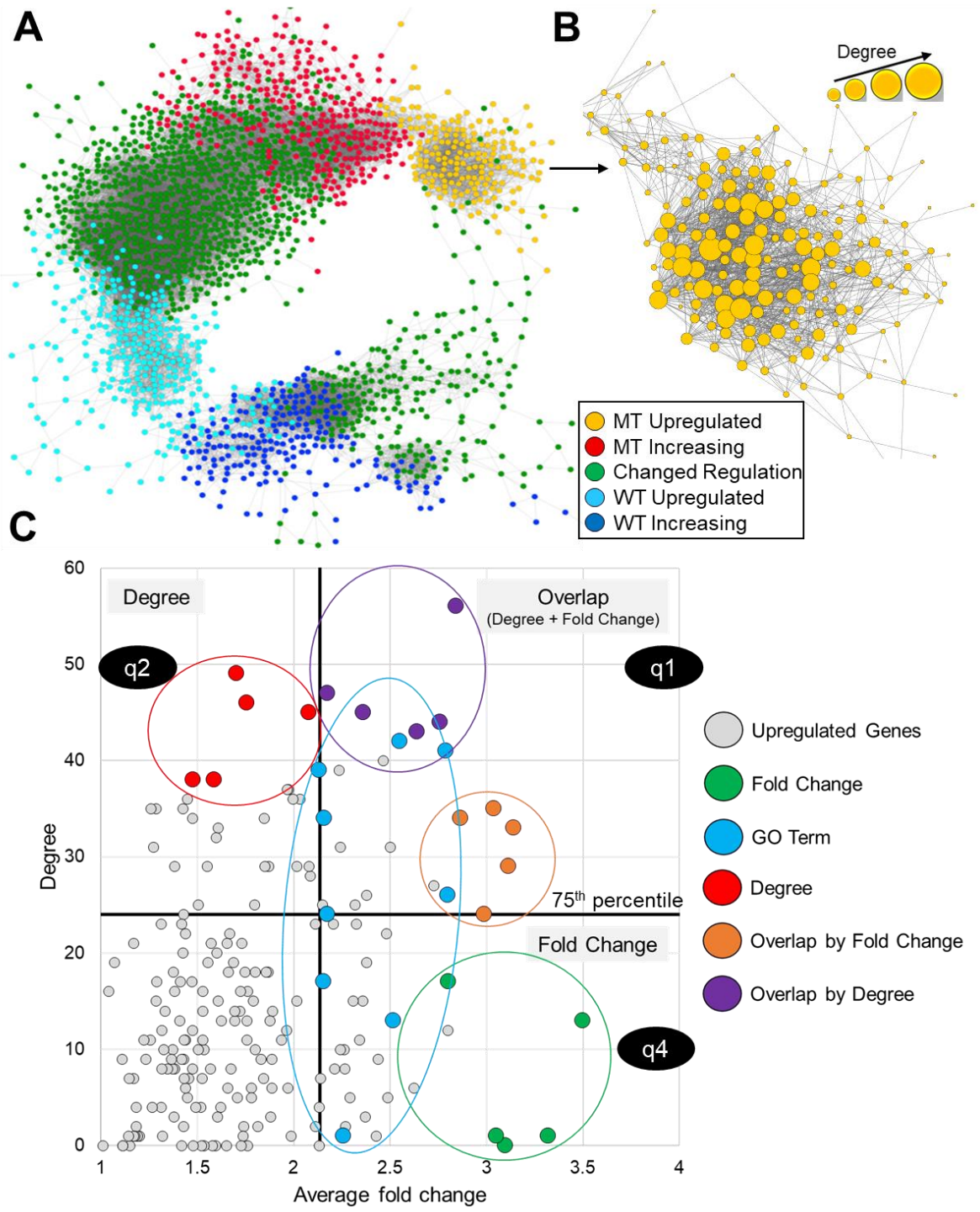


Figure 3- 5. Enhanced IL-responsive gene expression of the MT strain verses the WT strain in 8% IL. (**A**) Growth of MT and WT in 8% IL. (**B**) Distribution of classified genes. (**C**) Number of classified genes in each category of gene ontology. Stars represent ontology terms where the sum of upregulated and increasing genes in a biological condition is at least twice greater than the other biological condition.

Figure 3- 6. (**A**) Correlation (Co-expression) network of classified genes between the MT and WT in 8% IL. (**B**) Subnetwork of co-expressed genes upregulated by the MT in 8% IL illustrating gene connectivity by degree. (**C**) Genetic targets selected by degree, fold change or gene ontology.


A
B
Degree
MT Upregulated
MT Increasing
Changed Regulation
WT Upregulated
WT Increasing
C
Degree
Overlap
(Degree + Fold Change)
q2
q1
q4
75th percentile
Fold Change
Degree
Average fold change
Upregulated Genes
Fold Change
GO Term
Degree
Overlap by Fold Change
Overlap by Degree

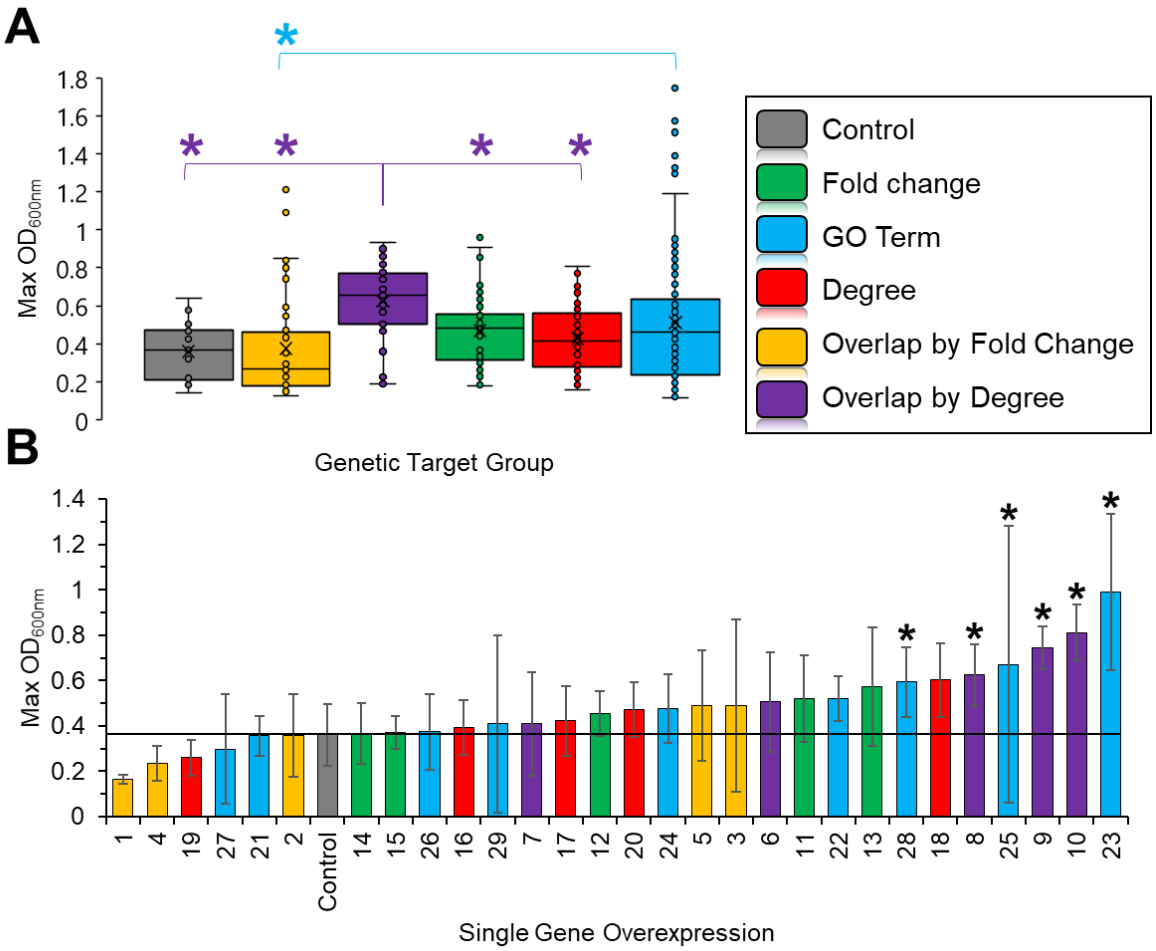


Figure 3- 7. Individual overexpression of genetic targets in 11% IL. (**A**) Box and whisker plot for each group of genetic targets depicting all replicate maximum OD 600nm for all associated genes contained within each group. (**B**) Max OD600nm for genetic targets cultured in 11% IL. *p-value ≤ 0.05 (n ≥ 9 for each gene and n=27 for the empty vector control).

1 Average Log_2(TPM) values

Gene	$BC1_{early}$	$BC1_{mid}$	$BC2_{early}$	$BC2_{mid}$
A	#	#	#	#
B	#	#	#	#
C	#	#	#	#

Pairwise Set	Biological condition 1 (BC1)	Biological condition 2 (BC2)
i	MT 0% IL	WT 0% IL
ii	WT 8% IL	WT 0% IL
iii	MT 8% IL	MT 0% IL
iv	MT 8% IL	WT 8% IL

2 Calculate X, Y and Z scores

$$X = \frac{BC1_{early} - BC2_{early}}{\sqrt{0.25 + (V/n)_{BC1_{early}} + (V/n)_{BC2_{early}}}}$$

$$Y = \frac{BC1_{mid} - BC2_{mid}}{\sqrt{0.25 + (V/n)_{BC1_{mid}} + (V/n)_{BC2_{mid}}}}$$

$$Z = \frac{(BC1_{mid} - BC1_{early}) - (BC2_{mid} - BC2_{early})}{\sqrt{0.25 + (V/n)_{BC1_{mid}} + (V/n)_{BC1_{early}} + (V/n)_{BC2_{mid}} + (V/n)_{BC2_{early}}}}$$

Variable	Definition
X	Fold change at earl- exponential sample
Y	Fold change at mid-exponential sample
Z	Regulation score
BC1	Average Log_2(TPM) of a gene for BC1
BC2	Average Log_2(TPM) of a gene for BC2
Early	Early-exponential sample
Mid	Mid-exponential sample
0.25	Pseudo variance
V	Variance
N	Number of replicates

3 Classify Genes

Classification	*X*	*Y*	*Z*
BC1 upregulated	X ≥ 1	Y ≥ 1	-
BC1 increasing	-1 < X < 1	Y ≥ 1	Z ≥ 1.5
BC2 upregulated	X ≤ -1	Y ≤ -1	-
BC2 increasing	-1 < X < 1	Y ≤ -1	Z ≤ -1.5
Changed regulation	Is not upregulated or increasing		| Z | ≥ 1.5
No change	Is not otherwise classified		

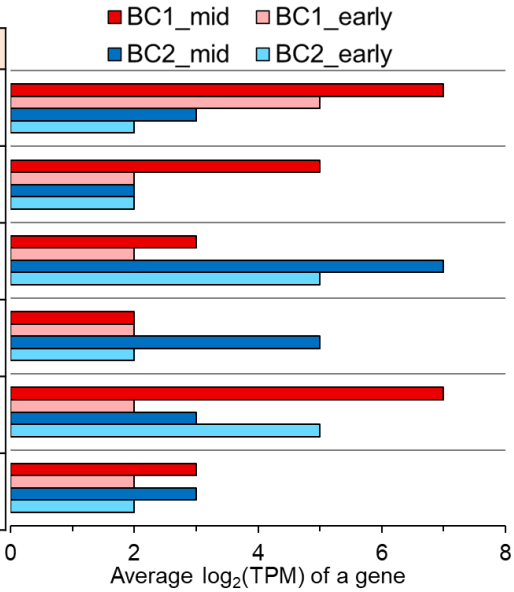


Figure 3- 8. Gene classification methodology. Step 1: For each pairwise set of biological conditions, gene TPM values are floored to a minimum of 5, averaged, and converted into log_2 scale. Step 2: Fold change (X and Y) and regulation (Z) scores are calculated between biological conditions. Step 3: Genes are classified based on the scores of X, Y and Z into upregulated or increasing for a biological condition, changed regulation, or no change.

Chapter IV

Exploring Proteomes of Robust *Yarrowia lipolytica* Isolates Cultivated in Biomass Hydrolysate Reveal Key Processes Impacting Complex Sugar Utilization, Lipid Accumulation, and Degradation

Authors for this manuscript include:

Caleb Walker, Bruce Dien, Richard J. Giannone, Patricia Slininger, Stephanie R. Thompson, and Cong T Trinh.

CTT conceived and supervised the study. CW, BR, RG and CTT designed the experiments, and analyzed the data. BR and ST performed the experiments. CW and CTT wrote and approved the manuscript.

ABSTRACT

Yarrowia lipolytica is an oleaginous yeast exhibiting robust phenotypes beneficial for industrial biotechnology. The phenotypic diversity found within the undomesticated *Y. lipolytica* clade from various origins illuminates desirable traits not found in the well-characterized laboratory strain CBS7504, which include superior xylose utilization, lipid accumulation, and growth on undetoxified biomass hydrolysates. Currently, the related phenotypes of lipid accumulation and degradation when metabolizing non-preferred sugars associated with biomass hydrolysates (i.e., xylose) are poorly understood, making it difficult to control and engineer in *Y. lipolytica*. To fill this knowledge gap, we analyzed the genetic diversity of five undomesticated *Y. lipolytica* strains and identified singleton genes and genes exclusively shared by strains exhibiting desirable phenotypes including superior growth, xylose utilization, lipid accumulation and/or degradation properties in undetoxified biomass hydrolysates. Strain characterizations from controlled bioreactor cultures revealed that the unconventional strain YB420 used xylose to support cell growth and maintained high lipid levels while the laboratory strain CBS7504 degraded cell biomass and lipids when xylose was the sole remaining carbon source. From proteomic analysis, we identified carbohydrate transporters, xylose metabolic enzymes and pentose phosphate pathway proteins stimulated during the xylose uptake stage for both strains. Furthermore, we distinguished proteins in lipid metabolism (e.g., lipase, NADPH generation, lipid regulators, β-oxidation) activated by YB420 (lipid maintenance phenotype) or CBS7504 (lipid degradation phenotype) when xylose was the sole remaining carbon source. Overall, the results relate genetic diversity of undomesticated *Y. lipolytica* strains to complex phenotypes of superior growth, sugar utilization, lipid accumulation and degradation in biomass hydrolysates.

IMPORTANCE

Yarrowia lipolytica is an important industrial oleaginous yeast due to its robust phenotypes for effective conversion of inhibitory lignocellulosic biomass hydrolysates into neutral lipids. While lipid accumulation has been well characterized in this organism, its interconnected lipid degradation phenotype is poorly understood during fermentation of biomass hydrolysates. Our investigation into the genetic diversity of undomesticated *Y. lipolytica* strains, coupled with detailed strain characterization and proteomic analysis, revealed metabolic processes and regulatory elements that are correlated with superior growth, sugar utilization, and lipid accumulation or degradation exhibited by these natural variants. This study provides a better understanding of the robust metabolism of *Y. lipolytica* and suggests potential metabolic engineering strategies to enhance its performance.

Key words: bioreactor characterization, proteomic analysis, xylose metabolism, xylose transporters, lipid accumulation, lipid degradation, lipid regulators.

INTRODUCTION

Yarrowia lipolytica is an important oleaginous yeast for industrial biotechnology. Wildtype strains can accumulate a remarkable 40% of cell weight in neutral lipids from lignocellulosic biomass or agricultural wastes (1). These microbial lipids are a promising alternative to petroleum and animal oils for the sustainable production of advanced fuels and oleochemicals. In addition, *Y. lipolytica* is exceptionally robust to chemical inhibitors and stressful environments, which are critical biocatalyst properties to achieve sustainable production of chemicals from low-cost biomass feedstocks. It can tolerate broad pH ranges (2), high salt concentrations (3), and organic solvents (e.g., ionic liquids) (4, 5); in fact, most *Y. lipolytica* isolates exhibit robust growth in up to 60% (v/v) undetoxified dilute acid-pretreated switchgrass hydrolysates that are normally inhibitory to microbes (6). Thus, a better understanding of the mechanisms that underpin *Y. lipolytica*' s natural robustness would not only enable development of niche strains for novel biocatalysis regimes but would also provide fundamental knowledge that may be applied to other industrially-relevant organisms.

Recently, significant research has focused on manipulating the metabolism of conventional laboratory strain CBS7504 (W29) for enhanced lipid production and utilization of pentose (e.g., xylose) and hexose (e.g., glucose) sugars in inhibitory lignocellulosic biomass hydrolysates. To increase lipid production in *Y. lipolytica*, numerous metabolic engineering strategies have been implemented that successfully redirected carbon flux to lipid metabolism (7) including overexpression of lipid biosynthesis enzymes (8) and/or disruption of the competitive β-oxidation pathway (9) and altered expression of regulators of lipid accumulation (10). The most lipogenic *Y. lipolytica* strain reported to date achieved 90% lipid content by simultaneous restoration of leucine and uracil biosynthesis, overexpression of diacylglycerol transferase (DGA1), deletion of peroxisome biogenesis enzyme peroxin-10 (PEX10), deletion of multifunctional β-oxidation enzyme (MFE1), and optimization of culture conditions (11).

To maximize lipogenesis from biomass hydrolysates requires efficient utilization of both hexose and pentose sugars. *Y. lipolytica* does not efficiently use xylose as a sole

carbon source, albeit it processes genes for the complete xylose catabolic pathway (12). Activation of this cryptic pathway has been accomplished through adaptive evolutionary approaches resulting in improved xylose utilization (13). Furthermore, overexpression of endogenous xylose catabolic genes (14-16) and heterologous expression of xylose reductase and xylose dehydrogenase from *Scheffersomyces stipitis* (16, 17) have successfully increased xylose consumption rates. Several transporters have also been identified in *Y. lipolytica* showing increased expression levels during xylose assimilation and combinatorial overexpression of the endogenous xylose dehydrogenase with several of these transporters has also achieved improved growth on xylose (18). While the production phenotypes are well characterized, fundamental understanding of complex phenotypes responsible for superior growth, sugar utilization, and lipid accumulation--or degradation--during fermentation of biomass hydrolysates are still lacking.

Complementary to these engineering efforts, recent investigation into the genetic diversity of undomesticated *Y. lipolytica* strains revealed emergent robust phenotypes not present in the conventional strain W29. Characterization of fifty-seven undomesticated *Y. lipolytica* isolates on inhibitory undetoxified biomass hydrolysates revealed select strains with superior growth, lipid production, and pentose-sugar assimilation relative to the conventional laboratory strain CBS7504 (6). In this study, *Y. lipolytica*' s natural genetic diversity is further explored using a combination of detailed strain characterization and proteomic analysis. Coupled together, these analyses uncover the underlying mechanisms behind these poorly understood complex phenotypes during fermentation of biomass hydrolysates. The results presented here will aid engineering efforts to better control lipid accumulation or degradation phenotypes to optimally produce advanced biofuels and/or oleochemicals from biomass hydrolysates.

RESULTS

Comparative genomics reveals unique genotypes of undomesticated Yarrowia strains

***Phylogenetic tree of Y. lipolytica isolates shows close similarity between genomes*.** Phylogenetic species analysis distinguished the first evolutionary split dividing the *Yarrowia* clade into two ancestral roots (Figure 1A). The first root contained the non-conventional YB419 and conventional strains CLIB122 with CBS7504 (W29). The second root contained the remaining four non-conventional isolates, depicting YB392 and YB420 as the most divergent from YB567 followed by YB566. This result was surprising since YB392, YB419 and YB420 were all isolated from corn milling plants within Illinois (19). Interestingly, the closest related species to the *Yarrowia* clade is *Sugiyamaella lignohabitans*, an efficient pentose utilizing and facultative anaerobic yeast (20).

***Undomesticated Y. lipolytica isolates contain unique genes not found in laboratory strains*.** The undomesticated strains were characterized for unique genes that may contribute to their distinctive phenotypes. A singleton gene signifies a gene appearing exclusively in one of the genomes within the pangenome (i.e., laboratory strains CBS7504 (W29) and CLIB122; undomesticated strains YB392, YB419, YB420, YB566 and YB567). Of the undomesticated *Y. lipolytica* strains, YB419 contained the most singletons (23 genes) followed by YB420 (17 genes), while the remaining 3 isolates contained 5 or less singletons (Figure 4-1B, Table 4-1).

Exclusively shared genes between strains exhibiting enhanced xylose assimilation and lipid accumulation. Two of the undomesticated strains, YB566 and YB567, outperformed all other *Y. lipolytica* strains in assimilating xylose from switchgrass hydrolysates (SGH) (6). We hypothesized that these two strains contain a shared genotype that was not found in other strains and was perhaps responsible for enhanced xylose assimilation. There are 10 genes exclusively shared between these two xylose-utilizing strains (Figure 4-1B, Table 4-2). Similarly, we sought exclusive genotypes associated with *Y. lipolytica* strains with superior lipid accumulation from SGH, namely

YB392, YB419 and YB420 (6). Six genes are exclusively shared among these three lipogenic strains (Figure 4-1B, Table 4-3).

***Y. lipolytica* strains of different origins thrive in undetoxified biomass hydrolysates and exhibit distinct phenotypes.** Growth characterization of CBS7504 and YB420 *Y. lipolytica* strains were performed in computer-controlled bioreactors (Figure 4-2A). The conventional *Y. lipolytica* strain CBS7504 grew well in undetoxified biomass hydrolysates, achieving maximum cell mass (39.3 ± 3.3 OD600nm) within 28 hours of fermentation from the co-utilization of 2.9 ± 0.05 (%w/v) glucose and 0.8 ± 0.03 (%w/v) xylose (Figure 4-2B). Lipids accumulated during growth, reaching a maximum of 2.1 ± 0.6 g/L at 21 hours into the fermentation (Figure 4-2D). Upon glucose exhaustion, cell mass and accumulated lipid levels steadily declined for the remaining 48 hours of fermentation despite the continued consumption of xylose. At 72 hours of fermentation, CBS7504 consumed a total of 2.1 ± 0.02 (%w/v) xylose and produced 0.91 ± 0.008 (%w/v) xylitol (yielding 0.44 ± 0.002 xylitol/xylose). The undomesticated *Y. lipolytica* strain YB420 also grew robustly in undetoxified biomass hydrolysates, but showed contrasting phenotypes with CBS7504. Over 45 hours of fermentation, YB420 showed less co-utilization of glucose (2.8 ± 0.03 [%w/v]) and xylose (0.4 ± 0.03 [%w/v]) and less lipid production (1.6 ± 0.5 g/L) than CBS7504 (Figure 4-2C, 4-2E). However, upon glucose depletion, YB420 maintained cell mass and lipids while consuming a total of 2.1 ± 0.01 (%w/v) xylose and producing 0.61 ± 0.04 (%w/v) xylitol (yielding 0.30 ± 0.006 xylitol/xylose). The differences observed in both the rate of xylose consumption and the maintenance of lipid levels prompted a systems-level comparison of the two strains.

Proteomic analysis reveals key processes impacting sugar utilization and lipid degradation

Proteome alterations in growth stages. Proteomic samples were collected during the exponential growth phase when glucose was assimilated (S1 and S2) and stationary phase when xylose was assimilated (S3 and S4) for both strains in biomass hydrolysate cultures. Differences in protein abundances were compared for S2, S3 and S4 against S1 for each strain (Figure 4-3A, 4-3B, 4-3D and 4-3E). As expected, there

were only minor proteome differences between glucose assimilation phase samples (S1 and S2) (Figure 4-3C, 4-3F). However, *Y. lipolytica* strains dramatically altered their proteomes during stationary phase samples (samples S3 and S4) when xylose was assimilated, and lipid levels were maintained by YB420 but degraded by CBS7504 (Figure 4-3B, 4-3E). For CBS7504, 673 protein abundances changed throughout the stationary phase (samples S3 and S4) while 800 protein abundances were changed in YB420. Since the largest proteome differences were found between exponential (S1 and S2) and stationary (S3 and S4) phase samples, we chose to characterize xylose assimilation and lipid degradation phenotypes using proteins with significant changes in abundance at S3 and/or S4 relative to S1.

Proteome alterations in xylose assimilation. CBS7504 consumed xylose faster than YB420 but converted more of it into xylitol (0.44 ± 0.002 %w/w of xylose) rather than maintaining cell mass or lipid content (Figure 4-2B, 4-2C). This led us to investigate protein abundances in the pentose phosphate pathway (PPP) where xylose is introduced into cellular central metabolism. Despite the quick assimilation of xylose, during xylose assimilation (S3 and S4) CBS7504 only upregulated the protein abundance of transketolase (TKL, YALI0D02277g) in the PPP (Figure 4-4A, Table 4-4). Interestingly, CBS7504 downregulated the protein abundance of ribose-phosphate pyrophosphokinase (PRS1, YALI0B13552g) which converts ribose-5-phosphate into 5-phosphoribosyl 1-pyrophosphate (PRPP) to feed downstream biosynthetic pathways (i.e., histidine, pyrimidine and purine metabolism) associated with cell growth, correlating with the decreased cell mass, increased xylitol production and lipid degradation phenotypes of CBS7504. Meanwhile, during xylose assimilation (S3 and S4) YB420 accumulated less xylitol (0.30 ± 0.006 %w/x of xylose) and maintained cell mass and lipid content (Figure 4-2C). Regarding the PPP, six proteins were upregulated, and none downregulated in YB420 during the stationary phase (Figure 4-4B, Table 4-4). Not surprisingly, these upregulated proteins include xylitol dehydrogenase (Xyl2, YALI0E12463g) and xylulokinase (Xyl3, YALI0F10923g) which together convert xylitol into xylulose-5P. Notably, YB420 also upregulated 2 proteins annotated (in panther database) for xylulose kinase (YALI0D15114g) and ribulokinase (YALI0E13321g) activities. Additionally, YB420

increased the protein abundance of D-arabinitol 2-dehydrogenase (ADH, YALI0F02211g).

Transporters. In total, 87 transporters were identified with statistically significant abundance changes when comparing stationary phase (S3 and/or S4) to exponential growth phase (S1) (Figure 4-4C, Table 4-4). While most are annotated for ion and inorganic molecular entity transmembrane transport activities, we focused on the 28 active transmembrane transporters to identify those with altered protein abundances during xylose assimilation. Specifically, six of these transporters are annotated with carbohydrate transmembrane transporter activity and have been studied for xylose assimilation in *Y. lipolytica* (Figure 4-4D) (13). YALI0C06424g is a carbohydrate symporter with the largest increase in abundance for both strains. This protein is similar to Snf3p and Rgt2p proteins of *Saccharomyces cerevisiae* that are involved in glucose sensing and signaling as well as fructose and mannose transport (21). Both CBS7504 and YB420 strains also increased the abundance of YALI0F06776g, an observation that is in agreement with previous transcriptomics data measuring Y. lipolytica's growth response to xylose as the sole carbon source (13). Though these two transporters exhibited similar upregulation patterns during growth on xylose across both strains, albeit with varied magnitudes, other transporters were more strain specific: YALI0D00363g, YALI0F25553g, and YALI0C04730g. YALI0D00363g was strongly upregulated in S4 for CBS7504, but not in YB420, while YB420 increased the protein abundance of YALI0F25553g at both S3 and S4 and YALI0C04730g S3 but not in CBS7504. Interestingly, individual overexpression of these 3 carbohydrate symporters (YALI0D00363g, YALI0F25553g or YALI0C04730g) supported growth on plates containing xylose as the sole carbon source (18). Lastly, YALI0D01111g was downregulated in CBS7504 during xylose assimilation S3 and S4 and at S4 in YB420.

Proteome alterations in lipid metabolism. CBS7504 demonstrated lipid degradation while YB420 maintained lipid content during the stationary/xylose assimilation growth stage samples (S3 and/or S4) (Figures 4-2D, 4-2E, and 4-5A). To understand the cause of lipid degradation/maintenance, we compared proteins involved in fatty acid degradation and triacylglycerol (TAG) metabolism. In the TAG synthesis

pathway, CBS7504 increased the abundance of both bifunctional glycerol-3-phosphate/glycerone-phosphate O-acyltransferase (SCT1, YALI0C00209g), converting glycerol-3-phosphate (gly-3P) into lysophosphatidic acid (LPA), and acyl-CoA dependent diacylglycerol acyltransferase I (DGA1, YALI0E32769g), which converts diacylglycerol (DAG) into TAG (Figure 4-5B, Table 4-5). Likewise, YB420 increased the of abundance DGA1, but also lysophosphatidate acyltransferase (ALE1, YALI0F19514g) and phosphatidic acid phosphohydrolase (PAP, YALI0D27016g), which together convert LPA into TAG (Figure 4-5C, Table 4-5). However, YB420 downregulated the abundance of diacylglycerol diphosphate phosphatase/phosphatidate phosphatase (LPP1, YALI0B14531g) which has the same metabolic function as PAP (KEGG E.C.3.1.3.4).

Considering the β-oxidation pathway, both strains showed increased protein abundance of acyl-coenzyme A oxidases 1 (POX1, YALI0E32835g), 2 (POX2, YALI0F10857g), and 6 (POX6, YALI0E06567g) and multifunctional β-oxidation protein (MFE1, YALI0E15378g) (Figure 4-5A, 4-5B). However, CBS7504 also showed increased abundance of POX3 (YALI0D24750g), POX4 (YALI0E27654g) and 3-oxyacyl-thiolase (POT1, YALI0E18568g), all involved in the breakdown of TAGs into free fatty acids (FFA). Both strains decreased the abundance of acetyl-CoA carboxylase (ACC1, YALI0C11407g) but CBS7504 also decreased abundance of fatty acid synthase subunit 1 (FAS1, YALI0B15059g). Interestingly, YB420 increased the abundance of malic enzyme (ME, YALI0E18634g), which generates NADPH pools that are required for the FAS complex in oleaginous organisms (22), but shows little to no involvement in lipid production in *Y. lipolytica* (23-25).

Lipase. In total, ten lipases were identified with statistically significant abundance changes during the stationary phase (S3 and/or S4) relative to exponential growth phase (S1) (Figure 4-5B, 4-5C; Table 4-5). While none include the well-studied triacylglycerol lipase 3 (TGL3, YALI0D17534g) or TLG4 (YALI0F10010g), four of the identified lipases are annotated for TGL activity (PTHR23025). Interestingly, CBS7504 increased the protein abundance of all four TGLs in stationary phase while YB420 only increased the protein abundance of one.

NADPH generation. We identified five differentially abundant proteins involved in the generation of NADPH, the reducing equivalent required to sustain fatty acid synthesis (Figure 4-5B, 4-5C; Table 4-5). While both YB420 and CBS7504 strains increased the protein abundance of sorbitol dehydrogenase (MnDH1, YALI0B16192g) during stationary phase. Only YB420 increased abundances of the other 4 proteins including malic enzyme (ME, YALI0E18634g), succinate semialdehyde dehydrogenase (UGA2, YALI0F26191g), 6-phosphogluconolactonase (SOL3, YALI0E11671g) and NADP-dependent glucose-6-phosphate dehydrogenase (ZWF1, YALI0E22649g).

Regulators of lipid synthesis. Nitrogen limitation (i.e., high carbon to nitrogen ratio) is a common technique to increase lipid synthesis from glucose (26), and numerous studies have reported regulators of lipid accumulation and genes affected by nitrogen limitation. Our analysis identified eight of these regulators with statistically significant abundance changes during the stationary phase (S3 and/or S4) relative to exponential growth phase (S1) – all of which have been previously reported to influence lipid accumulation (10, 27, 28). Of these, YB420 increased the protein abundance of HLH transcription factor YAS2 (YALI0E32417g), a subunit of the SWI/SNF chromatin remodeling complex POR1 (YALI0D12628g), AMP-activated serine/threonine protein kinase SNF1 (YALI0D02101g) and heat shock transcription factor HSF1 (YALI0E13948g) (Figure 5C, Table S5). Meanwhile, CBS7504 increased the protein abundance of cytoplasmic pre-60S factor REI1 (YALI0B08734g) and decreased the protein abundance of a zinc finger protein (YALI0E30789g), sterol regulatory element binding protein UPC2 (YALI0B15818g) and SNF1-activating kinase 1 SAK1 (YALI0D08822g) (Figure 4-5B, Table 4-5).

Proteome alterations in regulatory elements. In total, 46 gene-specific regulator proteins (from panther and TFDB) were identified with statistically significant abundance changes during the stationary phase (S3 and/or S4) relative to exponential growth phase (S1) (Figure 4-6). Thirteen of these regulatory proteins had significant changes in both CBS7504 and YB420. The protein with the largest increase in abundance, YALI0C07821g, is annotated as glucose transport transcription regulator RGT1-related (PTHR31668) but only has the conserved domain, GAL4 (smart00066) with

S. cerevisiae RGT1p. Interestingly, four of these regulatory proteins have distinct, strain specific abundance patterns: RFX1, ASG1, STP1, and FHL1. RFX1, regulatory factor X in *S. cerevisiae*, is a major transcriptional repressor of DNA-damage-regulated genes (29). ASG1, an activator of stress-related genes, activates genes in β-oxidation, glucogenesis, glyoxylate cycle, triacylglycerol breakdown, peroxisomal transport, and helps assimilate fatty acids in *S. cerevisiae* (30). STP1, involved in species-specific tRNA processing, activates transcription of amino acid permease genes and is directly involved in pre-tRNA splicing in *S. cerevisiae* (31, 32). Finally, FHL1 (fork-head like) in *S. cerevisiae* functions as a transcription regulator of ribosomal protein transcription (33).

DISCUSSION

Lipid accumulation and lipid degradation patterns from cultures fermenting xylose (a non-preferred sugar prevalent in biomass hydrolysates) are complex phenotypes making them difficult to control and engineer in *Y. lipolytica* (14, 34, 35). By comparing proteomes of natively robust unconventional *Y. lipolytica* strain YB420 with the conventional strain CBS7504, we identified key proteins supporting cell growth and lipid accumulation with xylose as the sole remaining carbon source.

Once all the glucose was consumed, YB420 continued to accumulate lipids and sustained cell mass from xylose, while CBS7504 degraded lipids, decreased cell mass and produced more xylitol (Figure 4-2). This more efficient use of xylose, demonstrated by YB420, is supported by the greater number of PPP proteins upregulated during xylose assimilation, including Xyl2 and Xyl3 which are critical for flux of xylose through the PPP (Figure 4-4B). Meanwhile, the almost unchanged abundances of proteins found in CBS7504 in the PPP agrees with the increased xylitol secretion, decreased cell mass and lipid degradation phenotypes observed (Figure 4-4A). Interestingly, both strains showed similar xylose uptake profiles despite varying cell mass and lipid profiles. This suggests that transporters YALI0C06424g and YALI0F06776g are likely responsible and/or specific for xylose uptake, as indicated by increased protein abundances for both transporters across each strain only during the xylose assimilation phase (Figure 4-4D).

In lipid metabolism, YB420 increased proteins involved in TAG biosynthesis while CBS7504 increased proteins involved in β-oxidation and TAG lipase activity, strongly supporting the lipid maintenance and lipid degradation phenotypes of YB420 and CBS7504 respectively (Figure 4-5). Accordingly, YB420 increased the abundance of NADPH-generating enzymes which supply critical reducing-equivalents for lipid synthesis and xylose assimilation, including SOL3, ZWF1, ME and UGA2. Previously, overexpression of SOL3 increased lipid yield, titer and content (28). SOL3 does not produce NADPH directly, but instead catalyzes the intermediate reaction of the oxidative PPP that feeds into NADPH-producing enzymes, ZWF1 and NADP+-dependent 6-phosphogluconate dehydrogenase (28). Additionally, the promoters of ZWF1 along with ME and UGA2 exhibited increased expression levels in response to nitrogen limitation, a condition which results in increased lipid accumulation (25). Taken together, the increased protein abundance of NADPH-producing enzymes supports the lipid maintenance phenotype observed by YB420.

Several regulators previously reported to influence lipid accumulation in *Y. lipolytica* were captured by the proteomic analysis. Of these, YB420 cultures showed increased protein abundances of YAS2, POR1, SNF1 and HSF1. In a previous report, overexpression of YAS2 did not increase lipid accumulation in glucose minimal media but did significantly increase lipogenesis when acetate was the sole carbon source, suggesting indirect involvements in lipid biosynthesis in *Y. lipolytica* (28). Overexpression of POR1 in *Y. lipolytica* resulted in ~18% increased lipid content in glycerol media but showed growth defects in glucose media (27). Conversely, deletion of SNF1 was reported to increase fatty acid accumulation without the need for nitrogen limitation (10) and overexpression of HSF1 resulted in decreased lipid accumulation with glycerol as the sole carbon source (27). We also identified one regulator (REI1) with increase protein abundance and three regulators (YALI0E30789g, SAK1 and UPC2) with decreased abundance by CBS7504 during lipid degradation. Previously, overexpression of UPC2 decreased lipid accumulation while overexpression of YALI0E30789g and REI1 increased lipid accumulation, which does not support the lipid degradation phenotype of CBS7504 in our study (27). Furthermore, deletion of SAK1 has been shown to result in

increased fatty acid content (10) but in our study, decreased SAK1 protein levels were accompanied by lipid degradation in CBS7504.

In conclusion, our characterization of *Y. lipolytica* isolates with phenotypic and genetic diversity sheds light on those proteins supporting lipid accumulation or degradation during fermentation of non-preferred biomass sugar xylose. However, these results suggest that the regulation machinery of pentose and lipid metabolism in *Y. lipolytica* is complex and multifaceted, with many aspects remaining undiscovered and unexplained.

MATERIALS AND METHODS

Strains

Y. lipolytica strains CBS7504 (W29), YB392, YB419, YB420, YB566 and YB567 were used. CBS7504 is from CBS-KNAW Culture Collection in Utrecht, the Netherlands. All other strains are from ARS (NRRL) Culture Collection, Peoria, IL. The strains were stored in 20% glycerol at -80°C.

Medium and culturing conditions

Switchgrass hydrolysate preparation. Liberty switchgrass, that had been pelleted and cut with a 4mm knife mill, was used. The biomass was hydrolyzed at 20% solids w/w (6). A 20-gram dry weight of biomass was added to stainless steel vessels. Then 80 mL of 0.936 % sulfuric acid w/ 3.72 g/L Pluronic F-68 was added to each vessel. Eleven vessels per oven run were filled. The 12th vessel was filled with 80 mL of water and contained the thermocouple. The vessels were placed in a Mathis Labomat Infrared Oven. The following settings were used for the program: Temperature = 160°C, Heat Ramp = 2.6°C, Mix settings = 50 rpm, 60 seconds to the left and 60 seconds to the right, and cooling temperature = 40°C. Once the vessels have cooled, 4.0 mL of 1.0 M citrate buffer was added to each vessel. Then pH of the pretreated biomass was adjusted to 4.5-5.0 with 30% calcium hydroxide. The vessels were placed back in the Mathis oven for mixing at room temperature. The contents of 11 vessels were transferred to a Fernbach flask with a solid rubber stopper. The following enzymes were added: 29.7 mL Cellic

Ctec3 and 5.5 mL Cellic NS-22244. The Fernbach was incubated at 50°C with shaking at 125 rpm for 3 days. After 3 days the solids were removed using a 0.2 mm filter unit. The liquid fraction was stored at 4°C. Multiple batches were made over the course of a week. The batches were pooled. The liquid was then pH adjusted to 6.0 using 10N sodium hydroxide. After pH adjustment the SGH was filter sterilized and frozen at -20°C until week of use. The SGH was thawed at 4°C overnight. Prior to use, it was amended to a whole hydrolysate with 2.31 g/L $(NH_4)_2SO_4$, 1.81g/L Difco vitamin assay casamino acids, 0.018 g/L DL-tryptophan, and 0.072 g/L L-cysteine, and then diluted to 50% v/v with water. The diluted and amended hydrolysate is referred to 50% SGH form here on.

Culturing conditions. The yeast stocks were streaked on YPD Agar plates and incubated at 28°C for 24-48 hours. The plates were stored at 4°C until use. YPD media, 2 mL in a 16 mL tube, was inoculated by loop for pre-seed cultures. Pre-seed cultures were incubated at 28°C with shaking at 250 rpm for 18 hours. 0.5 mL of pre-seed culture was transferred to 10 mL 50 % SGH in 50 mL baffled flasks for seed cultures. Seed cultures were incubated 24 hours at 28°C with 250 rpm. The seed cultures were centrifuged to remove supernatant and resuspended in sterile water to A_{600} = 50. 150 mL of 50% SHG was inoculated at an A_{600} = 0.75. DasGip DasBox bioreactors were used for experimental cultures. Each strain was inoculated in triplicate. The following settings were used for the bioreactors: Beginning volume = 150 mL; Vessel = 250 mL; Temperature = 28°C; pH Set Point = 6.0; Agitation = 900 rpm; Aeration = 9.0 L/h; Base/Acid control: use 2 M HCl and 2 M NaOH for automatic dosing; Data Collection = dissolved oxygen, temperature, and pH. Cognis Clerol FBA 3107 antifoam was used to control foaming. After inoculation, 200uL of antifoam was added to each vessel. Antifoam was then added by pipet as need for duration of culture growth. A 1.2 – 1.5mL sample was taken for A600, residual sugars and lipids 3 times a day. A 1.0mL aliquot was removed from each sample for residual sugars and lipid analysis. The 1.0mL aliquot was centrifuged to remove the supernatant for residual sugar analysis. The cell pellet was washed twice with deionized water and resuspended up to 1.0mL with water. The samples were frozen at -20°C until analysis. The remaining sample was diluted for A_{600} measurement. Duplicate samples (2.0 mL) for proteomic analysis were taken once the

OD reaches <4.0, 2-3 hours later, at 44-48 hours and a final sample at 68-72 hours. Samples were kept cold while processing. The samples were centrifuged to remove supernatant, washed with 1.0mL of chilled water and then centrifuged again to remove water. The washed cell pellets were the stored at -80°C.

Analytical methods

Lipid quantification. Lipid analysis was done using a sulfo-phospho-vanillin colorimetric assay as previously reported by Dien *et al.* (36). For each sample, 1.0mL of sulfuric acid was added to a glass tube and 50 mL of sample (diluted with water if needed). The tube was heated at 100°C in a dry bath for 10 minutes. After heating, the tubes were cooled in a room temp water bath for 10 minutes. Once cooled, 2.5 mL of the vanillin-phosphoric acid solution was added to each tube. The tubes were mixed and placed in a 37°C incubator for 15 minutes. They are then cooled in a room temp water bath. The absorbance is then measured at 530 nm. The Vanillin-phosphoric acid solution (0.12g vanillin, 20 mL water, and 80 mL 85% o-phosphoric acid) is made fresh daily for assays. A blank with 50uL water and four calibration standards are used for the standard curve. The calibration standards are dilutions of corn oil dissolved in 2:1 (v/v) chloroform/methanol and 50 mL of each standard was processed in duplicate along with the samples.

Metabolites and sugar quantification. Residual sugars were measured on a Thermo High-Performance Liquid Chromatography system. The system used a Biorad HPX-87H column and a refractive index detector. The column was kept at 65°C with 0.6mL/min of 5mM sulfuric acid as a mobile phase.

LC/MS for proteomic analysis. *Y. lipolytica* cells harvested at the time points detailed above were resuspended in 100 mM Tris-HCl, 10 mM dithiothreitol, pH 8.0 and bead beat with 0.5 mm zirconium oxide beads in a Geno/Grinder® 2010 (SPEX SamplePrep) for 5 min at high speed (1750 rpm). Samples were adjusted to 4% SDS and incubated at 95°C for 10 min. Crude lysate was then cleared via centrifugation (21,000 x g) and quantified by corrected absorbance (Scopes) at 205 nm (NanoDrop OneC; Thermo Fisher). Samples were then treated with 30 mM iodoacetamide for 20 min at room

temperature in the dark. Three hundred micrograms of crude protein were then processed by protein aggregation capture (PAC) (37). Briefly, 300 ug of magnetic beads (1 micron, SpeedBead Magnetic Carboxylate; GE Healthcare UK) was suspended in each sample and protein aggregation was induced by adjusting the sample to 70% acetonitrile. Aggregated proteins were then washed with 1 mL of neat acetonitrile followed by 70% ethanol, and aggregated protein pellet digested with 1:75 (w/w) proteomics-grade trypsin (Pierce) in 100 mM Tris-HCl, pH 8.0 overnight at 37°C and again for 4 h the following day. Tryptic peptides released from the beads were then acidified to 0.5% formic acid, filtered through a 10 kDa MWCO spin filter (Vivaspin500; Sartorius), and quantified by NanoDrop OneC.

Peptide samples were analyzed by automated 1D LC-MS/MS analysis using a Vanquish UHPLC plumbed directly in-line with a Q Exactive Plus mass spectrometer (Thermo Scientific) outfitted with a trapping column coupled to an in-house pulled nanospray emitter as previously described (38). The trapping column (100 µm ID) was packed with 10 cm of 5 µm Kinetex C18 RP resin (Phenomenex) while the nanospray emitter (75 µm ID) was packed with 15 cm of 1.7 µm Kinetex C18 RP resin. For each sample, 3 µg of peptides were loaded, desalted, and separated by uHPLC with the following conditions: sample injection followed by 100% solvent A (95% H_2O, 5% acetonitrile, 0.1% formic acid) chase from 0-30 min (load and desalt), linear gradient 0% to 30% solvent B (70% acetonitrile, 30% water, 0.1% formic acid) from 30-220 min (separation), and column re-equilibration at 100% solvent A from 220-240 min. Eluting peptides were measured and sequenced by data-dependent acquisition on the Q Exactive MS as previously described (39).

Bioinformatics and data analysis

Comparative genomics. The following genome assemblies of *Y. lipolytica* strains were downloaded from NCBI as genbank files: CBS7504/CLIB89/W29 (GCA_001761485.1), CLIB122 (GCA_000002525.1), YB392 (GCA_003367865.1), YB419 (GCA_003367925.1), YB420 (GCA_003367965.1), YB566 (GCA_003367945.1) and YB567 (GCA_003367845.1). These were imported individually into a KBase

narrative as genomes and combined into a genome set using the Build GenomeSet v1.0.1 application. A phylogenetic tree was constructed using Insert Set of Genomes Into Species Tree 2.1.10 application with neighbor public genome count of 20. Orthologue genes and unique genes shared between isolates were identified with the Compute Pangenome application using the genome set of the 7 isolates as the input. Finally, Pangenome Circle Plot - v1.2.0 application was used to produce a list of singleton genes for each base genome identified from the pangenome.

Proteomics. MS/MS spectra were searched against the *Y. lipolytica* CLIP122 proteome (UniProt; Nov18 build) appended with non-redundant proteins from strain YB-420 and common protein contaminants using the MS Amanda v.2.0 algorithm in Proteome Discoverer v.2.3 (ThermoScientific). Peptide spectrum matches (PSM) were required to be fully tryptic with 2 mis-cleavages; a static modification of 57.0214 Da on cysteine (carbamidomethylated) and a dynamic modification of 15.9949 Da on methionine (oxidized) residues. Peptide spectrum matches (PSM) were scored and filtered using the Percolator node in Proteome Discoverer and false-discovery rates initially controlled at < 1% at both the PSM- and peptide-levels. Peptides were then quantified by chromatographic area-under-the-curve, mapped to their respective proteins, and areas summed to estimate protein-level abundance. Proteins without 3 valid values in a minimum of 1 biological condition were removed and remaining protein abundances were log2 transformed. Missing values were imputed to simulate the mass spectrometer's limit of detection using Perseus v1.6.10.43 (i.e., normal distribution, width of 0.3, down shift of 1.9 and mode changed to total matrix) (40). Significant differences in protein abundance were calculated via T-test for each sample (S2, S3 and S4) against control group (S1 sample) for each strain using FDR of 0.05, 250 permutations and s0 of 1.

All raw mass spectra for quantification of proteins used in this study have been deposited in the MassIVE and ProteomeXchange data repositories under accession numbers MSV000085941 (MassIVE) and PXD020854 (ProteomeXchange), with data files available at ftp://massive.ucsd.edu/ MSV000085941.

Bioinformatics.

Pathway proteins were annotated using KEGG database (40) and from literature sources where cited. Ontology associations and orthologs for regulator proteins were identified using panther database (41).

ACKNOWLEDGEMENT

We would like to acknowledge financial support from the DOE BER Genomic Science Program (DE-SC0019412 and FWP 3ERKP921). The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the funding agencies. We also acknowledge the excellent technical support of Rebecca Splitt for the bioreactor studies.

References

1. Beopoulos A, Cescut J, Haddouche R, Uribelarrea J-L, Molina-Jouve C, Nicaud J-M. 2009. Yarrowia lipolytica as a model for bio-oil production. Progress in Lipid Research 48:375-387.
2. Epova E, Shevelev A, Belyakova A, Isakova E, Kovalyov L, Guseva M, Zylkova M, Deryabina Y. 2012. Identification of proteins involved in pH adaptation in extremophile yeast Yarrowia lipolytica. INTECH Open Access Publisher.
3. Andreishcheva E, Isakova E, Sidorov N, Abramova N, Ushakova N, Shaposhnikov G, Soares M, Zvyagilskaya R. 1999. Adaptation to salt stress in a salt-tolerant strain of the yeast *Yarrowia lipolytica*. BIOCHEMISTRY C/C OF BIOKHIMIIA 64:1061-1067.
4. Ryu S, Labbé N, Trinh CT. 2015. Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α-ketoglutaric acid by Yarrowia lipolytica. Applied Microbiology and Biotechnology 99:4237-4244.
5. Walker C, Ryu S, Trinh CT. 2019. Exceptional solvent tolerance in Yarrowia lipolytica is enhanced by sterols. Metabolic Engineering 54:83-95.
6. Quarterman J, Slininger PJ, Kurtzman CP, Thompson SR, Dien BS. 2017. A survey of yeast from the Yarrowia clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate. Applied Microbiology and Biotechnology 101:3319-3334.
7. Dulermo T, Nicaud J-M. 2011. Involvement of the G3P shuttle and β-oxidation pathway in the control of TAG synthesis and lipid accumulation in Yarrowia lipolytica. Metabolic Engineering 13:482-491.
8. Tai M, Stephanopoulos G. 2013. Engineering the push and pull of lipid biosynthesis in oleaginous yeast Yarrowia lipolytica for biofuel production. Metabolic Engineering 15:1-9.
9. Beopoulos A, Mrozova Z, Thevenieau F, Le Dall M-T, Hapala I, Papanikolaou S, Chardot T, Nicaud J-M. 2008. Control of Lipid Accumulation in the Yeast *Yarrowia lipolytica*. Applied and Environmental Microbiology 74:7779.

10. Seip J, Jackson R, He H, Zhu Q, Hong S-P. 2013. Snf1 Is a Regulator of Lipid Accumulation in *Yarrowia lipolytica*. Applied and Environmental Microbiology 79:7360.
11. Blazeck J, Hill A, Liu L, Knight R, Miller J, Pan A, Otoupal P, Alper HS. 2014. Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. Nature Communications 5:3131.
12. Rodriguez GM, Hussain MS, Gambill L, Gao D, Yaguchi A, Blenner M. 2016. Engineering xylose utilization in Yarrowia lipolytica by understanding its cryptic xylose pathway. Biotechnology for Biofuels 9:149.
13. Ryu S, Hipp J, Trinh CT. 2016. Activating and Elucidating Metabolism of Complex Sugars in *Yarrowia lipolytica*. Applied and Environmental Microbiology 82:1334.
14. Niehus X, Crutz-Le Coq A-M, Sandoval G, Nicaud J-M, Ledesma-Amaro R. 2018. Engineering Yarrowia lipolytica to enhance lipid production from lignocellulosic materials. Biotechnology for Biofuels 11:11.
15. Prabhu AA, Ledesma-Amaro R, Lin CSK, Coulon F, Thakur VK, Kumar V. 2020. Bioproduction of succinic acid from xylose by engineered Yarrowia lipolytica without pH control. Biotechnology for Biofuels 13:113.
16. Yao F, Liu S-C, Wang D-N, Liu Z-J, Hua Q, Wei L-J. 2020. Engineering oleaginous yeast Yarrowia lipolytica for enhanced limonene production from xylose and lignocellulosic hydrolysate. FEMS Yeast Research 20.
17. Ledesma-Amaro R, Lazar Z, Rakicka M, Guo Z, Fouchard F, Coq A-MC-L, Nicaud J-M. 2016. Metabolic engineering of Yarrowia lipolytica to produce chemicals and fuels from xylose. Metabolic Engineering 38:115-124.
18. Ryu S, Trinh CT. 2018. Understanding Functional Roles of Native Pentose-Specific Transporters for Activating Dormant Pentose Metabolism in *Yarrowia lipolytica*. Applied and Environmental Microbiology 84:e02146-17.
19. Walker C, Ryu S, Na H, Zane M, LaButti K, Lipzen A, Haridas S, Barry K, Grigoriev IV, Quarterman J, Slininger P, Dien B, Trinh CT. 2018. Draft Genome Assemblies of Five Robust *Yarrowia lipolytica* Strains Exhibiting High Lipid Production, Pentose Sugar Utilization, and Sugar Alcohol Secretion from Undetoxified

Lignocellulosic Biomass Hydrolysates. Microbiology Resource Announcements 7:e01040-18.

20. Bellasio M, Peymann A, Steiger MG, Valli M, Sipiczki M, Sauer M, Graf AB, Marx H, Mattanovich D. 2016. Complete genome sequence and transcriptome regulation of the pentose utilizing yeast Sugiyamaella lignohabitans. FEMS Yeast Research 16.
21. Ozcan S, Dover J, Rosenwald AG, Wölfl S, Johnston M. 1996. Two glucose transporters in Saccharomyces cerevisiae are glucose sensors that generate a signal for induction of gene expression. Proceedings of the National Academy of Sciences 93:12428.
22. Ratledge C. 2014. The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: a reappraisal and unsolved problems. Biotechnology Letters 36:1557-1568.
23. Zhang H, Zhang L, Chen H, Chen YQ, Ratledge C, Song Y, Chen W. 2013. Regulatory properties of malic enzyme in the oleaginous yeast, Yarrowia lipolytica, and its non-involvement in lipid accumulation. Biotechnology Letters 35:2091-2098.
24. Qiao K, Wasylenko TM, Zhou K, Xu P, Stephanopoulos G. 2017. Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism. Nature Biotechnology 35:173-177.
25. Liu H, Marsafari M, Deng L, Xu P. 2019. Understanding lipogenesis by dynamically profiling transcriptional activity of lipogenic promoters in Yarrowia lipolytica. Applied Microbiology and Biotechnology 103:3167-3179.
26. Bellou S, Triantaphyllidou IE, Mizerakis P, Aggelis G. 2016. High lipid accumulation in Yarrowia lipolytica cultivated under double limitation of nitrogen and magnesium. Journal of Biotechnology 234:116-126.
27. Leplat C, Nicaud J-M, Rossignol T. 2018. Overexpression screen reveals transcription factors involved in lipid accumulation in Yarrowia lipolytica. FEMS Yeast Research 18.

28. Silverman AM, Qiao K, Xu P, Stephanopoulos G. 2016. Functional overexpression and characterization of lipogenesis-related genes in the oleaginous yeast Yarrowia lipolytica. Applied Microbiology and Biotechnology 100:3781-3798.
29. Huang M, Zhou Z, Elledge SJ. 1998. The DNA Replication and Damage Checkpoint Pathways Induce Transcription by Inhibition of the Crt1 Repressor. Cell 94:595-605.
30. Jansuriyakul S, Somboon P, Rodboon N, Kurylenko O, Sibirny A, Soontorngun N. 2016. The zinc cluster transcriptional regulator Asg1 transcriptionally coordinates oleate utilization and lipid accumulation in Saccharomyces cerevisiae. Applied Microbiology and Biotechnology 100:4549-4560.
31. Wang SS, Stanford DR, Silvers CD, Hopper AK. 1992. STP1, a gene involved in pre-tRNA processing, encodes a nuclear protein containing zinc finger motifs. Molecular and Cellular Biology 12:2633.
32. Jørgensen MU, Gjermansen C, Andersen HA, Kielland-Brandt MC. 1997. STP1, a gene involved in pre-tRNA processing in yeast, is important for amino-acid uptake and transcription of the permease gene BAP2. Current Genetics 31:241-247.
33. Rudra D, Zhao Y, Warner JR. 2005. Central role of Ifh1p–Fhl1p interaction in the synthesis of yeast ribosomal proteins. The EMBO Journal 24:533-542.
34. Li H, Alper HS. 2016. Enabling xylose utilization in Yarrowia lipolytica for lipid production. Biotechnology Journal 11:1230-1240.
35. Yook SD, Kim J, Gong G, Ko JK, Um Y, Han SO, Lee S-M. 2020. High-yield lipid production from lignocellulosic biomass using engineered xylose-utilizing Yarrowia lipolytica. GCB Bioenergy 12:670-679.
36. B. S. Dien, P. J. Slininger, C. P. Kurtzman, B. R. Moser, P. J. O'Bryan, Identification of superior lipid producing Lipomyces and Myxozyma yeasts. *AIMS Environ Sci* **3**, 1-20 (2016).
37. T. S. Batth *et al.*, Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteomics* **18**, 1027-1035 (2019).

38. C. Walker, S. Ryu, R. J. Giannone, S. Garcia, C. T. Trinh, Understanding and Eliminating the Detrimental Effect of Thiamine Deficiency on the Oleaginous Yeast Yarrowia lipolytica. *Applied and environmental microbiology* **86**, e02299-02219 (2020).
39. S. M. Clarkson *et al.*, Construction and Optimization of a Heterologous Pathway for Protocatechuate Catabolism in Escherichia coli Enables Bioconversion of Model Aromatic Compounds. *Applied and Environmental Microbiology* **83**, e01313-01317 (2017).
40. S. Tyanova *et al.*, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731-740 (2016).
41. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27-30 (2000).
42. H. Mi *et al.*, PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research* **49**, D394-D403 (2021).

Appendix

Table 4- 1. Singleton genes of YB392, YB419, YB420, YB566 and YB567.

Genome	Locus Tag
YB392	B0I71DRAFT_23206
YB392	B0I71DRAFT_165597
YB392	B0I71DRAFT_132072
YB392	B0I71DRAFT_140658
YB392	B0I71DRAFT_39083
YB419	B0I72DRAFT_42591
YB419	B0I72DRAFT_173937
YB419	B0I72DRAFT_34279
YB419	B0I72DRAFT_7217
YB419	B0I72DRAFT_133747
YB419	B0I72DRAFT_148340
YB419	B0I72DRAFT_20302
YB419	B0I72DRAFT_176868
YB419	B0I72DRAFT_176347
YB419	B0I72DRAFT_164308
YB419	B0I72DRAFT_15024
YB419	B0I72DRAFT_2773
YB419	B0I72DRAFT_148966
YB419	B0I72DRAFT_138964
YB419	B0I72DRAFT_134698
YB419	B0I72DRAFT_128204
YB419	B0I72DRAFT_141486
YB419	B0I72DRAFT_6483
YB419	B0I72DRAFT_26882
YB419	B0I72DRAFT_159127
YB419	B0I72DRAFT_164433
YB419	B0I72DRAFT_172193
YB419	B0I72DRAFT_142595
YB420	B0I73DRAFT_171833
YB420	B0I73DRAFT_39464
YB420	B0I73DRAFT_141110
YB420	B0I73DRAFT_11141
YB420	B0I73DRAFT_166406
YB420	B0I73DRAFT_141163
YB420	B0I73DRAFT_129575
YB420	B0I73DRAFT_168289

Table 4-1 continued

YB420	B0I73DRAFT_10053
YB420	B0I73DRAFT_20235
YB420	B0I73DRAFT_129528
YB420	B0I73DRAFT_138424
YB420	B0I73DRAFT_26166
YB420	B0I73DRAFT_136916
YB420	B0I73DRAFT_162507
YB420	B0I73DRAFT_141781
YB420	B0I73DRAFT_144199
YB566	B0I74DRAFT_176766
YB566	B0I74DRAFT_42025
YB567	B0I75DRAFT_2109

Table 4- 2. Exclusively shared genes between xylose consuming strains YB566 and YB567.

YB567	**YB566**
B0I75DRAFT_157292	B0I74DRAFT_172414
B0I75DRAFT_40401	B0I74DRAFT_40877
B0I75DRAFT_11481	B0I74DRAFT_21615
B0I75DRAFT_140064	B0I74DRAFT_140479
B0I75DRAFT_154119	B0I74DRAFT_152545
B0I75DRAFT_155102	B0I74DRAFT_149309
B0I75DRAFT_128836	B0I74DRAFT_174097
B0I75DRAFT_134495	B0I74DRAFT_134394
B0I75DRAFT_169517	B0I74DRAFT_163910
B0I75DRAFT_137511	B0I74DRAFT_138950

Table 4- 3. Exclusively shared genes between lipid accumulating strains YB392, YB419 and YB420.

YB392	YB419	YB420
B0I71DRAFT_39511	B0I72DRAFT_38745	B0I73DRAFT_40681
B0I71DRAFT_12264	B0I72DRAFT_37180	B0I73DRAFT_14480
B0I71DRAFT_131081	B0I72DRAFT_138926	B0I73DRAFT_133629
B0I71DRAFT_9972	B0I72DRAFT_12030	B0I73DRAFT_6620
B0I71DRAFT_130434	B0I72DRAFT_134889	B0I73DRAFT_131291
B0I71DRAFT_129034	B0I72DRAFT_135679	B0I73DRAFT_132750

Table 4- 4. Proteomics analysis of xylose metabolism and transporters.

				Difference against S1 (t-test)			
Name	**Symbol**	**Locus Tag**	**Accession**	**CBS S3**	**CBS S4**	**YB420 S3**	**YB420 S4**
ribose-phosphate pyrophosphokinase	PRS1	YALI0B13552g	Q6CER7	-1.97	-4.05		
NADP+-dependent glucose-6-phosphate dehydrogenase	ZWF1	YALI0E22649g	Q6C4Y7			1.13	1.46
6-phosphogluconolactonase	SOL3	YALI0E11671g	Q6C682			1.12	1.08
ribose 5-phosphate isomerase B	RPIB	YALI0F01628g	Q6C393			3.65	2.05
Transketolase	TKL	YALI0D02277g	Q6CAJ3	6.77	6.92	3.14	3.98
Xylitol dehydrogenase	Xyl2	YALI0E12463g	Q6C648			1.73	1.98
D-arabinitol 2-dehydrogenase	ADH	YALI0F02211g	Q6C367			1.42	1.57
Xylulokinase	Xyl3	YALI0F10923g	Q6C246			2.56	3.45
ribulo and xylulose kinase		YALI0D15114g	Q6C916	1.68	1.69	2.48	2.40
ribulo kinase		YALI0E13321g	Q6C615	1.30	2.09	2.98	3.55

Table 4- 5. Proteomics analysis of lipid metabolism, lipid regulators, lipase and NADPH generating proteins.

				Difference against S1 (t-test)			
Name/Description	**Symbol**	**Locus Tag**	**Accession**	**CBS S3**	**CBS S4**	**YB420 S3**	**YB420 S4**
Lysophospholipid acyltransferase	ALE1	YALI0F19514g	Q6C138			2.66	2.12
Phosphatidate phosphatase	PAP	YALI0D27016g	Q6C7L9			2.06	2.37
Acyl-coenzyme A oxidase	POX6	YALI0E06567g	Q6C6T0	1.71	1.78	1.77	1.99
Acyl-coenzyme A oxidase 1	POX1	YALI0E32835g	O74934	3.36	3.71	0.822	2.54
Lipid phosphate phosphatase 1/Diacylglycerol pyrophosphate phosphatase 1	LPP1	YALI0B14531g	Q6CEM4			-3.62	-3.15
Acyl-coenzyme A oxidase 3	POX3	YALI0D24750g	O74936	1.37	1.15		
Acyl-coenzyme A oxidase	POX5	YALI0C23859g	F2Z630			1.95	1.57
Diacylglycerol acyltransferase	DGA1	YALI0E32769g	Q6C3R2	4.40	3.40	2.72	3.54
acylglycerol lipase	YJU3	YALI0C14520g	Q6CBY1			-0.204	-2.14
Acyl-coenzyme A oxidase 2	POX2	YALI0F10857g	O74935	2.72	2.72	1.34	1.71
Acyl-CoA:sterol acyltransferase	ARE1	YALI0F06358g	Q6C2N7	3.66	3.97	4.30	4.67
Acyl-coenzyme A oxidase	POX4	YALI0E27654g	F2Z627	1.31	1.51		
3-ketoacyl-CoA thiolase	POT1	YALI0E18568g	Q05493	1.16	1.44		
Glycerol-3-phosphate acyltransferase	SCT1	YALI0C00209g	Q6CDI7	1.57	1.45		
2nd step of β-oxidation	MFE1	YALI0E15378g	F2Z6I5	1.54	1.57	1.42	1.56
Acetyl-CoA Carboxylase	ACC1	YALI0C11407g	Q6CC91	-1.50	-1.63	-1.44	-1.73
Fatty acid synthase 1	FAS1	YALI0B15059g	P34229	-1.25	-1.22		
Malic enzyme	ME	YALI0E18634g	Q6C5F0			2.11	2.23
TAG lipase		YALI0B09361g	F2Z5Z7	7.28	8.52	0.356	3.98
TAG lipase		YALI0A20350g	F2Z685	2.88	2.03		
TAG lipase		YALI0E17655g	Q6C5J1	2.86	2.82		
TAG lipase		YALI0E08492g	Q6C6L5	2.87	1.08		
Mannitol dehydrogenase	MnDH1	YALI0B16192g	Q6CEE9	3.32	2.99	2.40	1.72
Succinate semialdehyde dehydrogenase	UGA2	YALI0F26191g	Q6C0B4			1.13	1.26
6-Phosphogluconolactonase	SOL3	YALI0E11671g	Q6C682			1.12	1.08
NADP+-dependent glucose-6-phosphate dehydrogenase	ZWF1	YALI0E22649g	Q6C4Y7			1.13	1.46
HLH transcription factor, upregulates cytochrome p450 genes in response to alkanes with YAS1; some similarity to INO4/INO2 from Saccharomyces cerevisiae	YAS2	YALI0E32417g	Q6C3S8			2.45	3.26

Table 4-5 continued

Y. lipolytica YALI1D12628gPOR1 Subunit of the SWI/SNF chromatin remodelingcomplex involved in transcriptional regulation	POR1	YALI0D12628g	Q6C9A9			0.255	1.31
Snf1 kinase	SNF1	YALI0D02101g	Q6CAK0			1.48	2.50
Snf1 Activating Kinase 1	SAK1	YALI0D08822g	Q6C9S1	-3.07	-1.83		
Cytoplasmic pre-60S factor, putative similar to S. cerevisiae YBR267W REI1 Cytoplasmic pre-60S factor	REI1	YALI0B08734g	Q6CFB1	2.14	1.51		
Conserved hypothetical protein weakly similar to Ajellomyces dermatitidis C2H2 finger domain-containingprotein		YALI0E30789g	Q6C407	-1.79	-2.30		
Conserved hypothetical protein some similarities with S. cerevisiae YGL073w HSF1 heat shock transcription factor	HSF1	YALI0E13948g	Q6C5Z0			1.55	1.51
Conserved hypothetical protein some similarities with S. cerevisiae YDR213W UPC2 Sterol regulatory element binding protein	UPC2	YALI0B15818g	Q6CEG5	-2.95	-3.92		

Figure 4- 1. (**A**) Phylogenetic tree of *Yarrowia lipolytica* isolates with 20 closest neighbor species. (**B**) Pangenome of *Y. lipolytica* reference strains CLIB122122 and CBS7504 (W29) and non-conventional strains YB392, YB419, YB420, YB566 and YB567. Bold, strain; parenthesis, total genes; outer petals, singleton genes, middle petals, uniquely shared genes between lipid producers (blue) and xylose consumers (green); middle circle, core genes.

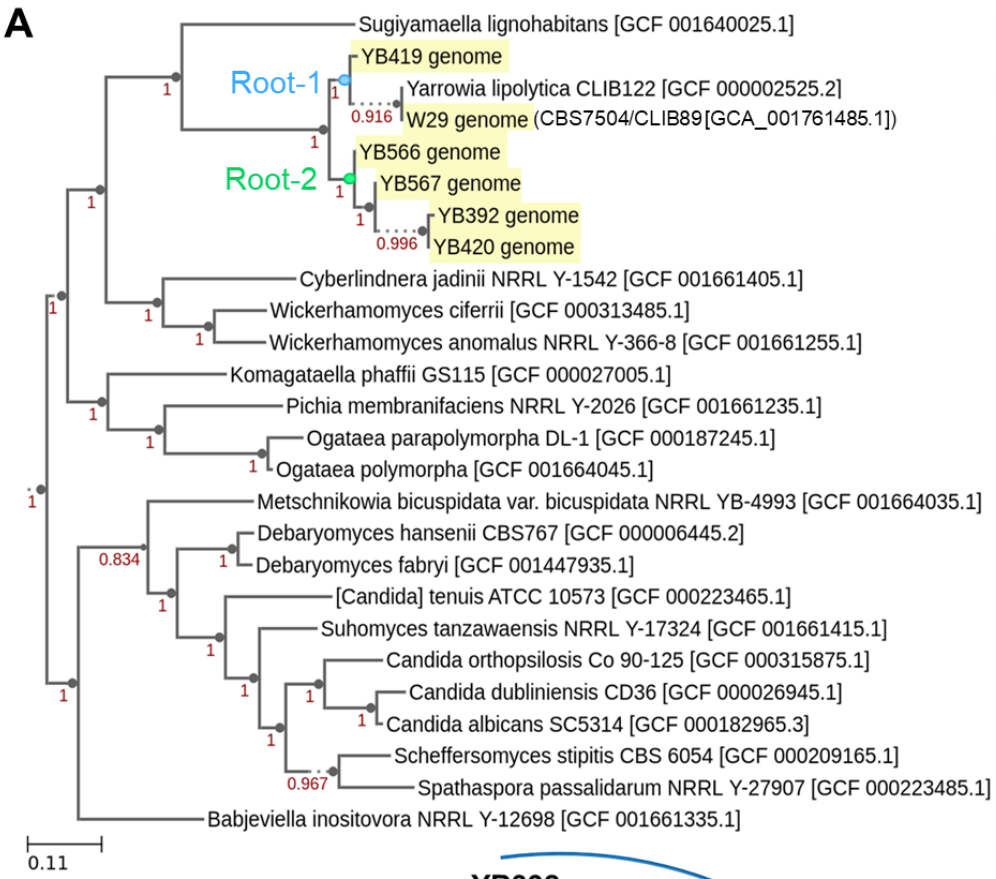
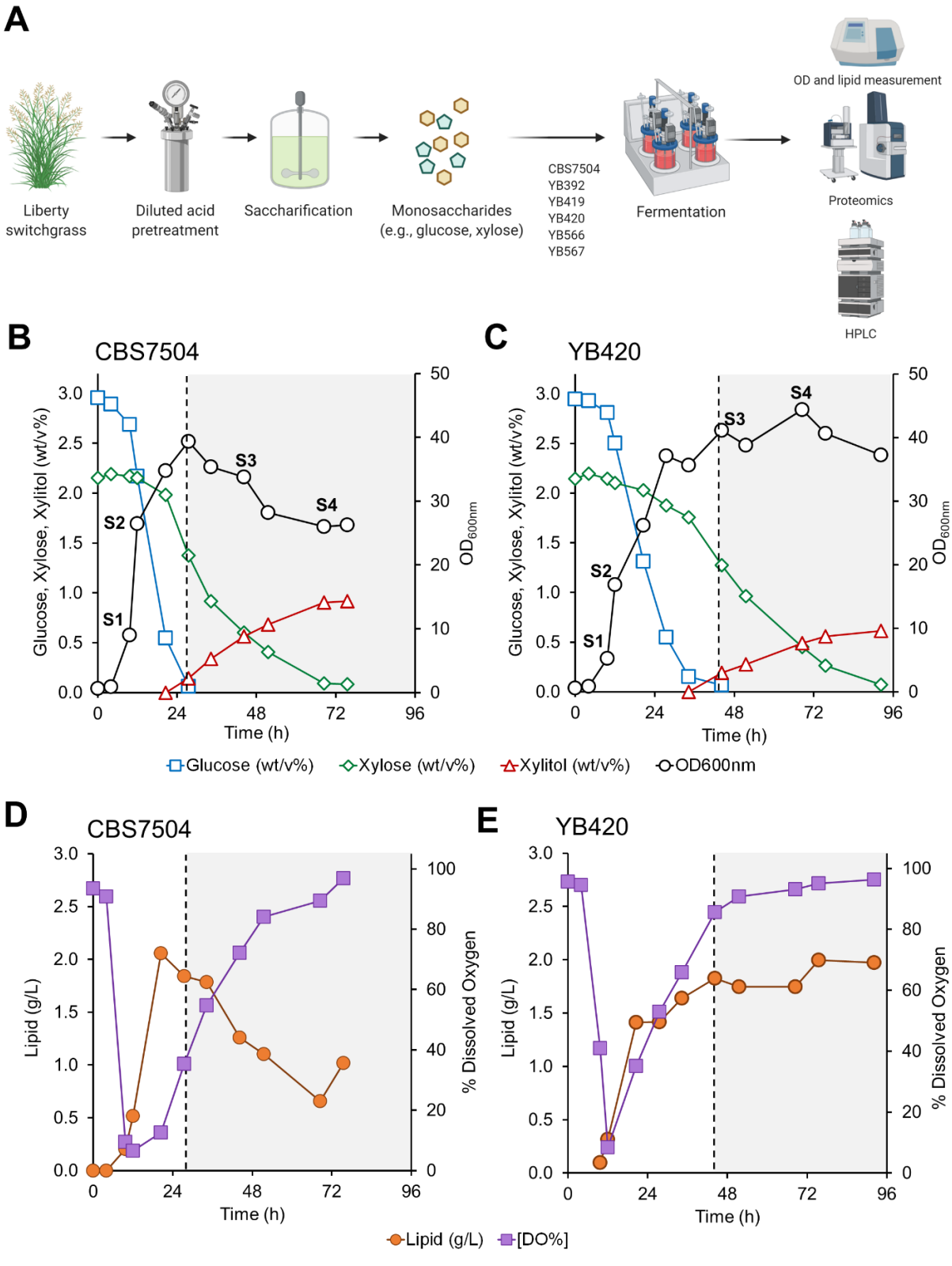

A
Sugiyamaella lignohabitans [GCF 001640025.1]
YB419 genome
Root-1
Yarrowia lipolytica CLIB122 [GCF 000002525.2]
W29 genome (CBS7504/CLIB89[GCA_001761485.1])
YB566 genome
Root-2
YB567 genome
YB392 genome
YB420 genome
Cyberlindnera jadinii NRRL Y-1542 [GCF 001661405.1]
Wickerhamomyces ciferrii [GCF 000313485.1]
Wickerhamomyces anomalus NRRL Y-366-8 [GCF 001661255.1]
Komagataella phaffii GS115 [GCF 000027005.1]
Pichia membranifaciens NRRL Y-2026 [GCF 001661235.1]
Ogataea parapolymorpha DL-1 [GCF 000187245.1]
Ogataea polymorpha [GCF 001664045.1]
Metschnikowia bicuspidata var. bicuspidata NRRL YB-4993 [GCF 001664035.1]
Debaryomyces hansenii CBS767 [GCF 000006445.2]
Debaryomyces fabryi [GCF 001447935.1]
[Candida] tenuis ATCC 10573 [GCF 000223465.1]
Suhomyces tanzawaensis NRRL Y-17324 [GCF 001661415.1]
Candida orthopsilosis Co 90-125 [GCF 000315875.1]
Candida dubliniensis CD36 [GCF 000026945.1]
Candida albicans SC5314 [GCF 000182965.3]
Scheffersomyces stipitis CBS 6054 [GCF 000209165.1]
Spathaspora passalidarum NRRL Y-27907 [GCF 000223485.1]
Babjeviella inositovora NRRL Y-12698 [GCF 001661335.1]
0.11
B
YB392
(6,750)
Lipid
Producers
5
Clib122
(6,471)
20
23
YB419
(6,751)
Reference
Strains
Singletons
6
Core
5,593
Uniquely
shared
CBS7504
(7,949)
1,343
17
YB420
(6,772)
10
YB567
(6,776)
1
2
YB566
(6,764)
Xylose
Consumers

Figure 4- 2**.** (**A**) Workflow of bioreactor characterization of *Yarrowia lipolytica* isolates CBS7504 (**B, D**) and YB420 (**C, E**) in 50% SGH. Dotted lines, time of glucose depletion where xylose is the sole remaining carbon source.


A
Liberty switchgrass
Diluted acid pretreatment
Saccharification
Monosaccharides (e.g., glucose, xylose)
CBS7504
YB392
YB419
YB420
YB566
YB567
Fermentation
OD and lipid measurement
Proteomics
HPLC
B
CBS7504
C
YB420
Glucose, Xylose, Xylitol (wt/v%)
OD600nm
Time (h)
S1
S2
S3
S4
Glucose (wt/v%)
Xylose (wt/v%)
Xylitol (wt/v%)
OD600nm
D
CBS7504
E
YB420
Lipid (g/L)
% Dissolved Oxygen
Time (h)
Lipid (g/L)
[DO%]

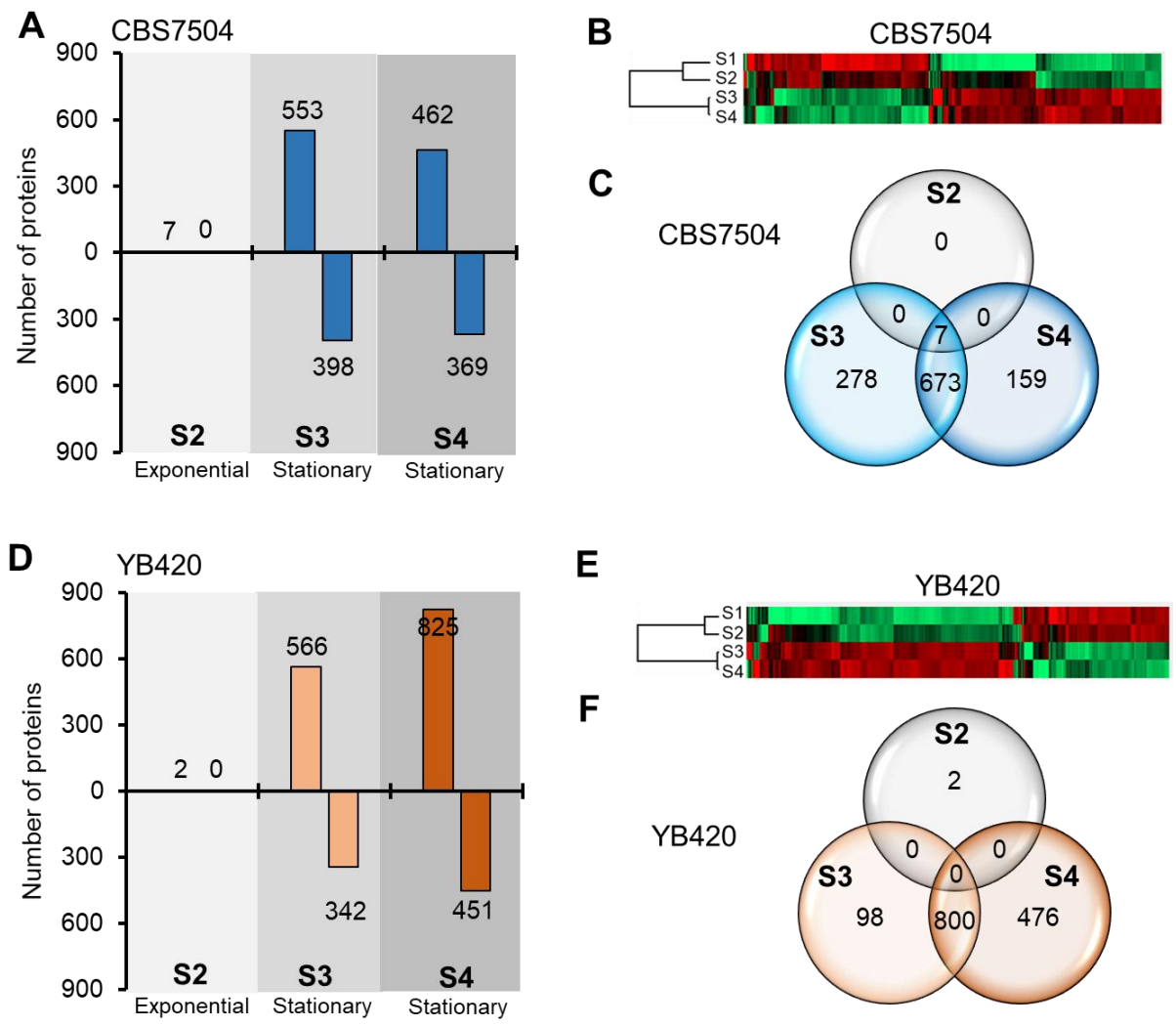


Figure 4- 3**.** Proteomic analysis of *Y. lipolytica* (**A-C**) CBS7504 and (**D-F**) YB420 strains. (**A, D**) Number of proteins with significant abundance changes relative to S1. (**B, E**) Heatmap of significant proteins with changed abundance relative to S1. (**C, F**) Venn diagram illustrating the number of proteins with significant abundance changes between samples.

Figure 4- 4. Pentose phosphate pathway proteome of (**A**) CBS7504 and (**B**) YB420. Increased protein abundance during xylose assimilation, red; decreased protein abundance during xylose assimilation, blue; pathways, rectangles; proteins, bold font; metabolites, plain text. (**C**) All transporters and (**D**) carbohydrate transporters with significant protein abundance changes in the xylose assimilation phase. CBS7504, blue; YB420, orange. Abbreviations: TKL (transketolase, YALI0D02277g); PRS1 (ribose-phosphate pyrophosphokinase, YALI0B13552g); Xyl1 (xylose reductase, YALI0D07634g); Xyl2 (xylitol dehydrogenase, YALI0E12463g); Xyl3 (xylulokinase, YALI0F10923g); ZWF1 (NADP+-dependent glucose-6-phosphate dehydrogenase, YALI0E22649g); SOL3 (6-phosphogluconolactonase, YALI0E11671g); RPIB (ribose 5-phosphate isomerase, YALI0F01628g); TAL (transaldolase, YALI0F15587g); RPE1 (ribulose-phosphate 3-epimerase, YALI0C11880g); GND1 (6-phosphogluconate dehydrogenase, YALI0B15598g).

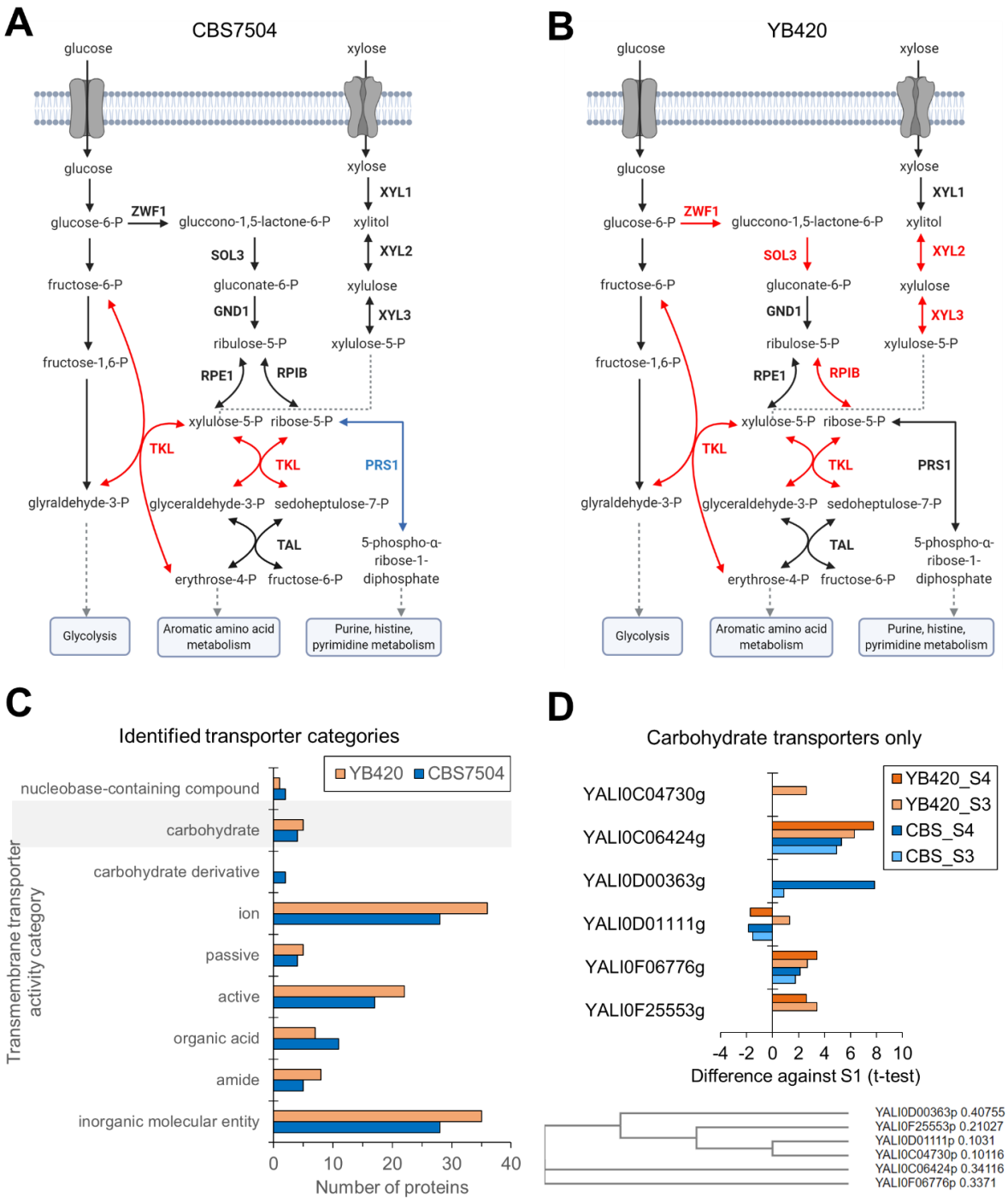
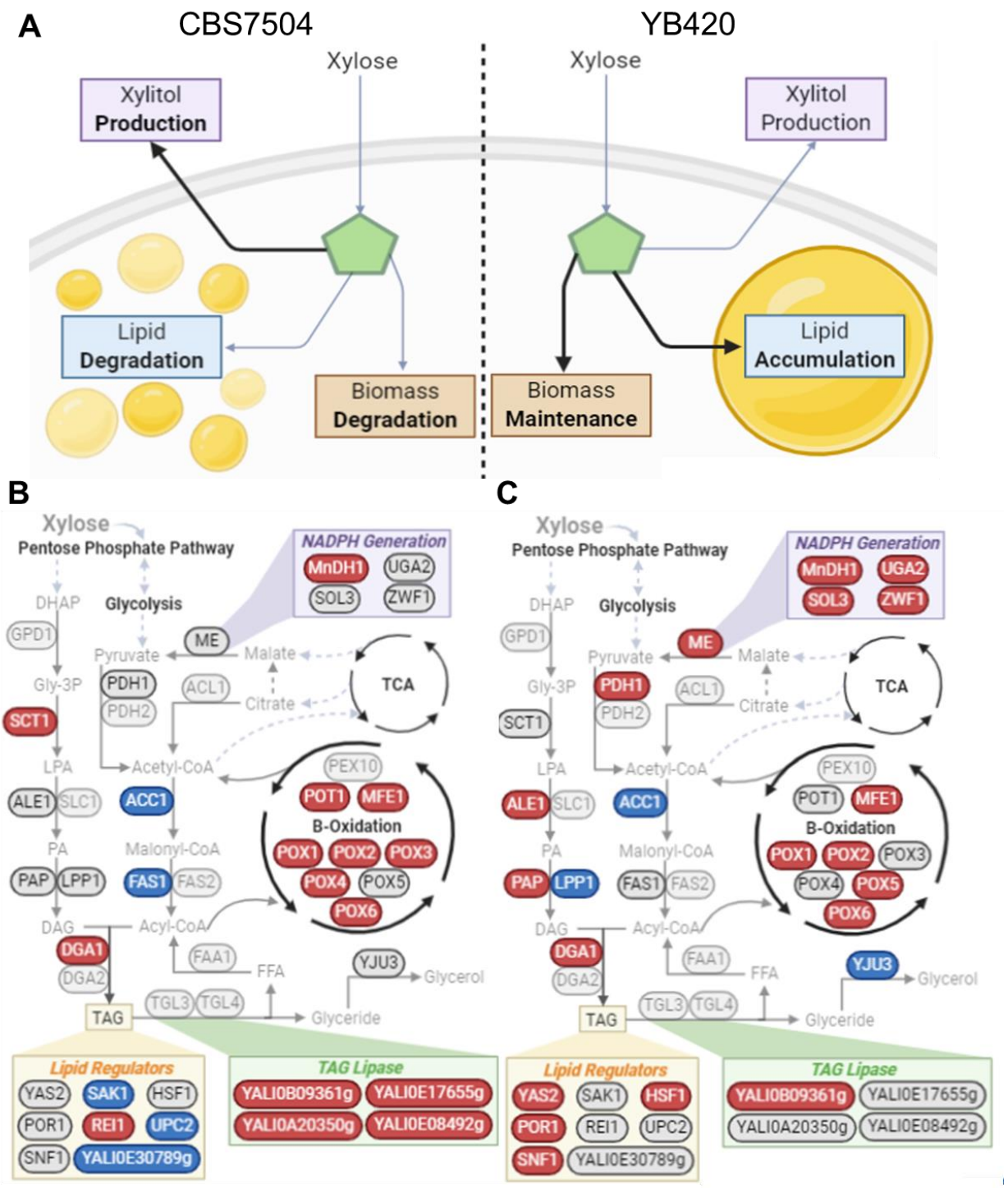

A
CBS7504
glucose
xylose
glucose
xylose
XYL1
glucose-6-P
ZWF1
gluccono-1,5-lactone-6-P
xylitol
SOL3
XYL2
fructose-6-P
gluconate-6-P
xylulose
GND1
XYL3
ribulose-5-P
xylulose-5-P
fructose-1,6-P
RPE1
RPIB
xylulose-5-P
ribose-5-P
TKL
TKL
PRS1
glyraldehyde-3-P
glyceraldehyde-3-P
sedoheptulose-7-P
TAL
5-phospho-α-ribose-1-diphosphate
erythrose-4-P
fructose-6-P
Glycolysis
Aromatic amino acid metabolism
Purine, histine, pyrimidine metabolism
B
YB420
glucose
xylose
glucose
xylose
XYL1
glucose-6-P
ZWF1
gluccono-1,5-lactone-6-P
xylitol
SOL3
XYL2
fructose-6-P
gluconate-6-P
xylulose
GND1
XYL3
ribulose-5-P
xylulose-5-P
fructose-1,6-P
RPE1
RPIB
xylulose-5-P
ribose-5-P
TKL
TKL
PRS1
glyraldehyde-3-P
glyceraldehyde-3-P
sedoheptulose-7-P
TAL
5-phospho-α-ribose-1-diphosphate
erythrose-4-P
fructose-6-P
Glycolysis
Aromatic amino acid metabolism
Purine, histine, pyrimidine metabolism
C
Identified transporter categories
YB420
CBS7504
nucleobase-containing compound
carbohydrate
carbohydrate derivative
ion
passive
active
organic acid
amide
inorganic molecular entity
Transmembrane transporter activity category
0
10
20
30
40
Number of proteins
D
Carbohydrate transporters only
YB420_S4
YB420_S3
CBS_S4
CBS_S3
YALI0C04730g
YALI0C06424g
YALI0D00363g
YALI0D01111g
YALI0F06776g
YALI0F25553g
-4
-2
0
2
4
6
8
10
Difference against S1 (t-test)
YALI0D00363p 0.40755
YALI0F25553p 0.21027
YALI0D01111p 0.1031
YALI0C04730p 0.10116
YALI0C06424p 0.34116
YALI0F06776p 0.3371

Figure 4- 5**.** (**A**) Schematic of growth characterization phenotypes for CBS7504 and YB420 when xylose was the sole remaining carbon source. Proteomic analysis of lipid metabolism showing increased (red) and decreased (blue) protein abundance when xylose was the sole remaining carbon source for CBS7504 (**B**) and YB420 (**C**). Proteins, circles; metabolites, regular font; pathways, bold font. Abbreviations: TAG (triacylglycerol); DHAP (dihydroxyacetone phosphate); Gly-3P (glycerol-3-phosphate); LPA (lysophosphatidic acid); PA, phosphatidic acid; DAG, diacylglycerol; PL, phospholipid; TAG, triacylglycerol; FFA, free fatty acid; TCA, tricarboxylic acid cycle; GPD1 (glycerol-3-phosphate dehydrogenase, YALI0B02948g); SCT1 (bifunctional glycerol-3-phosphate/glycerone-phosphate O-acyltransferase, YALI0C00209g); ALE1 (lysophosphatidate acyltransferase, YALI0F19514g); SLC1 (1-acyl-sn-glycerol-3-phosphate acyltransferase, YALI0E18964g); PAP (phosphatidic acid phosphohydrolase, YALI0D27016g); LPP1 (diacylglycerol diphosphate phosphatase/phosphatidate phosphatase , YALI0B14531g); DGA1 (acyl-CoA dependent diacylglycerol acyltransferase I, YALI0E32769g); DGA2 (acyl-CoA dependent diacylglycerol acyltransferase II, YALI0D07986g); TGL3 (triacylglycerol lipases 3, YALI0D17534g); TGL4 (triacylglycerol lipase 4, YALI0F10010g); YJU3 (monoglycerol lipase, YALI0C14520g); FAA1 (long-chain acyl-CoA synthetase, YALI0D17864g); FAS1 (fatty acid synthase 1,YALI0B15059g); FAS2 (fatty acid synthase 2; YALI0B19382g); ACC1 (acetyl-CoA carboxylase, YALI0C11407g); PDH1, PDH2; ACL1 (ATP-citrate lyase, YALI0E34793g); ME (malic enzyme, YALI0E18634g); MnDH1 (Mannitol dehydrogenase, YALI0B16192g); UGA2 (Succinate semialdehyde dehydrogenase, YALI0F26191g); SOL3 (6-Phosphogluconolactonase, YALI0E11671g); ZWF1 (NADP+-dependent glucose-6-phosphate dehydrogenase, YALI0E22649g); POT1 (3-oxyacyl-thiolase, YALI0E18568g); MFE1 (multifunctional β-oxidation protein, YALI0E15378g); PEX10 (peroxin-10, YALI0C01023g); POX1 (acyl-coenzyme A oxidase 1, YALI0E32835g); POX2 (acyl-coenzyme A oxidase 2, YALI0F10857g); POX3 (acyl-coenzyme A oxidase 3, YALI0D24750g); POX4 (acyl-coenzyme A oxidase 4, YALI0E27654g); POX5 (acyl-coenzyme A oxidase 5, YALI0C23859g); POX6 (acyl-coenzyme A oxidase 6, YALI0E06567g); YAS2 (HLH transcription factor, YALI0E32417g); POR1

(YALI0D12628g); SNF1 (YALI0D02101g); SAK1 (YALI0D08822g); REI1 (YALI0B08734g); HSF1 (YALI0E13948g); UPC2 (YALI0B15818g).


A
CBS7504
YB420
Xylose
Xylitol Production
Lipid Degradation
Biomass Degradation
Xylose
Xylitol Production
Lipid Accumulation
Biomass Maintenance
B
C
Xylose
Pentose Phosphate Pathway
NADPH Generation
MnDH1
UGA2
SOL3
ZWF1
DHAP
Glycolysis
GPD1
Pyruvate
ME
Malate
Gly-3P
PDH1
PDH2
ACL1
Citrate
TCA
SCT1
LPA
Acetyl-CoA
PEX10
ALE1
SLC1
ACC1
POT1
MFE1
B-Oxidation
POX1
POX2
POX3
POX4
POX5
POX6
PA
Malonyl-CoA
PAP
LPP1
FAS1
FAS2
DAG
Acyl-CoA
DGA1
DGA2
FAA1
FFA
YJU3
Glycerol
TAG
TGL3
TGL4
Glyceride
Lipid Regulators
YAS2
SAK1
HSF1
POR1
REI1
UPC2
SNF1
YALI0E30789g
TAG Lipase
YALI0B09361g
YALI0E17655g
YALI0A20350g
YALI0E08492g

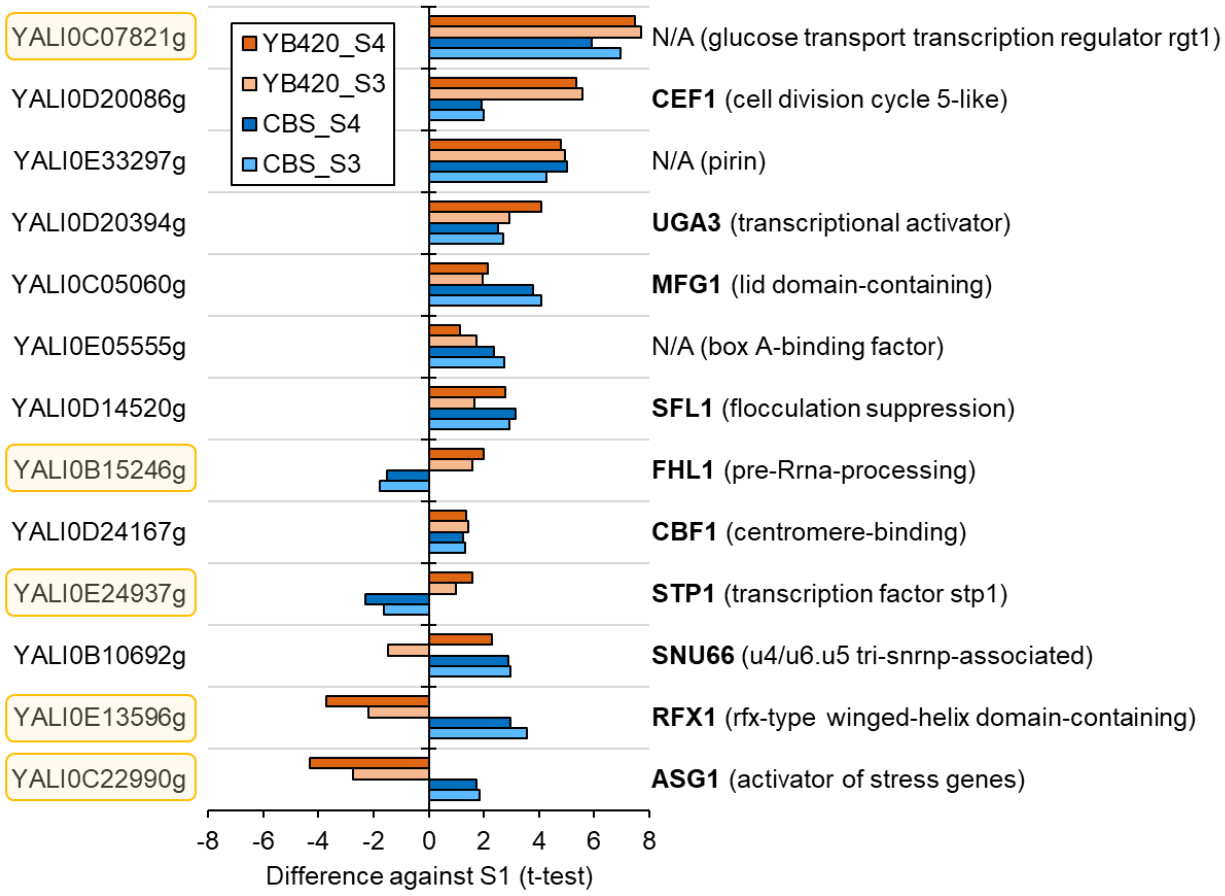


Figure 4- 6. Regulator proteins with significant protein abundance changes in the xylose assimilation phase. CBS7504, blue; YB420, orange.

Chapter V

Harnessing *Yarrowia lipolytica* for Biological Upcycling of Plastic Waste into Citric Acid and Lipids

Authors for this manuscript include:
Caleb Walker, Max Mortensen, Siris Laursen and Cong T Trinh.

CTT conceived and supervised the study. CW, MM, SL and CTT designed the experiments, and analyzed the data. CW and MM performed the experiments. CW and CTT wrote and approved the manuscript.

Abstract

Sustainable processes for recycling plastic waste are needed to reduce environmental pollution associated with increased plastic production, incineration, and landfill storage. Currently, most recycling strategies generate lower quality polymers that do not support an economic and sustainable return. Here, we present an alternative strategy using the yeast, *Yarrowia lipolytica*, for biological upcycling of catalytically depolymerized plastic (DP) waste into higher-value chemicals. While hydrophobic substrate (i.e., DP) metabolism is not fully understood, short-term adaptation of *Y. lipolytica* to the DP oil effectively increased utilization of the DP and sequential conversion into citric acid and triacylglycerol. Interestingly, adapted *Y. lipolytica* exhibited cellular adhesion to all hydrophobic substrates tested and significantly improved growth on the model alkane substrate hexadecane. Taken together, *Y. lipolytica* is a promising microbial host for upcycling plastic waste into higher-value chemicals and an emerging model organism to further elucidate mechanisms and genes underlying hydrophobic metabolism.

Keywords

Plastic waste; upcycling; *Yarrowia lipolytica*; hydrophobic substrate; catalytic depolymerization; alkane; hexadecane; organic acid production; lipid production; adaptive engineering; cell adhesion.

Introduction

Plastic waste pollutes all major ecosystems worldwide, negatively impacting wildlife and human health (1). Approximately 9% of all generated plastic waste was recycled, 12% incinerated, and 79% amassed in landfills as of 2015 (2). In 2050, an estimated 12,000 metric tons of plastic waste will occupy landfills if current trends in plastic production or plastic waste management are not altered (2). Yet, sustainable processes remain elusive

since current recycling concepts produce polymers of lower quality and cost (3). Hence, plastic waste must be converted into higher-value chemicals to render this process economically sustainable and cost competitive. Here, we demonstrate that biological upcycling of catalytically depolymerized plastic (DP) waste is an advantageous strategy due to the wide plethora of valuable chemicals microorganisms can produce.

One promising microbial host for this process is the oleaginous yeast *Yarrowia lipolytica* which can assimilate hydrophobic substrates (HS) including alkanes generated by catalytic depolarization of plastic (4). Additionally, *Y. lipolytica* efficiently produces valuable chemicals (e.g., organic acids and lipid-derived products), has an established genetic toolkit (5), is generally regarded as safe (GRAS certified) (6), and is extremely robust (e.g., broad ranging pH, high salinity and organic solvents) (7-9), which are all important characteristics for a bioproduction microbe.

Numerous studies demonstrating utilization of alkanes as a carbon source by *Y. lipolytica* date as far back as 1954, yet several aspects of this complex alkane metabolism are not fully understood (10). Specifically, i) how *Y. lipolytica* accesses the hydrophobic alkanes, ii) identification of a specific alkane transport system and iii) the carbon-number range of alkanes that support cell growth.

There are two proposed mechanisms enabling *Y. lipolytica* to access HS, secretion of an emulsifier named liposan (11) and direct adhering of the cells to the HS with a protrusion structure on the cell surface (12). However, identification of the genes encoding the proteins (i.e., genetic tractability) responsible for either of these mechanisms have not yet been discovered (13).

Further, identification of genes and mechanisms responsible for the transport of alkanes into *Y. lipolytica* cells remains unknown. However, disruption of two different genes altered growth phenotypes in a carbon-chain dependent manner. Specifically, disruption of an adenine nucleotide transporter protein ANT1 (providing ATP for acyl-coA synthase) diminished growth on short chain alkanes (mainly C10 and C11) but did not affect growth on C16 (14). Similarly, disruption of an ABC transporter ABC1 prohibited growth on long-chain alkanes (C14-C16), reduced growth on medium-chain alkanes (C12 and C13) but did not alter growth on short-chain alkanes (C10 and C11) (14).

Finally, the maximum carbon-number range of alkanes that support the growth of *Y. lipolytica* has not been explored. This is important because depending on the plastic feedstock and catalysis conditions, the depolymerized plastic will contain mixtures of alkanes and alkenes with a broad range of carbon chain lengths. While shorter alkanes (i.e., C1-C8 alkanes) do not support growth of *Y. lipolytica* cells (15), growth on longer chain alkanes (>C18) have not been reported to our knowledge.

Here, we investigate the biological upgrading of depolymerized plastic waste using *Y. lipolytica*. Specifically, we characterize the efficacy of *Y. lipolytica* to utilize the complex mixture of DP using short term adaptation, address the stability of the depolymerized plastic to identify potential limitations for biological upcycling, and demonstrate production of citric acid and lipids directly from the depolymerized plastic waste. We expect the results of this study to help guide catalytic selectivity goals, address potential stability issues with depolymerized plastic, and stimulate potential applications of biological upcycling plastic waste into higher-value chemicals.

Materials and Methods

Strains. Wildtype *Yarrowia lipolytica* Po1f was obtained from the American Type Culture Collection (ATCC MYA-2613) and was used as the parent strain for short-term adaptation to produce the adapted *Y. lipolytica* strain.

Medium and culturing conditions

Catalytic depolymerized plastic waste. Low-density polyethylene (LDPE) was used as the source of plastic waste. Batches of LDPE were catalytically depolymerized into DP oil and kindly provided by Dr. Laursen (UTK, CBE) and Max Mortensen after flushing vials containing the DP with nitrogen (Figure 5-1A). These vials of DP oil were stored at -20C until experimentation unless otherwise stated. The DP oil was composed of alkanes and alkenes ranging from 11 to 29 carbons (Figure 5-1B, 5-1C).

Media. Minimal defined media contained yeast nitrogen base without amino acids (Sigma #Y0626) and supplemented with 380 mg/L leucine and 76 mg/L uracil. Nitrogen limited

media contained yeast nitrogen base without amino acids and ammonium sulfate (Sigma #Y1251) and supplemented with 380 mg/L leucine, 76 mg/L uracil, 100 mM HEPES (Fisher #BP310), 90 mM dibasic phosphate, 10 mM monobasic phosphate, 0.73 g/L ammonium sulfate and adjusted to pH of 5 with 6 N hydrochloric acid. Glucose, DP, and purchased alkanes were used as carbon sources at concentrations mentioned throughout the text.

Culturing conditions. For each experiment, seed cultures were generated by inoculating a single colony from a fresh petri dish into a 14 mL culture tube in 3 mL of minimal media containing 20 g/L glucose and incubated overnight. Sub-seed cultures were generated by transferring 1.5 mL of the seed culture into 50 mL of minimal media containing 20 g/L glucose using a 500 mL baffled flask and incubated overnight until mid-exponential growth phase. For nitrogen limited experiments, sub-seed cultures were incubated for 2 days in nitrogen limited media containing 20 g/L glucose. Cells were centrifuged and washed once in sterile water prior to starting main experiments in 50 mL baffled flasks with 10 mL working volume and 3 technical replicates. Cells were incubated using a MaxQ6000 air incubator set to 250 rpms and 28 ℃.

Analytical methods

Lipid quantification. 1.2 mL of cell culture was transferred into a pre-weighted 1.5 mL centrifuge tube, washed twice with water, and resuspended to a total volume of 1.2 mL in water. 100 µL aliquots of each sample were transferred into a single well in duplicate using a 96-well plate. The remaining 1 mL of the cell culture sample was centrifuged at max speed for 5 minutes before removing the supernatant. The cell pellet was dried at 50 ℃ until a constant weight was obtained (e.g., 2-3 days) and used to calculate dry cell weight (DCW). A corn oil standard was prepared at 5 mg/mL l in ethanol and used to generate working standards ranging 0.05-1 mg/mL in water. 2 µL of 1 µg/mL Bodipy 435/503 (Fisher #D3922) in DMSO was added to each well and the plate was shaken in the dark for 15 minutes before measuring fluorescence (excitation 485nm; emission

528nm) to measure lipid concentrations. Lipid %DCW was calculated by dividing the average lipid concentration for 2 replicates/sample by the sample DCW.

Metabolites and sugar quantification. 1 mL cell culture samples were centrifuged at max speed for 2 minutes and the supernatant was transferred to a new centrifuge tube. Glucose and organic acids were measured after filtering supernatant samples with 0.2 µm filters using a Shimadzu high performance liquid chromatography system equipped with UV and RID detectors (Shimadzu Scientific Instruments, Inc. MD, USA) and an Aminex 87H column (Biorad, CA, USA) at a flow rate of 0.6 mL/min, oven temperature of 48 ℃ with 10 mM sulfuric acid mobile phase (16).

HS quantification by GCMS. 10 mL of sacrificial flask replicates were transferred into a 15 mL tube and stored at - 80 ℃. Samples were thawed before adding chloroform containing 800 mg/L ethyl pentadecanoate as the internal standard. Samples were incubated at 4 ℃ on a tube rotator overnight to ensure complete cell lysis and extraction. Samples were vortexed well and centrifuged at 3500 rpm for 5 minutes before transferring 250 µL of solvent into a GCMS vial through a 0.2 µM filter. Gas chromatography mass spectroscopy (GCMS) and nuclear magnetic resonance (NMR) were performed by Max Mortensen and advised Dr. Laursen (UTK, CBE).

Results

Growth and adaptation of Y. lipolytica on synthetic oil. Preliminary characterization with wildtype *Y. lipolytica* showed cell growth (i.e., turbidity) in culture tubes containing 5% (v/v) DP oil as the sole carbon source and in culture tubes containing a mixture of 5% (v/v) DP oil and 5 g/L glucose after 2 days of cultivation (Figure 5-2A). To confirm cell growth on DP oil, we repeated this experiment using baffled flasks to provide better aeration for cell growth. We found that 2% (v/v) DP oil as the sole carbon source supported growth of wildtype *Y. lipolytica* cells achieving 0.67 ± 0.09 OD_{600nm} in 4 days of cultivation (Figure 5-2B). However, cell growth was not detected using 5% or 10% (v/v) DP oil (data not shown). Notably, maximum cell mass (i.e., OD_{600nm}) in the mixture of

glucose and DP oil was reduced by ~53% in comparison to cell growth in glucose alone (Figure 2B). Attempting to improve cell growth on DP oil, we employed a short-course adaptation experiment by subjecting wildtype *Y. lipolytica* to 5 successive transfers in 2% (v/v) oil (Figure 5-2C). Short-term adaptation improved growth performance in DP oil, enabling the adapted strain (maximum OD600nm of 2.69 ± 0.55) to increase cell mass ~1.9-fold relative to the wildtype (maximum OD600nm of 1.43 ± 0.04) using 2% (v/v) DP oil as the sole carbon source (Figure 5-2D). Additionally, the adapted strain demonstrated superior growth on hexadecane (C16 alkane), which is a component of the plastic oil (Figure 5-2D). Of note, the adapted strain exhibited cellular adhesion to the hydrophobic layer of both DP oil and hexadecane even after centrifugation (Figure 5-2E).

Stability analysis of synthetic oil. During our experiments using plastic oil, we observed reduced cell growth from oil stored at room temperature (RT) for long periods of time (~2-3 months). To confirm this hypothesis, we combined fresh DP oil batches and partitioned half of the oil for storage at RT and half of the oil for storage at -20℃ (Figure 5-3A). Growth characterization of adapted *Y. lipolytica* at 3 and 6 weeks of storage showed no statistical differences in cell growth between the two storage conditions, albeit RT-stored oil resulted in greater deviation of growth between replicates (Figure 5-3B). After 9 weeks, the deviation in cell growth between replicates became prominent for RT-stored oil (Figure 5-3B). Finally, after 14 weeks of storage, the adapted strain was unable to grow in the RT-stored oil while growth was not affected from oil stored at -20℃ (Figure 3B). Qualitative physical observations of the RT-stored oil at 14 weeks included increased melting point and increased viscosity (Figure 5-3C). To understand the unstable components of the oil, we subjected oil from both storage conditions at 14 weeks to 2D NMR and GCMS which has not yet resulted in identification of differences in oil components.

Lipid and citric acid production from synthetic oil. Next, we aimed to upcycle plastic waste into higher-value chemicals, specifically triacylglycerol lipids and citric acid, using nitrogen limited media. Using 2% (v/v) DP oil, adapted *Y. lipolytica* achieved a maximum

cell growth of 3.34 ± 0.32 OD_{600nm} while producing 2.33 ± 0.12 g/L citric acid and accumulating 10.09 ± 0.42 lipid %DCW (Figure 5-4A-C). In 2% (v/v) hexadecane, maximum cell growth (11.74 ± 0.36 OD600nm) and citric acid production (14.6 ± 0.77 g/L citric acid) were significantly higher in comparison to the DP oil (Figure 5-4A, 5-4B). However, both hexadecane and DP oil carbon sources exhibited similar lipid levels by day 4 (Figure 5-4C). Cells grown in 20 g/L glucose showed the lowest lipid level (6.93 ± 1.17 lipid %DCW) with moderate citric acid titer (4.28 ± 0.37 g/L citric acid) and intermediate cell growth (7.34 ± 0.14 OD600nm) between the 3 carbon sources. Interestingly, both hexadecane and DP oil GCMS profiles showed complete consumption despite the differences in cell growth and citric acid production (Figure 5-4D).

Discussion

Increased demand for- and production of- plastic has led to a significant increase in waste. To avoid the negative environmental impacts associated with plastic waste necessitates economic strategies to recycle plastic into useful chemicals. Here, we aimed to upcycle plastic waste by harnessing *Y. lipolytica* for the bioconversion of catalytically DP. By maintaining the DP in frozen storage conditions and adapting *Y. lipolytica* to the DP, we demonstrated production of citric acid and triacylglycerol from plastic waste as the sole carbon source.

Although well-studied for utilization of HS, wildtype *Y. lipolytica* exhibited poor growth on the DP oil (Figure 5-2B, 5-2D). Further, cell growth in the mixture of glucose and DP oil was reduced in comparison to cell growth in glucose only, indicating that the plastic oil contains toxic components that hinder cell growth (Figure 5-2B). Short term adaptation, however, not only improved cell growth on DP oil but also on hexadecane (Figure 5-2D), suggesting alkane metabolism was cryptic and required activation as previously observed in xylose metabolism (17). Yet, the question arises about the drastic differences in cell growth between pure hexadecane and the mixture of alkanes contained in the DP. Is reduced growth because of inhibitory components in the DP oil? Or does *Y. lipolytica* lack the metabolic capacity required to support cell growth on the longer

carbon-chain alkanes contained in the DP? To answer these questions, we are currently testing growth of *Y. lipolytica* on alkanes and alkenes ranging from C11-C28 and performing detailed NMR on the DP to identify potential inhibitory components.

During our experiments using DP, we observed inhibition of cell growth (even for the adapted strain) from oil stored at room temperature for long periods of time (~3 months), indicating unstable components were contained within the DP (Figure 5-3B). Fortunately, the oil remained stable if stored frozen immediately after catalytic depolymerization, but the question arises concerning what makes the DP unstable. Specifically, what chemical(s) arise that inhibits growth of *Y. lipolytica* and can the catalytic depolymerization processing be optimized to prevent or minimize this unstably? Due to the complexity of the DP mixture, we were only able to detect small shifts in NMR spectra between inhibitory- and fresh-DP oil that have not provided any clear hypotheses to date. In parallel, we are analyzing the inhibitory- and fresh-DP oil using GCMS to help resolve compositional changes in the unstable DP oil.

Albeit with room for improvement, *Y. lipolytica* showed promise as the biological host for upcycling plastic waste into higher-value chemicals by producing citric acid and neutral lipids directly from DP (Figure 5-4A-C). Cells grown with hexadecane as the sole carbon source consumed all hexadecane, produced ~6-fold more citric acid, and achieved ~3.5 greater cell mass in comparison to cells grown on DP oil. However, GCMS analysis revealed most of the DP components (i.e., <C26 alkanes) were consumed where the question arises about the carbon flow (Figure 5-4D). In other words, DP-grown cells assimilated most of the DP yet did not achieve similar cell growth or citric acid titers in comparison to hexadecane-grown cells. We hypothesize that assimilated DP is either converted and secreted as an extracellular emulsifier (e.g., liposan) (18) or functionalized into an intermediate metabolite of alkane degradation (i.e., aldehyde, alcohol or fatty acid). To date, GCMS has not identified any alcohol, aldehyde or fatty acid intermediate metabolites. However, fatty acids are not readily detectable by GCMS without first converting them into esters. Hence, we are performing transesterification on extracted fatty acids to conclude whether fatty acids are accumulating in DP-grown cells. In parallel,

we are analyzing emulsification activity of these cultures to determine if DP is being converted into the compositionally controversial emulsifier (13).

Additionally, for all HS tested, the adapted strain grew as a mixture of adhering cells (i.e., located in the HS top layer) and planktonic cells (i.e., located in the aqueous phase) (Figure 5-2E) which are understood to emerge once the hydrophobic layer has become saturated with adhering cells (13). While the mechanism(s) underlying HS accessibility and transport are not fully understood, we observed evidence via microscopic images of both surface-mediated (i.e., extracellular emulsifier acting as surfactant, Figure 5-4E) and direct interfacial (i.e., cell attachment to HS droplets, Figure 5-2E) transport mechanisms enabling cells to access and assimilate the HS (19).

In conclusion, our work demonstrates a promising platform for upcycling plastic waste by catalytic depolymerization and subsequent bioconversion of the DP into higher-value chemicals with the bioenergy microbe, *Y. lipolytica*. These results suggest further adaptation engineering of *Y. lipolytica* could improve cell growth and chemical production from DP. Further work concerning mechanisms and genes governing HS accessibility, uptake, and regulation of HS metabolism is critical to unlock the full potential of *Y. lipolytica* for upcycling plastic waste (20). To address this challenge, we have conducted proteomic samples of *Y. lipolytica* (planktonic and adhering) cells cultured in glucose, hexadecane, or DP oil to elucidate proteins required for- and mechanisms underlying- HS metabolism.

References

1. K. L. Law *et al.*, The United States' contribution of plastic waste to land and ocean. *Science Advances* **6**, eabd0288 (2020).
2. R. Geyer, J. R. Jambeck, K. L. Law, Production, use, and fate of all plastics ever made. *Science Advances* **3**, e1700782 (2017).
3. K. Beydoun, J. Klankermayer, Efficient Plastic Waste Recycling to Value-Added Products by Integrated Biomass Processing. *ChemSusChem* **13**, 488-492 (2020).
4. P. Fickers *et al.*, Hydrophobic substrate utilisation by the yeast Yarrowia lipolytica, and its potential applications. *FEMS Yeast Research* **5**, 527-543 (2005).
5. K. A. Markham, H. S. Alper, Synthetic Biology Expands the Industrial Potential of Yarrowia lipolytica. *Trends in Biotechnology* **36**, 1085-1095 (2018).
6. M. Groenewald *et al.*, Yarrowia lipolytica: Safety assessment of an oleaginous yeast with a great industrial potential. *Critical Reviews in Microbiology* **40**, 187-206 (2014).
7. C. Walker, S. Ryu, C. T. Trinh, Exceptional solvent tolerance in Yarrowia lipolytica is enhanced by sterols. *Metabolic Engineering* **54**, 83-95 (2019).
8. S. Ryu, N. Labbé, C. T. Trinh, Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α-ketoglutaric acid by Yarrowia lipolytica. *Applied Microbiology and Biotechnology* **99**, 4237-4244 (2015).
9. P. Xu, K. Qiao, G. Stephanopoulos, Engineering oxidative stress defense pathways to build a robust lipid production platform in Yarrowia lipolytica. *Biotechnology and Bioengineering* **114**, 1521-1530 (2017).
10. G. Barth *et al.*, "Functional genetics of Yarrowia lipolytica" in Functional Genetics of Industrial Yeasts*,* J. H. de Winde, Ed. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003), 10.1007/3-540-37003-X_8, pp. 227-271.
11. S. S. Zinjarde, A. Pant, Emulsifier from a tropical marine yeast, Yarrowia lipolytica NCIM 3589. *Journal of Basic Microbiology* **42**, 67-73 (2002).

12. K. Mlíčková *et al.*, Lipid Accumulation, Lipid Body Formation, and Acyl Coenzyme A Oxidases of the Yeast <em>Yarrowia lipolytica</em&gt. *Applied and Environmental Microbiology* **70**, 3918 (2004).
13. F. Thevenieau *et al.*, "Uptake and assimilation of hydrophobic substrates by the oleaginous yeast Yarrowia lipolytica" in Handbook of hydrocarbon and lipid microbiology. (2010).
14. F. Thevenieau *et al.*, Characterization of Yarrowia lipolytica mutants affected in hydrophobic substrate utilization. *Fungal Genetics and Biology* **44**, 531-542 (2007).
15. M. J. Klug, A. J. Markovetz, Degradation of Hydrocarbons by Members of the Genus <em>Candida</em&gt. *Applied Microbiology* **15**, 690 (1967).
16. S. Ryu, C. T. Trinh, Understanding Functional Roles of Native Pentose-Specific Transporters for Activating Dormant Pentose Metabolism in <span class="named-content genus-species" id="named-content-1">Yarrowia lipolytica</span&gt. *Applied and Environmental Microbiology* **84**, e02146-02117 (2018).
17. S. Ryu, J. Hipp, C. T. Trinh, Activating and Elucidating Metabolism of Complex Sugars in <span class="named-content genus-species" id="named-content-1">Yarrowia lipolytica</span&gt. *Applied and Environmental Microbiology* **82**, 1334 (2016).
18. M. H. Vance-Harrop, N. B. d. Gusmão, G. M. d. Campos-Takaki, New bioemulsifiers produced by Candida lipolytica using D-glucose and babassu oil as carbon sources. *Brazilian journal of microbiology* **34**, 120-123 (2003).
19. T.-H. Kim, Y.-S. Oh, S.-J. Kim, The possible involvement of the cell surface in aliphatic hydrocarbon utilization by an oil-degrading yeast, Yarrowia lipolytica 180. *Journal of microbiology and biotechnology* **10**, 333-337 (2000).
20. R. Fukuda, Metabolism of Hydrophobic Carbon Sources and Regulation of It in n-Alkane-Assimilating Yeast Yarrowia lipolytica. *Bioscience, Biotechnology, and Biochemistry* **77**, 1149-1154 (2013).

Appendix

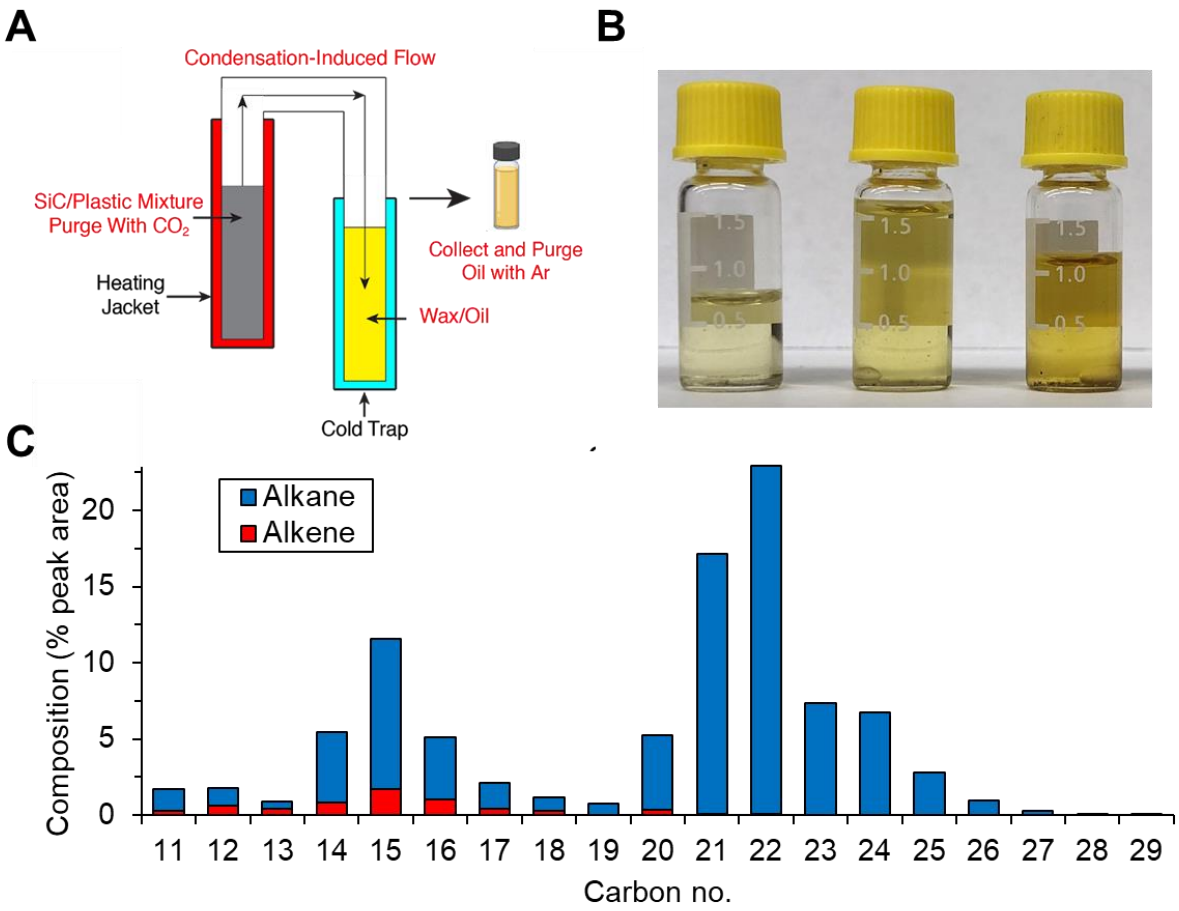
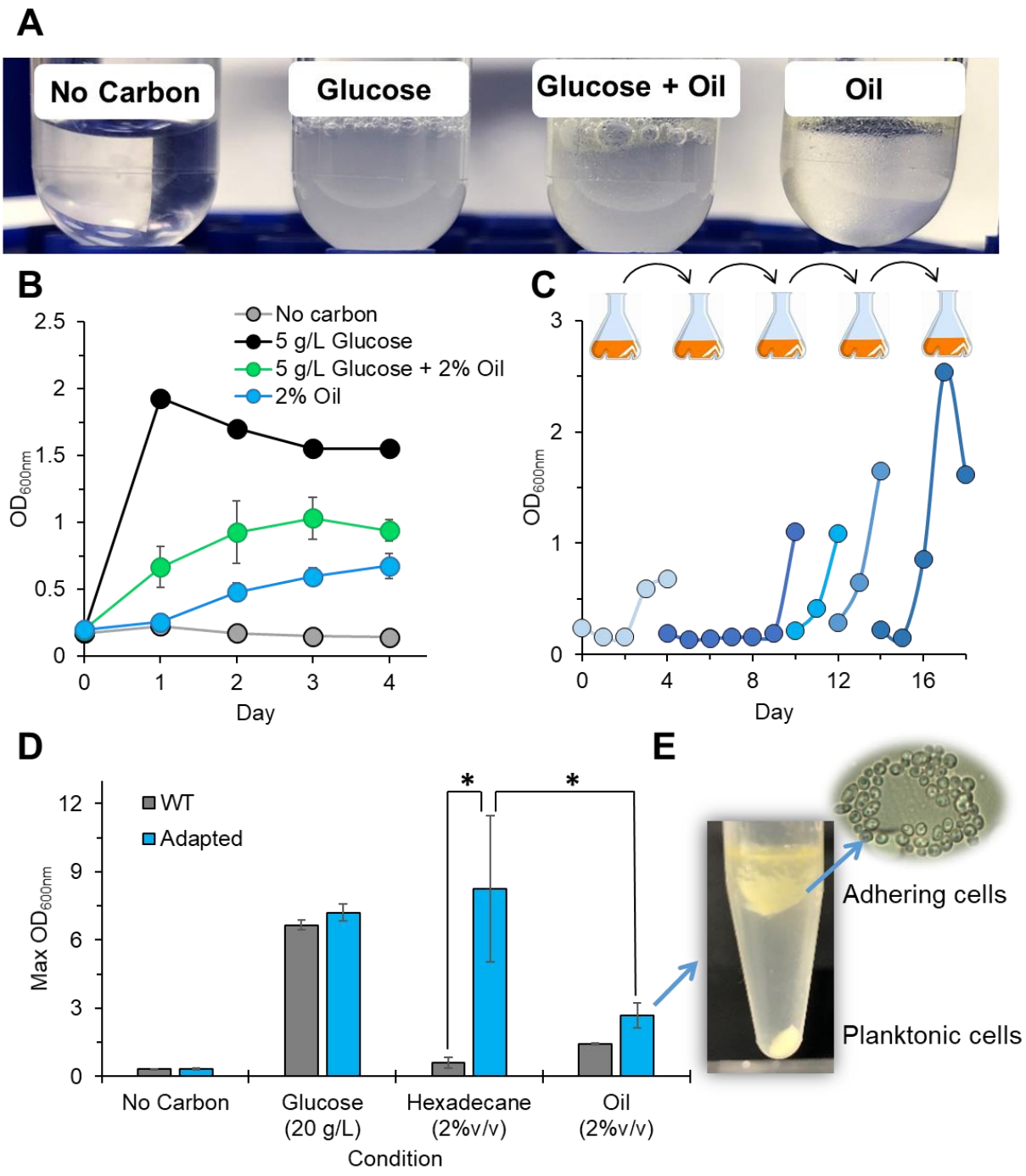


Figure 5- 1. (**A**) Condensation flow reactor scheme to generate DP. (**B**) Images and (**C**) composition of DP oil.

Figure 5- 2. Growth of wildtype *Y. lipolytica* in (**A**) culture tubes and (**B**) baffled flasks. (**C**) Short-term adaptation of *Y. lipolytica* in DP oil to generate adapted strain. (**D**) Growth comparison of wildtype and adapted *Y. lipolytica* in glucose, hexadecane and DP oil. (**E**) Image of adapted *Y. lipolytica* cultured in DP oil after centrifugation and microscopic image of adhering cells. *p-value < 0.001 using one way analysis of variance Holm-Sidak method.


A
No Carbon
Glucose
Glucose + Oil
Oil
B
No carbon
5 g/L Glucose
5 g/L Glucose + 2% Oil
2% Oil
OD600nm
Day
C
OD600nm
Day
D
WT
Adapted
Max OD600nm
No Carbon
Glucose (20 g/L)
Hexadecane (2%v/v)
Oil (2%v/v)
Condition
E
Adhering cells
Planktonic cells

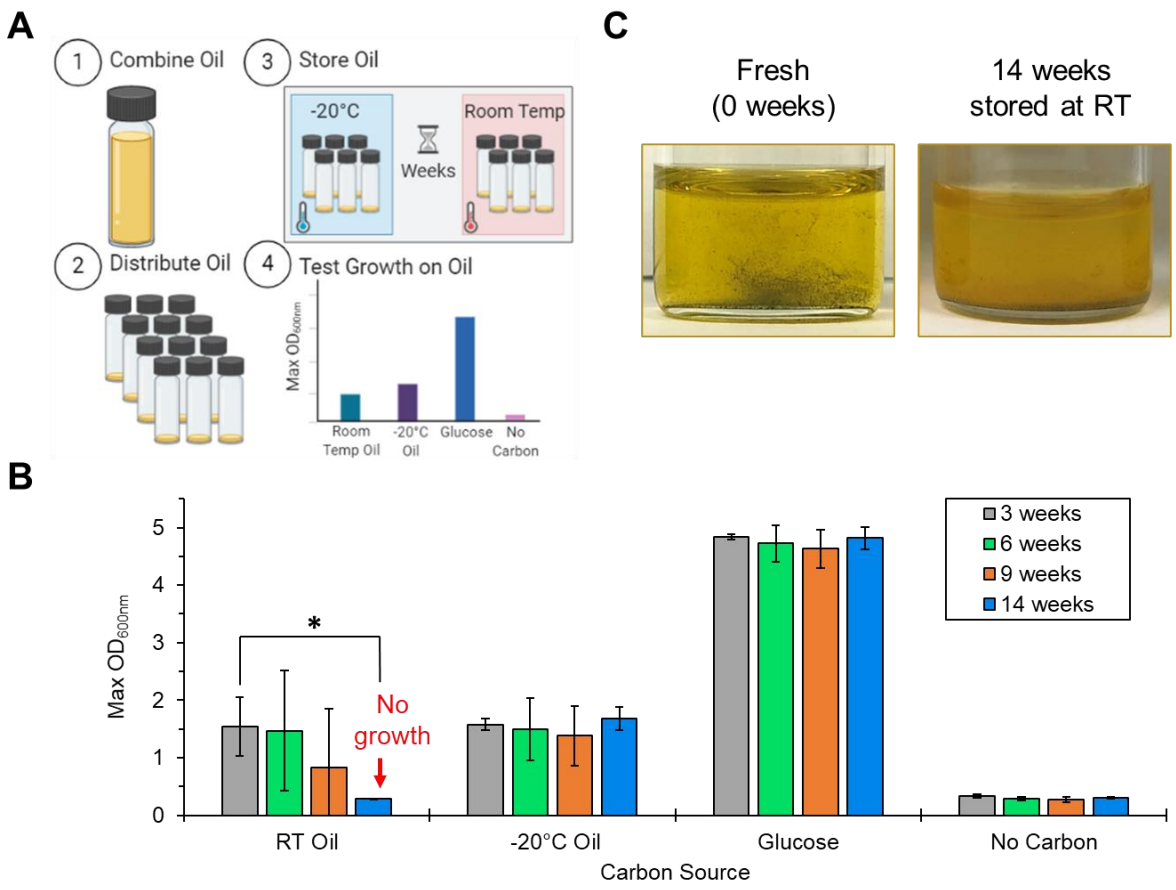


Figure 5- 3. (**A**) Workflow to generate DP oil stored at RT or -20°C for stability characterization. (**B**) Growth of adapted *Y. lipolytica* on DP oil stored at RT or -20°C after 3, 6, 9 and 14 weeks. (**C**) Images of DP oil stored in RT at week 0 and week 14. *p-value < 0.001 using one way analysis of variance Holm-Sidak method.

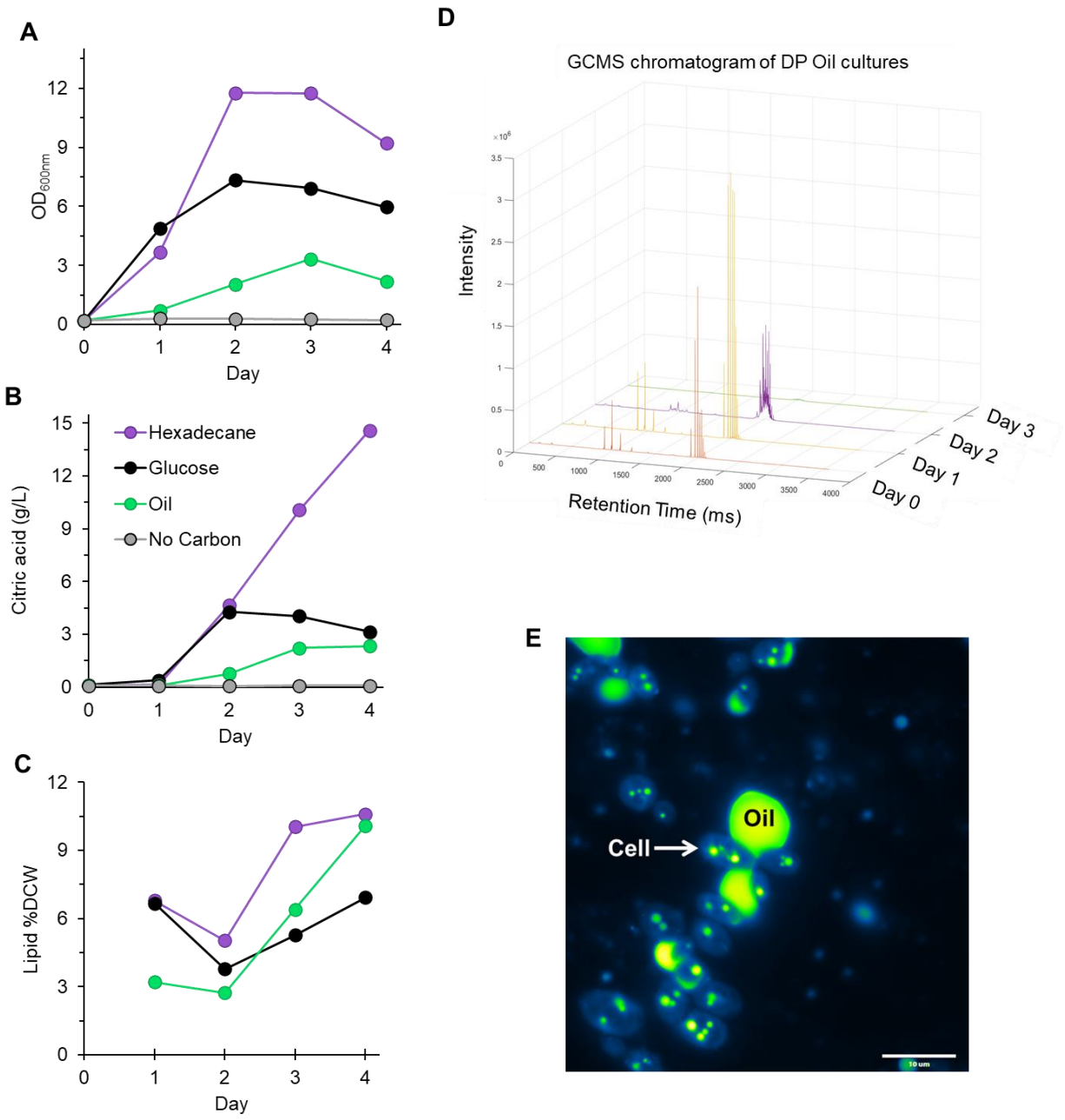


Figure 5- 4. Production of citric acid and lipids from DP oil using adapted *Y. lipolytica*. (**A**) Growth, (**B**) citric acid production, (**C**) lipid accumulation, (**D**) and DP oil consumption profiles. (**E**) Fluorescent image of adapted *Y. lipolytica* cultured in DP oil and stained with Bodipy (as described in lipid quantification method).

CONCLUSION AND FUTURE DIRECTION

Knowledge of genotypes and mechanisms underlying robustness are essential for reverse engineering these traits, but also for understanding processes affected by- or attributing to- these complex phenotypes. The goal of this dissertation was to elucidate genotypes and cellular processes responsible for robust phenotypes displayed by *Y. lipolytica* by ascertaining differences between robust and non-robust strains using multi-OMICS strategies.

In the case of an absent phenotype, thiamine auxotrophy, comparative genomics with thiamine prototrophic yeast *S. cerevisiae* and knowledge of the thiamine biosynthesis pathway proved sufficient for identification of the missing gene causing *Y. lipolytica*'s thiamine auxotroph. However, restoring thiamine biosynthesis with this gene required a thiamine-regulated promoter to efficiently express this gene, highlighting the complexity of gene regulation underlying seemingly simple phenotypes.

In the case of divergent phenotypes by members of the *Y. lipolytica* clade, comparative proteomics revealed global changes and allowed for speculation of proteins governing xylose metabolism and lipid degradation or accumulation. We found singleton or exclusive genes that might be responsible for the diverging lipid and xylose metabolic phenotypes, but unfortunately our proteomics analysis did not identify these novel proteins. This study also highlights the importance of further investigation of gene regulation (e.g., cis and trans regulatory elements) in controlling cellular phenotypes.

In the case of native phenotypes (i.e., the phenotype already exists), we found adaptive laboratory engineering particularly useful which produced platform strains by which to study superior mechanisms of robustness. Specifically, short-term adaptation on depolymerized plastic oil generated a strain with dramatically improved utilization of hydrophobic substrates (HS) but also with pronounced characteristics hypothesized to enable cellular accessibility of the HS. With this strain, we expect to identify genes responsible for cellular adhesion to the hydrophobic layer and perhaps other mechanisms empowering HS metabolism. Multi-OMICS analysis will be critical for addressing these fundamental knowledge gaps.

Regarding solvent tolerance, long-term adaption in ionic liquid (IL) generated a superior solvent tolerant strain that exhibited remarkable and broad tolerance to all ILs tested. Extensive characterization of this strain led us to multiple mechanistic hypotheses of solvent tolerance such as lipid remodeling but relied on previous findings to establish the mechanism (i.e., sterols strengthening membrane) attributing to our success in identifying a genotype (i.e., overexpression of the sterol transcription factor) for reverse engineering solvent-tolerance. In this project, the system studied was the whole cell and numerous characterization experiments were necessary to identify sterols were affected by ILs and critical for IL-tolerance. However, what if we could produce genes conferring IL-tolerance from a single transcriptomics experiment? If so, this would minimize characterization efforts by significantly reducing the size of the system from the whole cell to handful of genes that would allow for rapid determination of genotypes and mechanisms underlying robust phenotypes. However, our temporal transcriptomics experiment resulted in a large list of candidate genes (i.e., 202 genes) leading to another scientific question - which genes should we chose and how should we choose them?

Traditional approaches advocate selection of the most differentially expressed gene(s), but is the most abundant gene transcript truly the best predictor of genetic targets? And with this question an idea was formulated: Since genes that are co-expressed can infer a relationship (e.g., controlled by the same regulation machinery, involved in the same metabolic pathway, functionally related etc.), could we predict the genotype underlying a phenotype by choosing genes with the highest connectivity in a co-expression network of upregulated genes during the phenotype? In other words, we hypothesized that the most connected genes (of genes upregulated by the robust mutant) would have the greatest influence on conferring to the robust phenotype. Though limited to 29 gene candidates, our results discovered gene connectivity as a promising metric to predict genetic targets conferring solvent tolerance in *Y. lipolytica.* To better understand the implications of our results, future studies could address other phenotypes, different strains, and test more genetic targets with alternative metrics of gene selection.

Overall, this work extensively contributes to experimental strategies and analyses aiming to elucidate processes and genotypes underlying robustness for reverse

engineering either absent (thiamine auxotroph), genetically diverse (lipid and xylose metabolism), or native phenotypes (HS metabolism and IL tolerance). Understanding and harnessing gene functions and regulation are critical to control metabolic processes for metabolic engineering and synthetic biology applications.

VITA

Caleb Walker was born to Jeff and Jane Walker on April 4, 1991 in Memphis, Tennessee. Caleb graduated from Sullivan South High School in 2009. He attended the University of Tennessee Chattanooga, earning a Bachelor of Science degree in Chemical Engineering and a minor in Chemistry in December 2013. He was employed as a process and development engineer at Steward Advanced Materials before continuing studies at the University of Tennessee, Knoxville. He plans to continue his career in the field of metabolic engineering and synthetic biology by pursuit of an industry position.