bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Reconstructing the evolution of metabolism in budding yeasts

Kevin Correia[1], Shi M. Yu[1], and Radhakrishnan Mahadevan[1,2,*]

[1]**Department of Chemical Engineering and Applied Chemistry, University of Toronto, Canada, ON**

[2]**Institute of Biomaterials and Biomedical Engineering, University of Toronto, Ontario, Canada**

Corresponding author:

Radhakrishnan Mahadevan*

Email address: krishna.mahadevan@utoronto.ca

ABSTRACT

Budding yeasts are a diverse group of microbes that inhabit a range of environments by exploiting various metabolic traits. The complete genetic basis of these traits are mostly unknown, preventing their addition or removal in a chassis organism for fermentation, including the removal of aerobic ethanol accumulation in *Saccharomyces cerevisiae* for biomass-derived products. To help understand the molecular evolution of aerobic fermentation in yeasts, we created Analyzing Yeasts by Reconstructing Ancestry of Homologs (AYbRAH), an open source database of manually curated and predicted orthologs in 33 Dikarya fungi (https://github.com/kcorreia/aybrah). AYbRAH was used to evaluate the gain and loss of orthologs in budding yeast central metabolism and refine the topology of recently published yeast species trees. The shift from a lipogenic-citrogenic to ethanologenic lifestyle is distinguished by the emergence of the pyruvate dehydrogenase (PDH) bypass in budding yeasts, with a glucose-inducible acetyl-CoA synthetase. Flux balance analysis (FBA) indicates the PDH bypass has a higher protein and phospholipid yield from glucose than ATP citrate lyase, the ancestral acetyl-CoA source in budding yeasts. Complex I subunits and internal alternative NADH dehydrogenase (NDI) orthologs have been independently lost in multiple yeast lineages. FBA simulations demonstrate Complex I has a significantly higher biomass yield than NDI. We propose the repeated loss of Complex I in yeasts is a result of Complex I inhibition by a mitochondrial metabolite. Existing and future genome annotations can be improved by annotating genomes against manually curated ortholog databases, which we believe can improve the understanding of physiology and evolution across the tree of life.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 INTRODUCTION

Yeasts are unicellular fungi that exploit diverse habitats on every continent, including the gut of wood boring beetles, insect frass, tree exudate, rotting wood, rotting cactus tissue, soil, brine solutions, and fermenting juice (Kurtzman et al., 2011). The most widely studied yeasts are true budding yeasts, which span roughly 400 million years of evolution within the subphylum Saccharomycotina (Hedges et al., 2015), and possess a broad range of traits important to metabolic engineering. These include citrate and lipid accumulation in *Yarrowia* sp. (Aiba and Matsuoka, 1979) and *Lipomyces* sp. (Boulton and Ratledge, 1983), thermotolerance in multiple lineages (Ryabova et al., 2003; Banat et al., 1996), acid tolerance in *Pichia* sp. (Rush and Fosmer, 2013), methanol utilization in *Komagataella* sp. (Hazeu and Donker, 1983), osmotolerance in *Debaromyces* sp. (Larsson and Gustafsson, 1993), xylose fermentation to ethanol in multiple yeast lineages (Schneider et al., 1981; Slininger et al., 1982; Toivola et al., 1984), alternative nuclear codon reassignments (Mühlhausen et al., 2016), glucose and acetic acid co-consumption in *Zygosaccharomyces* sp. (Sousa et al., 1998), and aerobic ethanol production (the Crabtree effect) in multiple lineages (De Deken, 1966; Van Urk et al., 1990; Blank et al., 2005; Christen and Sauer, 2011). Most of these traits do not have a complete known genetic basis, preventing their addition or removal in a chassis organism to improve fermentation techno-economics.

The Crabtree effect is one notable hindrance associated with *Saccharomyces cerevisiae* for biomass-derived products, such as heterologous proteins (Hensing et al., 1995). The volumetric productivity of these fermentations could be improved if aerobic ethanol production were eliminated in a two-stage fermentation process (Venayak et al., 2015), yet a complete understanding of the molecular mechanisms underlying the Crabtree effect remains unknown. At a fundamental level, the presence or absence of the Crabtree effect can be explained by a gain of function, a loss of function, or some combination of gains and losses of function in an ancestral yeast species without the Crabtree effect. However, most studies take a *S. cerevisiae*-centric view (Vemuri et al., 2007; Nilsson and Nielsen, 2016; Huberts et al., 2012; Vazquez and Oltvai, 2016), or do not consider molecular evolution of multiple species (Blank et al., 2005; De Deken, 1966; Christen and Sauer, 2011).

If we could study the physiology of the mother of all budding yeasts, which we refer to as the Proto-Yeast, along with a Crabtree-negative and Crabtree-positive yeast, we could undertake model-based studies to reverse engineer what gains and losses of function have occurred in each lineage. The Proto-Yeast has evolved from its original state, making this direct study impossible, but we can study the metabolism of Proto-Yeast's living descendants. In recent years, dozens of diverse yeast genomes have been sequenced (Riley et al., 2016), paving the way for greater insight into the evolution of metabolism in the yeast pan-genome.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genome-scale metabolic reconstructions provide a scaffold to study metabolism in yeasts (Österlund et al., 2013), but many of these reconstructions are understandably skewed to reflect the well-studied metabolism of *S. cerevisiae*. The two types of errors in these metabolic reconstructions are commission and omission. With commission, enzymes present in the metabolism of *S. cerevisiae* are often included in a non-conventional organism's metabolic reconstruction, despite a lack of genomic evidence. Omission occurs when enzymes are not included in a non-conventional organism's metabolic reconstruction, likely since they are absent in *S. cerevisiae*. Specific examples of both types of errors in various yeast reconstructions can be found in Supplemental Information. These errors can significantly alter flux analysis and impair our understanding of these yeasts' physiologies. The root of this problem is a lack of enzyme characterization in non-model organisms and a lack of high-quality orthology assignments. Every enzyme in every organism is unlikely to be characterized, hence, we argue that a transparent and open curated ortholog database is critical for creating higher quality genome-scale reconstructions.

The distinction between orthologs, paralogs, ohnologs and xenologs is an important aspect of comparative genomics and its role in the evolution of physiology. Briefly, orthologs arise from speciation and often have conserved function; paralogs and ohnologs emerge from duplication events and may have novel function; xenologs derive from horizontal gene transfer between organisms (Jensen, 2001). Commonly used ortholog databases, such as eggNOG (Huerta-Cepas et al., 2016b), InParanoid (Sonnhammer and Östlund, 2015), KEGG orthology (Mao et al., 2005), PIRSF (Wu et al., 2004), Ensembl (Vilella et al., 2009), OrthoMCL (Fischer et al., 2011), OrthoDB (Kriventseva et al., 2015), and OMA (Altenhoff et al., 2015), often fail to separate homologs into true orthologs in gene families with many duplications (see Supplemental Information). Higher quality ortholog databases include PANTHER and Yeast Gene Order Browser, but these are limited to a smaller selection of organisms (Mi et al., 2012; Byrne and Wolfe, 2005). In addition to these databases, ad hoc ortholog relationships are often assigned in genomics studies or genome-scale metabolic reconstructions, but they lack transparency and are not continuously improved. To solve these problems, and ultimately improve our understanding of yeast physiology and its evolution, we present Analyzing Yeasts by Reconstructing Ancestry of Homologs (AYbRAH). AYbRAH, derived from the Hebrew name Abra, mother of many, is an open source collection of predicted and manually curated orthologs, their function, and origin. Using this database, we reviewed the gains and losses of orthologs in central metabolism within Dikarya, refined the topology of published yeast species trees, and focus on how the evolution of the pyruvate dehydrogenase (PDH) bypass and NADH dehydrogenase may have played a role in the Crabtree effect. These studies, combined with modeling, provide an improved understanding of how the evolution of acetyl-CoA metabolism and the electron transport chain (ETC) impacted yeast physiology.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2 METHODS

AYbRAH curation. AYbRAH was created by combining several databases and programs in a pipeline to identify ortholog groups (Figure 1). Protein sequences from 33 organisms in Dikarya were downloaded from UniProt and MycoCosm (Table 1) (Grigoriev et al., 2013). 212 836 proteins were fed into OrthoMCL (Fischer et al., 2011) to cluster genes into putative Fungal Ortholog Group's (FOG's) with default parameters in BLASTP and OrthoMCL. FOG's were coalesced into Homolog Groups (HOG's) using Fungi-level OrthologDB identification codes (Kriventseva et al., 2015). MAFFT 7.245 (Katoh and Standley, 2013) was used to align the HOG sequences using a gap and extension penalty of 1.5. 100 bootstrap trees were constructed for each HOG with PhyML 3.2.0 (Guindon et al., 2010), optimized for tree topology and branch length. Consensus phylogenetic trees were generated with SumTrees in DendroPy 4.1.0 (Sukumaran and Holder, 2010), and figures drawn with ETE3 (Huerta-Cepas et al., 2016a). Phylogenetic trees were reviewed in central metabolism when OrthoMCL failed to create monophyletic ortholog groups caused by underclustering and overclustering (see Supplemental Information). New ortholog groups were created from polyphyletic FOG's when there was evidence of a monophyletic ortholog group either by manual inspection of its phylogenetic tree or with the help of an ETE3 script. Evolview (He et al., 2016) was used to map ortholog groups to the yeast species trees. Additional steps were required to assign proteins to a FOG due to a lack of assignment by OrthoMCL or an incomplete protein annotation. A protein from each FOG of the species closest relative was queried against its genome sequence with TBLASTN, using an expect threshold less than 1e-20. Annotated proteins were then queried against the TBLASTN hits to determine which proteins were annotated but not assigned to a FOG by OrthoMCL, and which proteins were unannotated despite a match in its genome. Both groups of proteins were assigned to a HOG by its best match, and to a FOG with pplacer (Matsen et al., 2010) via the MAFFT add alignment option.

Metabolic simulations Flux balance analysis (FBA) was carried out with the consensus *S. cerevisiae* genome-scale metabolic model (Heavner and Price, 2015) in MATLAB to quantify yield differences with acetyl-CoA pathways and matrix-orientated NADH dehydrogenase. Carbohydrate, protein, RNA, DNA, and phospholipid yields with glucose were computed from three acetyl-CoA pathways: the *S. cerevisiae* wild-type PDH bypass via pyruvate decarboxylase (PDC), NADP-acetaldehyde dehydrogenase (ALD), and acetyl-CoA synthetase (ACS); the PDH bypass via NAD-ALD; and ATP citrate lyase (ACL). Biomass yield was computed with Complex I and internal NADH dehydrogenase (NDI).

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3 RESULTS & DISCUSSION

Impact of acetyl-CoA source on biomass precursor yields. We maximized the production of carbohydrates, protein, DNA, RNA, and phospholipids with glucose using a *S. cerevisiae* genome-scale model to help understand the impact of acetyl-CoA sources on their yields. No changes were observed with carbohydrate, RNA, and DNA yields for the PDH bypass with NADP-ALD, the PDH bypass with NAD-ALD, and ACL. Protein and phospholipid yields were highest with the PDH bypass, at 0.53 g/g and 0.02 g/g, respectively. The PDH bypass via NAD-ALD marginally reduced both yields by 0.01%. ACL reduced the protein and phospholipid yields by 2.9% and 0.07%, respectively (Figure 4). Hence, there are clear differences between the different acetyl-CoA synthesis pathways, suggesting a link between the different pathways, and their impact on the physiology of the yeasts.

Evolution of PDH bypass The PDH bypass is a well-known pathway in *S. cerevisiae* (Pronk et al., 1996; van Rossum et al., 2016), but its phylogenetic origin has never been explored or why the ancestral ACL was lost in most budding yeasts. The primary source of cytosolic acetyl-CoA in Proto-Yeast was likely ACL (Hynes and Murray, 2010); it exists as a homo-oligomer in Basidiomycota (Shashi et al., 1990), and a hetero-oligomer in Ascomycota following an ancient gene duplication (Figure 3). ACL was lost in the lineage sister to *Nadsonia fulvescens* following the acquisition of ACS2 (Figure 3). Another source of acetyl-CoA in Proto-Yeast was the phosphoketolase (PHK) pathway , which was lost in budding yeasts sister to *Lipomyces starkeyi*. PHK also exists as a heterooligomer in the Pezizomycotina-Saccharomycotina clade. Proto-Yeast likely had one copy of ACS, ACS1, an enzyme expressed when grown on acetate (Connerton et al., 1990). A duplication in an ancestral species sister to *Yarrowia lipolytica*, which we refer to as the Proto-Fermenter, led to ACS2 (Figure 4A), the glucose-inducible ACS present in the PDH bypass (van den Berg et al., 1996; Zeeman and Steensma, 2003). ACS3, a propionyl-CoA synthetase, derived from an additional duplication of ACS1 in Pezizomycotina; it may have been present in Proto-Yeast. *Schizosaccharomyces pombe* (Van Urk et al., 1990), *Blastobotrys adeninivorans*, and *N. fulvescens* are the only yeasts in this study with ACL but use the PDH bypass as an acetyl-CoA source with glucose; it is not known when each source is preferred for cytosolic acetyl-CoA production in these species. Expression of PDC under aerobic glucose growth would have been a prerequisite for the PDH bypass in yeast, but its expression was likely limited to oxygen-limiting conditions in Proto-Yeast (Sanchis et al., 1994). All budding yeasts in this study have homologs of the PDH bypass genes (PDC, NAD or NADP-ALD, and ACS), but the delimiting gene for an active PDH bypass in Saccharomycotina, and ultimately glucose fermentation to ethanol, appears to be the ACS2 paralog. The absence of an active PDH bypass in *L. starkeyi* and *Y. lipolytica* indicate that microbes do not maximize their biomass yields given their metabolic network (Segre et al., 2002; Fong et al., 2003), but likely require chance duplications

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to rewire function.

Expansion of aldehyde dehydrogenase family. NADP-ALD was absent in Proto-Yeast but emerged as cytosolic and mitochondrial enzymes from several duplications in budding yeasts (Figure 3). It's unclear if ALD6.1 or ALD5 originated first from ALD2, but the *S. cerevisiae*'s ALD6 gene, annotated as ALD6.2 in AYbRAH, is paralogous to the first cytosolic NADP-ALD in the budding yeast pan-genome, ALD6.1. ALD2.4, another putative cytosolic NADP-ALD, also emerged within Taphrinomycotina but has not been characterized directly (Van Urk et al., 1990). The loss of cytosolic NADP-ALD in the CTG clade, with the exception of *Debaromyces hansenii*, *Cephaloascus albidus* and *Cephaloascus fragrans* (Grigoriev et al., 2013), is puzzling as ALD6 has been shown to be indispensable for *S. cerevisiae* (Grabowska and Chelstowska, 2003). An alternative NADPH source may have evolved in the CTG clade that reduced the need for cytosolic NADP-ALD. The presence and absence of ALD2.3, the only mitochondrial ALD in Pezizomycotina covered in this study, appears to divide the fungal species into ethanologenic and citrogenic species but further analysis is required with more species; the PDH bypass may be active in *Neurospora crassa* and *Trichoderma reesei* under some conditions.

Dekkera bruxellensis and *D. hansenii* are two Crabtree-positive yeasts with peculiar recent duplications in the ALD family. First, *D. bruxellensis* has a mitochondrial NADP-ALD, ALD6.3, originating from a cytosolic NADP-ALD duplication, ALD6.1; this complements the older mitochondrial NAD(P) paralog, ALD5, present in most yeasts sister to *N. fulvescens*. ALD6.3 is also present in other Crabtree-positive *Dekkera* species, but absent in *Brettanomyces naardenensis*, a closely related Crabtree-negative species (results not shown). Second, *D. hansenii* has a cytosolic NAD(P)-ALD, ALD5.2, derived from a mitochondrial NAD(P)-ALD duplication, ALD5; ALD5.2 is the only known cytosolic NADP-ALD source in the CTG clade covered in this study. These yeasts also deviate from the typical Crabtree effect observed in Saccharomycetaceae: both yeasts possess Complex I and alternative oxidase, while not having any NDI orthologs; *D. bruxellensis* accumulates large amounts of acetic acid when grown aerobically with glucose, but for reasons not entirely understood; fructose 1,6-bisphosphate (F1,6P), an important metabolite elevated during overflow metabolism in many species, was found at a lower concentration in *D. hansenii* than *S. cerevisiae* following glucose addition to non-fasted cells (Sánchez et al., 2006); and finally, both yeasts exhibit the Crabtree effect at a lower glycolytic flux than Saccharomycetaceae yeasts (Van Urk et al., 1990). It's uncertain how these paralogs may directly impact the Crabtree effect, but there is evidence NADP(H) may play an underappreciated role. Most Crabtree-positive yeasts cannot metabolize xylose via NADPH-dependent xylose reductase; *Kluyveromyces* sp. are the only known lineage in Saccharomycetaceae that have an NAD(P)-dependent glyceraldehyde 3-phosphate dehydrogenase (Verho et al., 2002) and can consume xylose; NADPH-dependent nitrate reductase impacts aerobic

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ethanol and acetic acid production in *Pachysolen tannophilus* and *D. bruxellensis* (Galafassi et al., 2013; Moktaduzzaman et al., 2015; Jeffries, 1983). Recent studies have shown the importance metabolites play in allosteric regulation and may offer a template to study how NAD(P)H *directly* or *indirectly* impacts the Crabtree effect (Hackett et al., 2016; Gerosa et al., 2015).

Evolution of phosphofructokinase and its possible connection to the PDH bypass. The emergence of ACS2 was followed by a duplication of PFK2, causing the ancestral homo-oligomeric 6-phosphofructokinase (PFK) to become a hetero-oligomeric enzyme with altered allosteric regulation (Habison et al., 1983; Reuter et al., 2000; Flores et al., 2005). Both forms are initially activated by ATP and subsequently inhibited at higher concentrations. PFK in Pezizomycotina fungi and citrogenic yeasts are inhibited by citrate or phosphoenolpyruvate, in contrast to the recently discovered pyruvate kinase inhibition by citrate in *S. cerevisiae* (Hackett et al., 2016); it is not known if citrate inhibition of pyruvate kinase exists in other budding yeasts. Hetero-oligomeric PFK's in Saccharomycotina are also activated by AMP and fructose 2,6-bisphosphate (Bär et al., 1997; Lorberg et al., 1999; Kirchberger et al., 2002), which has not been found in *Y. lipolytica*'s PFK (Flores et al., 2005). The physiological significance of these changes in allostery has not been explored in the literature, but the emergence of the PDH bypass may mark a shift from negative control of upper glycolysis by citrate under nitrogen limitation with citrogenic/lipogenic yeasts, to the positive control of upper glycolysis by AMP under excess glucose in ethanologenic yeasts. Crabtree-positive yeasts in Saccharomycetaceae have been shown to have elevated AMP levels, a product of acetyl-CoA synthase, relative to Crabtree-negative yeasts (Christen and Sauer, 2011); however, further characterization is needed in other Crabtree-positive and citrogenic yeast lineages. The reduced demand for cytosolic citrate from the loss of ACL and the absence of citrate inhibition in upper glycolysis in *S. cerevisiae* may have enabled F1,6P accumulation, which is an important source of metabolic control (Díaz-Ruiz et al., 2008) during excess glucose and nitrogen-limiting conditions (Hackett et al., 2016). Coarse-grained kinetic studies, similar to a recent investigation on the glycolytic imbalance in *S. cerevisiae* (van Heerden et al., 2014), could help elucidate the advantages of either allosteric interactions given their metabolic networks and environments. Ideal organisms for these studies include *L. starkei* as a proxy for Proto-Yeast, *B. adeninivorans* as a proxy for Proto-Fermenter since it uses the PDH bypass, retains ACL, and has a homo-oligomeric PFK, and finally *N. fulvescens* which gained a hetero-oligomeric PFK.

Biomass yield as a function of NADH dehydrogenase sink. The impact of using Complex I versus NDI as an NADH sink was assessed by optimizing for biomass production (40% protein content) with FBA. The proton-translocating Complex I biomass yielded 0.63 g biomass/g glucose in contrast to 0.50 g biomass/g glucose with non-proton translocating NDI (Figure 5D), which are in agreement with

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

experimental yields when yeasts are grown with glucose (Van Hoek et al., 1998; Van Urk et al., 1990). These results suggest there is a clear disadvantage associated with the loss of Complex I with respect to the biomass yield. While such a loss in yield may not reflect the growth rate, a low yield might suggest a loss in biomass production.

Orientation of alternative NADH dehydrogenase. Phobius plots indicate the position of transmembrane domains in an alternative NADH dehydrogenase (NDH2) can predict its membrane orientation (Figure 5B and Figure 5C). NDI orthologs, which orients towards the matrix, have C-terminal hydrophobic domains; external alternative NADH dehydrogenases (NDE), which orient towards the cytosol, have dominant N-terminal hydrophobic domains and sometimes a C-terminal hydrophobic domain (Figure 5C). *S. pombe* does not encode any NDI orthologs but it can oxidize NADH within the mitochondrial matrix (Crichton et al., 2007); alternative translation start sites in its two NDE orthologs, which are conserved across other *Schizosaccharomyces* species, may translate additional isoforms (see Supplemental information).

Expansion and loss of Type II NADH dehydrogenase. Proto-Yeast likely had three orthologs of NDH2 (Figure 5A): NDI0, the ancestral NDI (Duarte et al., 2003) orthologous to *Escherichia coli*'s NDH-2 gene; NDE0, an ancestral NDE originating from an NDI0 duplication after endosymbiosis; NDE1, an ortholog of *S. cerevisiae*'s NDE1, derived from an NDE0 duplication before the emergence of fungi. NDE1 is conserved in all species reviewed in this study, but there is a repeated loss of internal NADH dehydrogenase in yeasts. NDI0 was lost in the sister clade of *L. starkeyi*. Another ortholog of internal NADH dehydrogenase re-emerged as NDI1 from a duplication of NDE1 in the clade sister to *Ascoidea*; NDI1 was subsequently lost in the *Pichiaceae* and CTG clades. NDI also independently re-emerged in *N. fulvescens*, likely before the loss of its Complex I. NDE1.2 in *S. pombe* originated from an NDE1 duplication in *Schizosaccharomyces*. Unlike *S. cerevisiae*, Proto-Yeast's NDE enzymes likely had NADPH dehydrogenase activity (Tarrío et al., 2006a,b; Melo et al., 2001; Carneiro et al., 2004). It is not known where the NADPH dehydrogenase activity was lost in *S. cerevisiae* evolution, or whether its loss contributed to the Crabtree effect.

The loss of Complex I in the two most widely studied Crabtree-positive yeasts, *S. cerevisiae* and *S. pombe* , led some researchers to believe the loss of Complex I was associated with aerobic fermentation (Gabaldón et al., 2005). Kerscher (2000) claimed it is more advantageous for an organism to synthesize a single polypeptide of NDI than Complex I, one of the largest known proteins, when a carbon source is abundant and rapid growth is essential (Katz et al., 1971; Schwitzguebel and Palmer, 1982). Repression of the ETC in fermentative yeasts would have lead to the loss of Complex I, which was functionally replaced by NDI. However, the genome-sequencing of *D. bruxellensis*, a lesser studied Crabtree positive

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

yeast, contained subunits of Complex I prompting researchers to reconsider this relationship (Procházka et al., 2010). The Kerscher (2000) hypothesis is consistent with *S. cerevisiae* 's physiology but does not consider the evolution and physiology of other yeasts (see Supplemental Information).

Repression and inhibition of internal NADH dehydrogenase may both lead to aerobic ethanol accumulation, but the underlying physiology and evolution can be different. Inhibition may lead to repression, but there can be many reasons why an organism represses an enzyme, such as repressing the ETC to accumulate ethanol to inhibit bacterial growth (Merico et al., 2007). Several studies indicate inhibition is relevant to the ETC. First, substrates with the adenosine moiety were found to inhibit Complex I from bovine heart (Birrell and Hirst, 2013). Second, Complex I is expressed but inactive in exponentially growing cells when grown on ethanol (Ohnishi, 1972); this is also compatible with post-translation modification (PTM), but protein phosphorylation has only been found to stimulate respiration in glioma cells (Pasdois et al., 2003). Third, the short-term Crabtree effect in *S. cerevisiae* results in overflow metabolism within seconds of a glucose pulse in glucose-limited chemostats (Van Urk et al., 1988), but also after more than 40 minutes with *Kluyveromyces lactis* (Kiers et al., 1998) and 60 minutes with *Cyberlindnera jadinii*, formally *Candida utilis* (Van Urk et al., 1988). The ETC is not repressed under these conditions, yet ethanol accumulates before any changes in the proteome can occur; this is also consistent with metabolite inhibition or a PTM in the ETC or TCA cycle. Fourth, NADP was found to inhibit the oxidation of NADH in purified mitochondria from a relative of *S. stipitis* (Camougrand et al., 1988). Finally, neuronal cells exposed to rotenone have a reduction of glucose-derived acetyl-CoA and shift to lipogenesis from glutamate, which are characteristics of the Warburg effect (Worth et al., 2014). These results suggest an underappreciated role of enzyme kinetics in the Crabtree-effect.

F1,6P is a potent inhibitor that is frequently elevated during overflow metabolism in multiple organisms. F1,6P has been found to inhibit Complex III and IV in *S. cerevisiae* at physiological concentrations, but not in *C. jadinii*, a well-studied Crabtree-negative yeast (Díaz-Ruiz et al., 2008; Hammad et al., 2016). Although the evidence for F1,6P inhibition on the ETC is strong, it is likely a single interaction in a cascade of events since there must be a reason for F1,6P accumulation. Crabtree-positive and Crabtree-negative yeasts have been used as a binary classification, but large-scale analysis within Saccharomycetaceae demonstrates aerobic ethanol production occurs over a gradient (Hagman et al., 2013). *C. jadinii* shows early symptoms of the Crabtree effect since it accumulates ethanol after a glucose pulse in glucose-limited chemostats, and has an inactive Complex I during growth on glucose in batch and chemostats. An inactive Complex I in *C. jadinii*, due to inhibition or repression, forces it to rely on Type II NADH dehydrogenase, leading to an increase in its glucose uptake to compensate for its lower P/O ratio relative to Complex I. The increase in glucose uptake plays a role in the imbalance between upper and lower glycolysis, and

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

hence F1,6P accumulation. An inactive Complex I during growth on non-fermentable carbon sources in *C. jadinii*, and lack of increase in F1,6P concentration in *D. hansenii* (Sánchez et al., 2006), precludes F1,6P's inhibition of Complex I or as a master regulator in all Crabtree-positive species. F1,6P may inhibit the ETC of some Saccharomycetaceae species, but there are likely additional mitochondrial inhibitors to the ETC or Complex I. If a yeast consistently found itself in an environment where Complex I was persistently inactivated by inhibition, a single deleterious mutation in one subunit would have been enough to cause Complex I to become defunct, rendering its more than 37 genes into pseudogenes, and eventually junk DNA without a trace of Complex I, despite having a higher biomass yield than NDI.

Additional results and discussions on the consensus species tree, gene duplications, subfunctionalization, neofunctionalization and horizontal gene transfers in central metabolism can be found in Supplemental Information. These include ribosome subunit duplications, phosphoglycerate mutase, citrate synthase, glyceraldehyde 3-phosphate dehydrogenase, UDP-glucose-4-epimerase, isocitrate dehydrogenase, malate synthase, malate dehydrogenase, fumarase, fumarate reductase, succinate semialdehyde, homocitrate synthase, trifunctional enzyme C1- tetrahydrofolate synthase, NAD-glycerol 3-phosphate dehydrogenase, and alcohol dehydrogenase.

Benefits of an ortholog database for genomics. AYbRAH was used to evaluate the gains and losses of genes that may have contributed to the Crabtree effect, but an open source curated ortholog database has additional benefits from providing gene targets for functional characterization to functional gene annotation to streamlining genome-scale reconstructions. First, a curated ortholog database can serve as a repository for orthologs that have been screened and orthologs that require screening. If we want to engineer non-conventional organisms, we need to understand the orthologs that do not exist in model organisms and the set of orthologs that do not have conserved function with model organisms, rather than characterizing all the orthologs in a handful of model organisms. Second, a curated ortholog database can be used to improve and simplify genome annotation. Genes from newly sequenced organisms can be mapped to curated ortholog groups, similar to what has been implemented with eggNOG-mapper (Huerta-Cepas et al., 2017) and YGAP (Proux-Wéra et al., 2012), and create new ortholog groups for de novo genes or genes from recent duplications. Pulling annotations from a curated ortholog database has the advantage of unifying the names and descriptions of genes between organisms, as has been proposed for ribosomal subunits (Ban et al., 2014), and can reduce the number of genes that are unannotated, misannotated, or annotated as hypothetical proteins. And finally, a curated ortholog database can be used to increase the quality and quantity of genome-scale reconstructions. Genome-scale reconstructions inherently require a great deal of curation for orthology assignment and functional annotation, which is often not transparent. Refocusing this effort to curate ortholog groups and their function in an open

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

source database for a pan-genome can allow for improvements to be pushed to all reconstructions, and for genome-scale reconstructions to be compiled for any taxonomic level, from Kingdom to Genus species.

4 CONCLUSION

We developed AYbRAH, an open source ortholog database for Fungi, to determine what gains and losses of metabolic genes may be relevant to the Crabtree effect in budding yeasts, with a focus on the PDH bypass and NADH dehydrogenase. Manual curation was required for gene families with high sequence similarity, often arising from recent gene duplications. FBA simulations indicate the PDH bypass has a higher yield of protein and phospholipids from glucose than ACL, the ancestral acetyl-CoA source in Proto-Yeast. All yeasts in this study have homologs of the PDH bypass but the delimiting gene for its expression with glucose appears to be ACS2, a paralog of the ancestral acetate-inducible ACS. FBA simulations demonstrate Complex I results in a 26% higher biomass yield than internal alternative NADH dehydrogenase when cells are grown with glucose. Complex I is bypassed for NDI in some yeast lineages, and both enzymes have been lost in many independent yeast lineages. We propose the repeated loss of Complex I is from inhibition by a mitochondrial metabolite, which is consistent with the physiology of Crabtree-positive and negative yeasts. By reconstructing the evolution of homologs into orthologs at the pan-genome level, bioinformaticians can improve existing and future genome annotations and improve the accuracy of species trees, which can both help biologists understand the evolution of physiology throughout the tree of life.

5 ACKNOWLEDGMENTS

The authors gratefully acknowledge Prof. Belinda Chang and Ryan Schott for their advice with the phylogenetic analysis. K.C. was supported by Bioconversion Network and NSERC CREATE M3.

Author contributions: K.C. and S.M.Y. performed the phylogenetic analysis. K.C. and R.M. discussed the results. K.C. wrote the manuscript. K.C. and R.M. reviewed the manuscript.

6 ABBREVIATIONS

ACS: acetyl-CoA ALD: aldehyde dehydrogenase ACL: ATP citrate lyase AYbRAH: Annotating Yeasts by Reconstructing Ancestry of Homologs ETC: electron transport chain F1,6P: fructose 1,6-bisphosphate FBA: flux balance analysis FOG: Fungal Ortholog Group HOG: Homolog Group NDE: external alternative NADH dehydrogenase NDI: internal alternative NADH dehydrogenase NDH2: alternative NADH dehydrogenase NUO: Complex I PDC: pyruvate decarboxylase PDH: pyruvate dehydrogenase PFK: 6-phosphofructokinase PHK: phosphoketolase PTM: post translational modification

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

Aiba, S. and Matsuoka, M. (1979). Identification of metabolic model: citrate production from glucose by *Candida lipolytica*. *Biotechnology and Bioengineering*, 21(8):1373–1386.

Altenhoff, A. M., Škunca, N., Glover, N., Train, C.-M., Sueki, A., Piližota, I., Gori, K., Tomiczek, B., Müller, S., Redestig, H., et al. (2015). The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Research*, 43(Database issue):D240–249.

Ban, N., Beckmann, R., Cate, J. H., Dinman, J. D., Dragon, F., Ellis, S. R., Lafontaine, D. L., Lindahl, L., Liljas, A., Lipton, J. M., et al. (2014). A new system for naming ribosomal proteins. *Current Opinion in Structural Biology*, 24:165–169.

Banat, I., Singh, D., and Marchant, R. (1996). The use of a thermotolerant fermentative *Kluyveromyces marxianus* IMB3 yeast strain for ethanol production. *Acta Biotechnologica*, 16(2-3):215–223.

Bär, J., Schellenberger, W., and Kopperschläger, G. (1997). Purification and characterization of phosphofructokinase from the yeast *Kluyveromyces lactis*. *Yeast*, 13(14):1309–1317.

Birrell, J. A. and Hirst, J. (2013). Investigation of NADH binding, hydride transfer, and NAD(+) dissociation during NADH oxidation by mitochondrial Complex I using modified nicotinamide nucleotides. *Biochemistry*, 52(23):4048–4055.

Blank, L. M., Lehmbeck, F., and Sauer, U. (2005). Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Research*, 5(6-7):545–558.

Boulton, C. A. and Ratledge, C. (1983). Use of transition studies in continuous cultures of *Lipomyces starkeyi*, an oleaginous yeast, to investigate the physiology of lipid accumulation. *Microbiology*, 129(9):2871–2876.

Byrne, K. P. and Wolfe, K. H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, 15(10):1456–1461.

Camougrand, N., Cheyrou, A., Henry, M., and Guérin, M. (1988). The alternative respiratory pathway of the yeast *Candida parapsilosis*: oxidation of exogenous NAD(P)H. *Microbiology*, 134(12):3195–3204.

Carneiro, P., Duarte, M., and Videira, A. (2004). The main external alternative NAD(P)H dehydrogenase of *Neurospora crassa* mitochondria. *Biochimica et Biophysica Acta*, 1608(1):45–52.

Christen, S. and Sauer, U. (2011). Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics. *FEMS Yeast Research*, 11(3):263–272.

Connerton, I., Fincham, J., Sandeman, R., and Hynes, M. (1990). Comparison and cross-species expression of the acetyl-CoA synthetase genes of the Ascomycete fungi, *Aspergillus nidulans* and *Neurospora crassa*. *Molecular Microbiology*, 4(3):451–460.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Crichton, P. G., Affourtit, C., and Moore, A. L. (2007). Identification of a mitochondrial alcohol dehydrogenase in *Schizosaccharomyces pombe*: new insights into energy metabolism. *Biochemical Journal*, 401(2):459–464.

De Deken, R. (1966). The Crabtree effect: a regulatory system in yeast. *Microbiology*, 44(2):149–156.

Díaz-Ruiz, R., Avéret, N., Araiza, D., Pinson, B., Uribe-Carvajal, S., Devin, A., and Rigoulet, M. (2008). Mitochondrial oxidative phosphorylation is regulated by fructose 1,6-bisphosphate. A possible role in Crabtree effect induction? *Journal of Biological Chemistry*, 283(40):26948–26955.

Duarte, M., Peters, M., Schulte, U., and Videira, A. (2003). The internal alternative NADH dehydrogenase of *Neurospora crassa* mitochondria. *Biochemical Journal*, 371(3):1005–1011.

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., and Stoeckert, C. J. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, Chapter 6:1–9.

Flores, C.-L., Martínez-Costa, O. H., Sanchez, V., Gancedo, C., and Aragon, J. J. (2005). The dimorphic yeast *Yarrowia lipolytica* possesses an atypical phosphofructokinase: characterization of the enzyme and its encoding gene. *Microbiology*, 151(5):1465–1474.

Fong, S. S., Marciniak, J. Y., and Palsson, B. Ø. (2003). Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale in silico metabolic model. *Journal of Bacteriology*, 185(21):6400–6408.

Gabaldón, T., Rainey, D., and Huynen, M. A. (2005). Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *Journal of Molecular Biology*, 348(4):857–870.

Galafassi, S., Capusoni, C., Moktaduzzaman, M., and Compagno, C. (2013). Utilization of nitrate abolishes the "Custers effect" in *Dekkera bruxellensis* and determines a different pattern of fermentation products. *Journal of Industrial Microbiology & Biotechnology*, 40(3-4):297–303.

Gerosa, L., van Rijsewijk, B. R. H., Christodoulou, D., Kochanowski, K., Schmidt, T. S., Noor, E., and Sauer, U. (2015). Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. *Cell Systems*, 1(4):270–282.

Grabowska, D. and Chelstowska, A. (2003). The ALD6 gene product is indispensable for providing NADPH in yeast cells lacking glucose-6-phosphate dehydrogenase activity. *Journal of Biological Chemistry*, 278(16):13984–13988.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., et al. (2013). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*, 42(Database issue):699–704.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3):307–321.

Habison, A., Kubicek, C. P., and Röhr, M. (1983). Partial purification and regulatory properties of phosphofructokinase from *Aspergillus niger*. *Biochemical Journal*, 209(3):669–676.

Hackett, S. R., Zanotelli, V. R., Xu, W., Goya, J., Park, J. O., Perlman, D. H., Gibney, P. A., Botstein, D., Storey, J. D., and Rabinowitz, J. D. (2016). Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science*, 354(6311).

Hagman, A., Säll, T., Compagno, C., and Piskur, J. (2013). Yeast ”make-accumulate-consume” life strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS One*, 8(7):e68734.

Hammad, N., Rosas-Lemus, M., Uribe-Carvajal, S., Rigoulet, M., and Devin, A. (2016). The Crabtree and Warburg effects: Do metabolite-induced regulations participate in their induction? *Biochimica et Biophysica Acta*, 1857(8):1139–1146.

Hazeu, W. and Donker, R. (1983). A continuous culture study of methanol and formate utilization by the yeast *Pichia pastoris*. *Biotechnology Letters*, 5(6):399–404.

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W.-H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Research*, pages W236–241.

Heavner, B. D. and Price, N. D. (2015). Comparative analysis of yeast metabolic network models highlights progress, opportunities for metabolic reconstruction. *PLoS Computational Biology*, 11(11):e1004530.

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., and Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, pages 835–845.

Hensing, M., Rouwenhorst, R., Heijnen, J., Van Dijken, J., and Pronk, J. (1995). Physiological and technological aspects of large-scale heterologous-protein production with yeasts. *Antonie van Leeuwenhoek*, 67(3):261–279.

Huberts, D. H., Niebel, B., and Heinemann, M. (2012). A flux-sensing mechanism could regulate the switch between respiration and fermentation. *FEMS Yeast Research*, 12(2):118–128.

Huerta-Cepas, J., Forslund, K., Pedro Coelho, L., Szklarczyk, D., Juhl Jensen, L., von Mering, C., and Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8):2115–2122.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016a). ETE 3: Reconstruction, analysis, and visualization of

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phylogenomic data. *Molecular Biology and Evolution*, 33(6):1635–1638.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., et al. (2016b). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44(D1):D286–293.

Hynes, M. J. and Murray, S. L. (2010). ATP-citrate lyase is required for production of cytosolic acetyl coenzyme A and development in *Aspergillus nidulans*. *Eukaryotic Cell*, 9(7):1039–1048.

Jeffries, T. W. (1983). Effects of nitrate on fermentation of xylose and glucose by *Pachysolen tannophilus*. *Nature Biotechnology*, 1(6):503–506.

Jensen, R. A. (2001). Orthologs and paralogs - we need to get it right. *Genome Biology*, 2(8):interactions1002–1.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780.

Katz, R., Kilpatrick, L., and Chance, B. (1971). Acquisition and loss of rotenone sensitivity in *Torulopsis utilis*. *European Journal of Biochemistry*, 21(3):301–307.

Kerscher, S. J. (2000). Diversity and origin of alternative NADH:ubiquinone oxidoreductases. *Biochimica et Biophysica Acta*, 1459(2):274–283.

Kiers, J., Zeeman, A.-M., Luttik, M., Thiele, C., Castrillo, J. I., Steensma, H., Van Dijken, J. P., and Pronk, J. T. (1998). Regulation of alcoholic fermentation in batch and chemostat cultures of *Kluyveromyces lactis* CBS 2359. *Yeast*, 14(5):459–469.

Kirchberger, J., Bär, J., Schellenberger, W., Dihazi, H., and Kopperschläger, G. (2002). 6-phosphofructokinase from *Pichia pastoris*: purification, kinetic and molecular characterization of the enzyme. *Yeast*, 19(11):933–947.

Kriventseva, E. V., Tegenfeldt, F., Petty, T. J., Waterhouse, R. M., Simão, F. A., Pozdnyakov, I. A., Ioannidis, P., and Zdobnov, E. M. (2015). OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Research*, 43(D1):D250–D256.

Kurtzman, C., Fell, J. W., and Boekhout, T. (2011). *The Yeasts: a taxonomic study*. Elsevier.

Larsson, C. and Gustafsson, L. (1993). The role of physiological state in osmotolerance of the salt-tolerant yeast *Debaryomyces hansenii*. *Canadian Journal of Microbiology*, 39(6):603–609.

Lorberg, A., Kirchrath, L., Ernst, J. F., and Heinisch, J. J. (1999). Genetic and biochemical characterization of phosphofructokinase from the opportunistic pathogenic yeast *Candida albicans*. *European Journal of Biochemistry*, 260(1):217–226.

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

identification using the KEGG orthology (KO) as a controlled vocabulary. *Bioinformatics*, 21(19):3787–3793.

Matsen, F., Kodner, R., and Armbrust, E. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1):538.

Melo, A. M., Duarte, M., Møller, I. M., Prokisch, H., Dolan, P. L., Pinto, L., Nelson, M. A., and Videira, A. (2001). The external calcium-dependent NADPH dehydrogenase from *Neurospora crassa* mitochondria. *Journal of Biological Chemistry*, 276(6):3947–3951.

Merico, A., Sulo, P., Piškur, J., and Compagno, C. (2007). Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *FEBS Journal*, 274(4):976–989.

Mi, H., Muruganujan, A., and Thomas, P. D. (2012). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(D1):D377–D386.

Moktaduzzaman, M., Galafassi, S., Capusoni, C., Vigentini, I., Ling, Z., Piškur, J., and Compagno, C. (2015). Galactose utilization sheds new light on sugar metabolism in the sequenced strain *Dekkera bruxellensis* CBS 2499. *FEMS Yeast Research*, 15(2).

Mühlhausen, S., Findeisen, P., Plessmann, U., Urlaub, H., and Kollmar, M. (2016). A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Research*, 26(7):945–955.

Mühlhausen, S. and Kollmar, M. (2014). Molecular phylogeny of sequenced Saccharomycetes reveals polyphyly of the alternative yeast codon usage. *Genome Biology and Evolution*, 6(12):3222–3237.

Nilsson, A. and Nielsen, J. (2016). Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. *Scientific Reports*, 6:22264.

Ohnishi, T. (1972). Factors controlling the occurrence of site I phosphorylation in *C. utilis* mitochondria. *FEBS Letters*, 24(3):305–309.

Österlund, T., Nookaew, I., Bordel, S., and Nielsen, J. (2013). Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Systems Biology*, 7(1):1.

Pasdois, P., Deveaud, C., Voisin, P., Bouchaud, V., Rigoulet, M., and Beauvoit, B. (2003). Contribution of the phosphorylable Complex I in the growth phase-dependent respiration of C6 glioma cells in vitro. *Journal of Bioenergetics and Biomembranes*, 35(5):439–450.

Procházka, E., Poláková, S., Piškur, J., and Sulo, P. (2010). Mitochondrial genome from the facultative anaerobe and petite-positive yeast *Dekkera bruxellensis* contains the NADH dehydrogenase subunit genes. *FEMS Yeast Research*, 10(5):545–557.

Pronk, J. T., Yde Steensma, H., and van Dijken, J. P. (1996). Pyruvate metabolism in *Saccharomyces*

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cerevisiae. *Yeast*, 12(16):1607–1633.

Proux-Wéra, E., Armisén, D., Byrne, K. P., and Wolfe, K. H. (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*, 13(1):237.

Reuter, R., Naumann, M., Bär, J., Haferburg, D., and Kopperschläger, G. (2000). Purification, molecular and kinetic characterization of phosphofructokinase-1 from the yeast *Schizosaccharomyces pombe*: evidence for an unusual subunit composition. *Yeast*, 16(14):1273–1285.

Riley, R., Haridas, S., Wolfe, K. H., Lopes, M. R., Hittinger, C. T., Göker, M., Salamov, A. A., Wisecaver, J. H., Long, T. M., Calvey, C. H., et al. (2016). Comparative genomics of biotechnologically important yeasts. *Proceedings of the National Academy of Sciences*, 113(35):9882–9887.

Rush, B. J. and Fosmer, A. M. (2013). Methods for succinate production. US Patent App. 14/374,464.

Ryabova, O. B., Chmil, O. M., and Sibirny, A. A. (2003). Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast *Hansenula polymorpha*. *FEMS Yeast Research*, 4(2):157–164.

Sánchez, N., Calahorra, M., González-Hernández, J., and Peña, A. (2006). Glycolytic sequence and respiration of *Debaryomyces hansenii* as compared to *Saccharomyces cerevisiae*. *Yeast*, 23(5):361–374.

Sanchis, V., Vinas, I., Roberts, I. N., Jeenes, D. J., Watson, A. J., and Archer, D. B. (1994). A pyruvate decarboxylase gene from *Aspergillus parasiticus*. *FEMS Microbiology Letters*, 117(2):207–210.

Schneider, H., Wang, P., Chan, Y., and Maleszka, R. (1981). Conversion of D-xylose into ethanol by the yeast *Pachysolen tannophilus*. *Biotechnology Letters*, 3(2):89–92.

Schwitzguebel, J. and Palmer, J. (1982). Properties of mitochondria as a function of the growth stages of *Neurospora crassa*. *Journal of Bacteriology*, 149(2):612–619.

Segre, D., Vitkup, D., and Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117.

Shashi, K., Bachhawat, A. K., and Joseph, R. (1990). ATP:citrate lyase of *Rhodotorula gracilis*: purification and properties. *Biochimica et Biophysica Acta*, 1033(1):23–30.

Shen, X.-X., Zhou, X., Kominek, J., Kurtzman, C. P., Hittinger, C. T., and Rokas, A. (2016). Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. *G3: Genes, Genomes, Genetics*, 6(12):3927–3939.

Slininger, P., Bothast, R., Van Cauwenberge, J., and Kurtzman, C. (1982). Conversion of D-xylose to ethanol by the yeast *Pachysolen tannophilus*. *Biotechnology and Bioengineering*, 24(2):371–384.

Sonnhammer, E. L. and Östlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Research*, 43(D1):D234–D239.

Sousa, M. J., Rodrigues, F., Coôrte-Real, M., and Leão, C. (1998). Mechanisms underlying the transport and intracellular metabolism of acetic acid in the presence of glucose in the yeast *Zygosaccharomyces*

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bailii. *Microbiology*, 144(3):665–670.

Sukumaran, J. and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, 26(12):1569–1571.

Tarrío, N., Becerra, M., Cerdán, M. E., and Siso, M. I. G. (2006a). Reoxidation of cytosolic NADPH in *Kluyveromyces lactis*. *FEMS Yeast Research*, 6(3):371–380.

Tarrío, N., Cerdán, M. E., and Siso, M. I. G. (2006b). Characterization of the second external alternative dehydrogenase from mitochondria of the respiratory yeast *Kluyveromyces lactis*. *Biochimica et Biophysica Acta*, 1757(11):1476–1484.

Toivola, A., Yarrow, D., Van Den Bosch, E., Van Dijken, J. P., and Scheffers, W. A. (1984). Alcoholic fermentation of D-xylose by yeasts. *Applied and Environmental Microbiology*, 47(6):1221–1223.

van den Berg, M. A., de Jong-Gubbels, P., Kortland, C. J., van Dijken, J. P., Pronk, J. T., and Steensma, H. Y. (1996). The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. *Journal of Biological Chemistry*, 271(46):28953–28959.

van Heerden, J. H., Wortel, M. T., Bruggeman, F. J., Heijnen, J. J., Bollen, Y. J., Planqué, R., Hulshof, J., O'Toole, T. G., Wahl, S. A., and Teusink, B. (2014). Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174):1245114.

Van Hoek, P., Van Dijken, J. P., and Pronk, J. T. (1998). Effect of specific growth rate on fermentative capacity of baker's yeast. *Applied and Environmental Microbiology*, 64(11):4226–4233.

van Rossum, H. M., Kozak, B. U., Pronk, J. T., and van Maris, A. J. (2016). Engineering cytosolic acetyl-coenzyme A supply in *Saccharomyces cerevisiae*: pathway stoichiometry, free-energy conservation and redox-cofactor balancing. *Metabolic Engineering*, 36:99–115.

Van Urk, H., Mark, P. R., Scheffers, W. A., and Van Dijken, J. P. (1988). Metabolic responses of *Saccharomyces cerevisiae* CBS 8066 and *Candida utilis* CBS 621 upon transition from glucose limitation to glucose excess. *Yeast*, 4(4):283–291.

Van Urk, H., Voll, W. L., Scheffers, W. A., and Van Dijken, J. P. (1990). Transient-state analysis of metabolic fluxes in Crabtree-positive and Crabtree-negative yeasts. *Applied and Environmental Microbiology*, 56(1):281–287.

Vazquez, A. and Oltvai, Z. N. (2016). Macromolecular crowding explains overflow metabolism in cells. *Scientific Reports*, 6:31007.

Vemuri, G., Eiteman, M., McEwen, J., Olsson, L., and Nielsen, J. (2007). Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 104(7):2402–2407.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Venayak, N., Anesiadis, N., Cluett, W. R., and Mahadevan, R. (2015). Engineering metabolism through dynamic control. *Current Opinion in Biotechnology*, 34:142–152.

Verho, R., Richard, P., Jonson, P. H., Sundqvist, L., Londesborough, J., and Penttilä, M. (2002). Identification of the first fungal NADP-GAPDH from *Kluyveromyces lactis*. *Biochemistry*, 41(46):13833–13838.

Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, 19(2):327–335.

Worth, A. J., Basu, S. S., Snyder, N. W., Mesaros, C., and Blair, I. A. (2014). Inhibition of neuronal cell mitochondrial Complex I with rotenone increases lipid β-oxidation, supporting acetyl-coenzyme A levels. *Journal of Biological Chemistry*, 289(39):26895–26903.

Wu, C. H., Nikolskaya, A., Huang, H., Yeh, L.-S. L., Natale, D. A., Vinayaka, C., Hu, Z.-Z., Mazumder, R., Kumar, S., Kourtesis, P., et al. (2004). PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Research*, 32(Database issue):D112–D114.

Zeeman, A.-M. and Steensma, H. (2003). The acetyl co-enzyme A synthetase genes of *Kluyveromyces lactis*. *Yeast*, 20(1):13–23.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Fungal species included in AYbRAH

Species	Strain	Database
Schizosaccharomyces pombe	972h-	UniProt
Aspergillus niger	CBS 513.88	UniProt
Neurospora crassa	CBS708.71	UniProt
Trichoderma reesei	QM6a	UniProt
Yarrowia lipolytica	CLIB 122	UniProt
Komagataella pastoris	GS115	UniProt
Kuraishia capsulata	CBS 1993	UniProt
Ogataea polymorpha	NRRL Y-7560	UniProt
Pichia kudriavzevii	SD108	UniProt
Meyerozyma guilliermondii	CBS 566	UniProt
Debaryomyces hansenii	CBS 767	UniProt
Scheffersomyces stipitis	CBS 6054	UniProt
Spathaspora passalidarum	NRRL Y-27907	UniProt
Kluyveromyces lactis	CBS 2359	UniProt
Lachancea thermotolerans	CBS 6340	UniProt
Zygosaccharomyces rouxii	CBS 732	UniProt
Saccharomyces cerevisiae	S288C	UniProt
Vanderwaltozyma polyspora	DSM 70294	UniProt
Rhodotorula graminis	WP1	MycoCosm
Saitoella complicata	NRRL Y-17804	MycoCosm
Lipomyces starkeyi	NRRL Y-11557	MycoCosm
Blastobotrys adeninivorans	LS3	MycoCosm
Nadsonia fulvescens var. elongata	DSM 6959	MycoCosm
Pachysolen tannophilus	NRRL Y-2460	MycoCosm
Ogataea arabinofermentans	NRRL YB-2248	MycoCosm
Dekkera bruxellensis	CBS 2499	MycoCosm
Pichia membranifaciens	NRRL Y-2026	MycoCosm
Ascoidea rubescens	NRRL Y17699	MycoCosm
Babjeviella inositovora	NRRL Y-12698	MycoCosm
Metschnikowia bicuspidata	NRRL YB-4993	MycoCosm
Wickerhamomyces anomalus	NRRL Y-366-8	MycoCosm
Cyberlindnera jadinii	NRRL Y-1542	MycoCosm
Hanseniaspora valbyensis	NRRL Y-1626	MycoCosm

Table 2. AYbRAH statistics following an curation.

	AYbRAH	
	v0.1	v0.1.5
Genes	169 118 (79%)	183 359 (86%)
Ortholog groups	14 249	22 473
Homolog groups	0	15 391
Curated ortholog groups	0	522
Electronically modified ortholog groups	0	2508

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

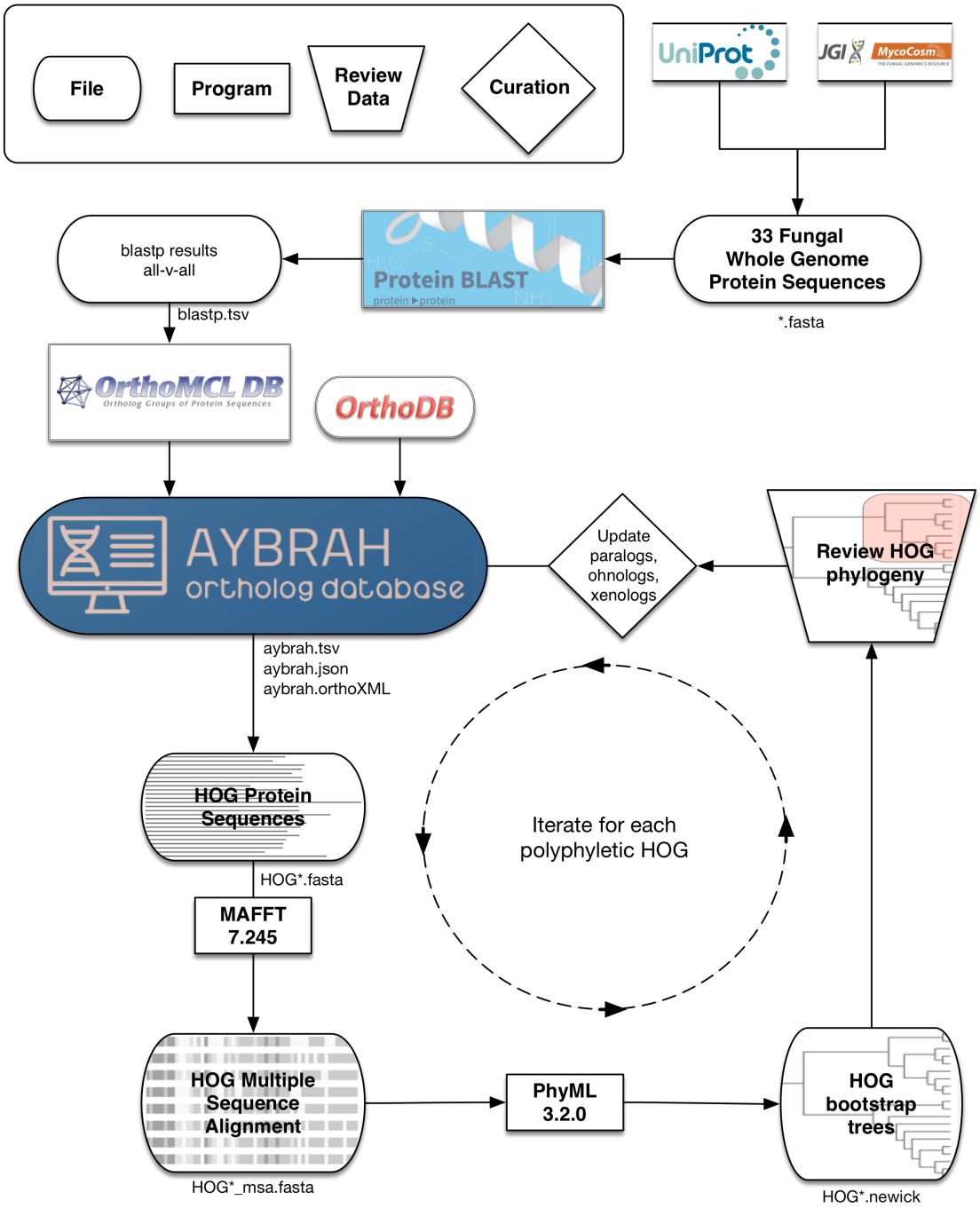


Figure 1. AYbRAH workflow for ortholog curation. Protein sequences from UniProt and MycoCosm were fed into OrthoMCL using standard parameters. Putative ortholog groups were assigned Fungi-level OrthologDB identification codes to cluster ortholog groups into Homolog Groups (HOG's). Multiple sequence alignments were performed with MAFFT for each HOG, and 100 bootstrap phylogenetic trees computed with PhyML. Consensus phylogenetic trees were reviewed for polyphyletic Fungal Ortholog Groups (FOG) in central metabolism to assigned curated ortholog groups.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

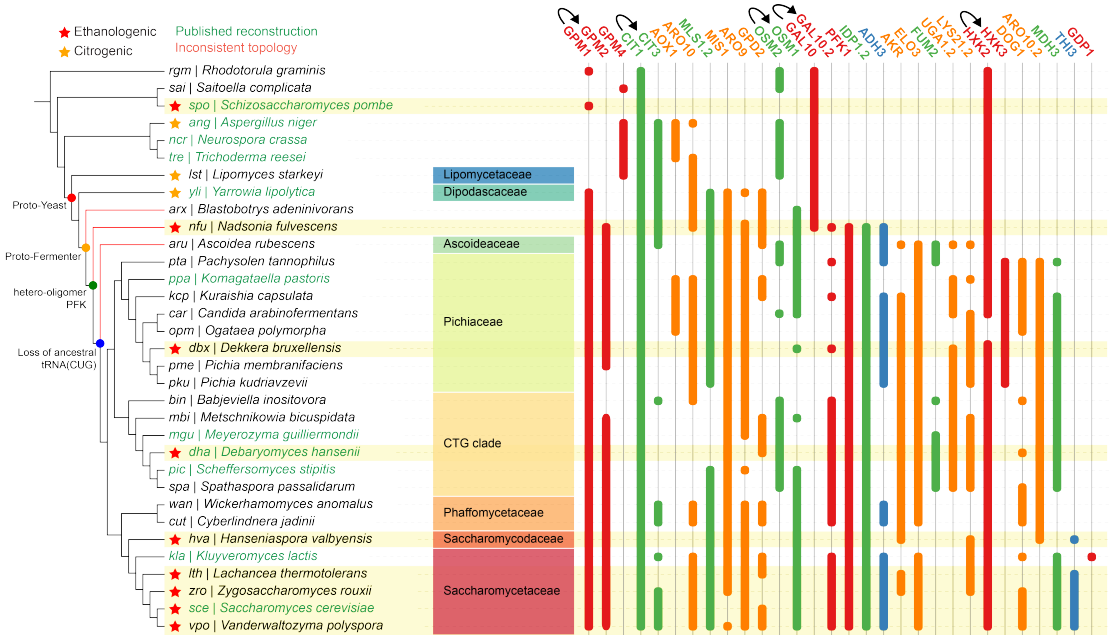


Figure 2. Manually curated glycolytic orthologs (red columns), TCA orthologs (green columns), pyruvate metabolism-related orthologs (blue columns), and other metabolic orthologs (orange) mapped to a budding yeast species tree; a Basidiomycete yeast, Taphrinomycotina yeasts, and Pezizomycotina fungi were used as outgroups. Family/clade names are shown next to species names. Proto-Yeast is the ancestor of all budding yeasts in Saccharomycotina; Proto-Fermenter is the first budding yeast to have used the PDH bypass following the duplication of ACS1 to ACS2, gaining the ability to ferment sugars to ethanol aerobically or with oxygen limitation; hetero-oligomer PFK denotes the transition from the ancestral homo-oligomeric PFK to the hetero-oligomeric PFK with altered allostery in budding yeasts; loss of ancestral tRNA(CUG) indicates the earliest known point for the loss of ancestral CUG tRNA in budding yeasts and subsequent reassignment in *P. tannophilus* and the CTG clade. Ethanologenic and Crabtree-positive yeasts are labeled with red stars and citrogenic fungi/yeasts with yellow stars. Species with published genome-scale reconstructions are labeled with green text. Gene names reflect *S. cerevisiae* assignments except for GPM4, the cofactor-independent phosphoglycerate mutase; AOX1, alcohol oxidase; MLS1.2, the peroxisomal malate synthase paralog from the ancestral cytoplasmic and peroxisomal malate synthase; OSM2, an uncharacterized fumarate reductase paralogous to OSM1 in *S. cerevisiae*; GAL10.2, UDP-glucose-4-epimerase orthologous to GAL10 in *S. cerevisiae*; IDP1.2, mitochondrial isocitrate dehydrogenase orthologous to IDP1 in *S. cerevisiae*; FUM2, an uncharacterized cytoplasmic fumarase paralogous to FUM1 in *S. cerevisiae*; UGA1.2, a mitochondrial gamma-aminobutyrate transaminase; LYS21.2, a cytoplasmic homocitrate synthase; HXK3, a paralog of *S. cerevisiae's* HXK2; ARO10.2, an uncharacterized transaminated amino acid decarboxylase. Arrows indicate the predicted direction of the gene duplication. The topology of our tree is based on recently published species trees (Mühlhausen and Kollmar, 2014; Riley et al., 2016; Shen et al., 2016), but with a modified topology for *B. adeninivorans*, *N. fulvescens* and *A. rubescens* to minimize homoplasy of gene duplications. Ortholog columns are sorted by their earliest duplication (except for the gene families for GPM, CIT, OSM, GAL10, HXK) to show support for our topology. We find support for *B. adeninivorans*' topology with OSM1 and ACS2; *N. fulvescens*' placement with GAL10.2, PFK1, IDP1.2, ADH3; *A. rubescens*' topology with AKR, ELO3, LYS21.2, and to a lesser extent FUM2 and UGA1.2. *Kluyveromyces lactis* is the only yeast in Saccharomycetaceae which is Crabtree-negative and can grow with xylose; it differs from its family by the gain of the GDP1 paralog and loss of THI3.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

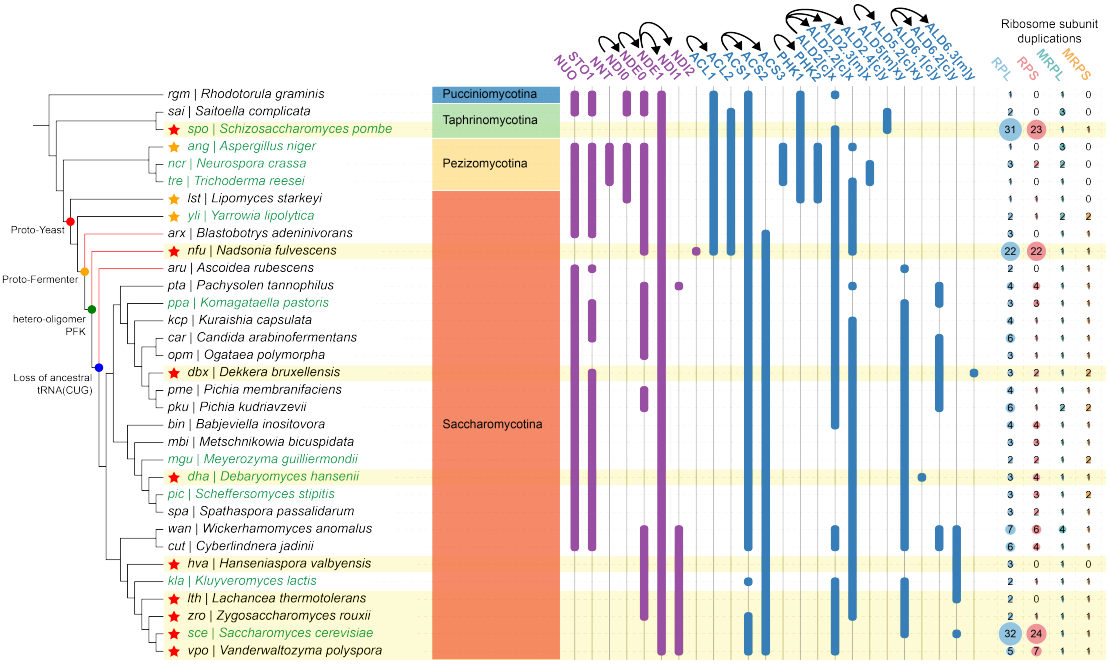
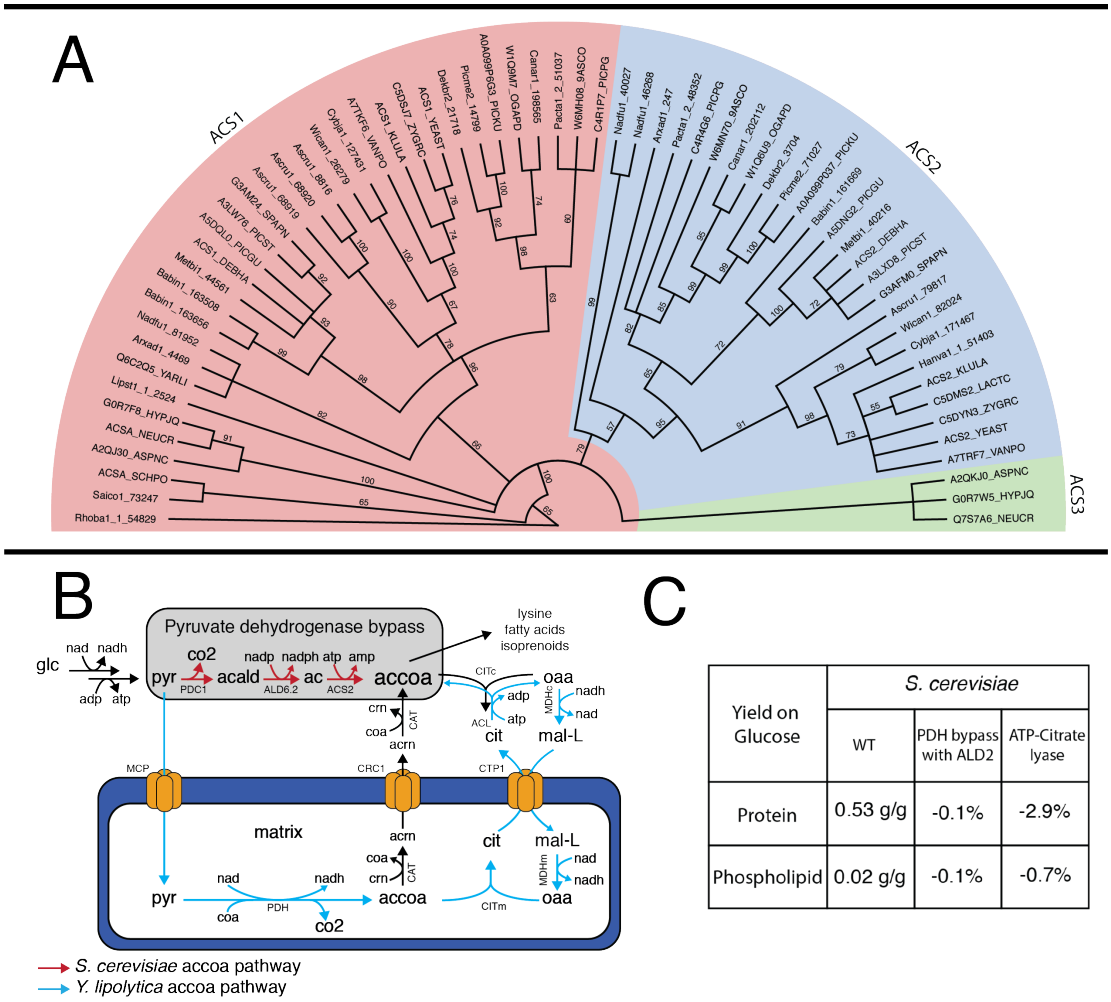


Figure 3. Manually curated ETC orthologs (blue columns), acetyl-CoA-related orthologs (purple columns) and ribosomal protein subunit duplications mapped to a budding yeast species tree. See Figure 2 for shared annotations. Subphylum names are shown next to species names. Enzymes with predicted cytoplasmic and mitochondrial localizations are indicated with [c] and [m], respectively. Oxidoreductases using NAD(H) and NADP(H) are designated with x and y, respectively. NUO, Complex I; STO1, alternative oxidase; NNT, membrane-bound transhydrogenase; NDI, internal NADH dehydrogenase; NDE, external NAD(P)H dehydrogenase; ALD, aldehyde dehydrogenase; PHK, phosphoketolase; ACL, ATP-citrate lyase; ACS, acetyl-CoA synthetase; RPL, ribosome protein of the large subunit; RPS, ribosome protein of the small subunit; MRPL, mitochondrial ribosomal protein of the large subunit; MRPS, mitochondrial ribosomal protein of the small subunit. No yeasts have maintained alternative oxidase while having lost Complex I; however, *P. tannophilus* and *O. polymorpha* have both lost alternative oxidase yet maintain Complex I. Transhydrogenase is absent in Rhodotorula, Taphrinomycotina, and Saccharomycotina, the three independent yeast lineages in this study. The ancestral internal alternative NADH dehydrogenase (NDI0) that was present in Proto-Yeast is only maintained in *L. starkeyi* within Saccharomycotina, while the older external NADH dehydrogenase (NDE0) is present in most clades except the CTG clade. NDE1, an external NADH dehydrogenase, is the only NDH2 ortholog maintained in all yeast lineages, although its activity with NADPH is not conserved in *S. cerevisiae*. Internal NADH dehydrogenase re-emerged as NDI1 from an NDE1 duplication, but was not maintained in the CTG clade or most of the Pichiaceae; NDI2 independently evolved in *N. fulvescens*. *L. starkeyi* is the only budding yeast that has kept all orthologs from the three cytosolic acetyl-CoA sources: ACL, ACS, and PHK. ACL and PHK both had duplications in Dikarya that led to hetero-oligomeric enzymes. Cytoplasmic NAD-ALD was present in Proto-Yeast, but duplications led to mitochondrial localizations (ALD2.3, ALD5, ALD6.3) and NADP activity (ALD2.4, ALD5, ALD5.2, ALD6.1, ALD6.2, ALD6.3). ALD5.2 and ALD6.3 are recent duplications that are present in *D. bruxellensis* and *D. hansenii*, two independent lineages of Crabtree-positive yeasts. *S. pombe*, *N. fulvescens*, and *S. cerevisiae* are three Crabtree-positive yeasts that have significant independent duplications in their cytoplasmic ribosome subunits.

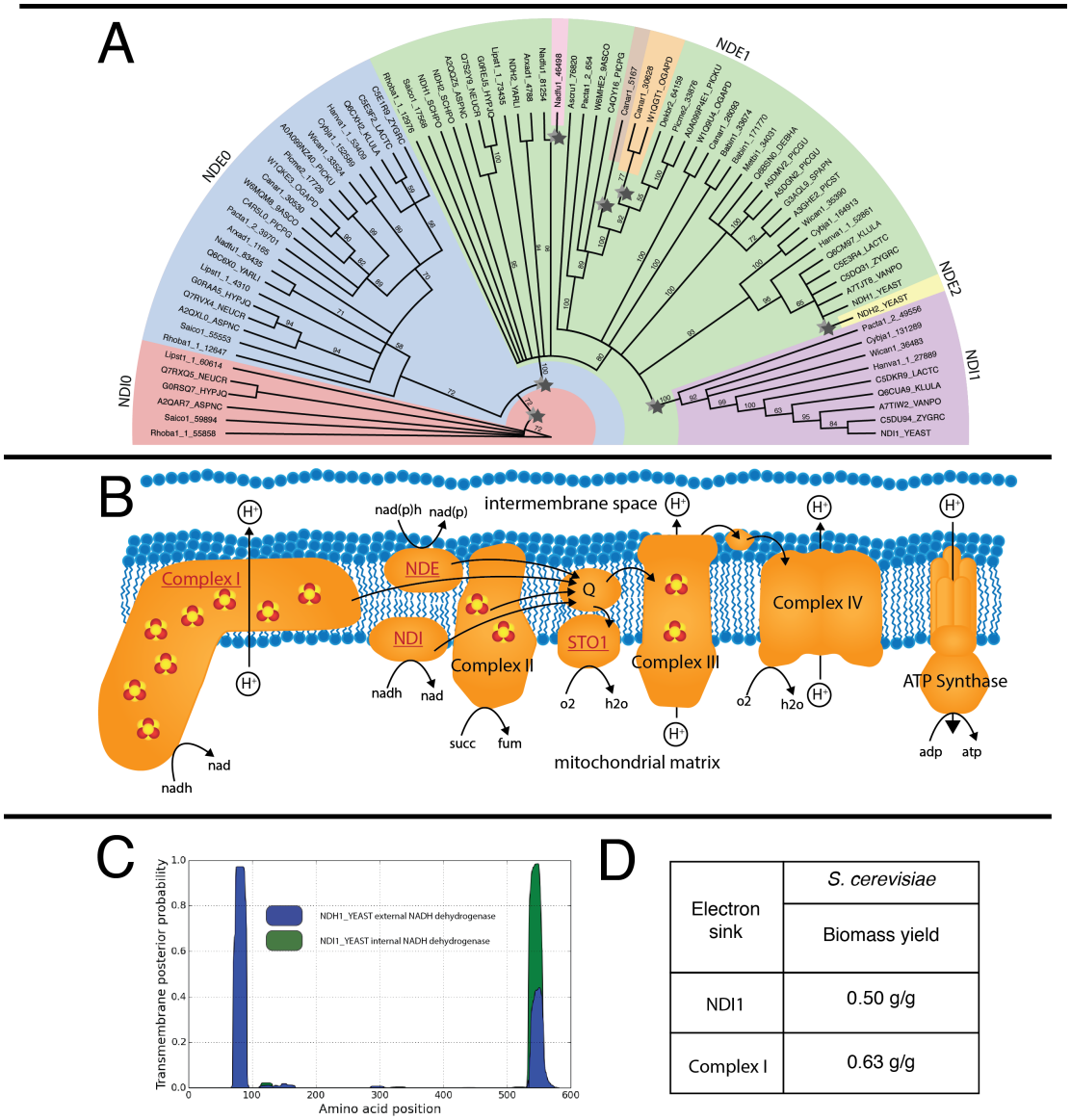
bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Yield on Glucose	*S. cerevisiae*		
	WT	PDH bypass with ALD2	ATP-Citrate lyase
Protein	0.53 g/g	-0.1%	-2.9%
Phospholipid	0.02 g/g	-0.1%	-0.7%

Figure 4. **(A)** Phylogenetic tree showing the three major ortholog groups in the ACS family: ACS1, acetate-inducible ACS; ACS2, glucose-inducible acetyl-CoA synthease; ACS3, propionyl-CoA synthetase. **(B)** Metabolic pathways for acetyl-CoA production in *S. cerevisiae* via the PDH bypass and *Y. lipolytica* via ACL. Gene annotations not consistent with *S. cerevisiae*'s genes: MCP, mitochondrial pyruvate carrier; PDH, PDH complex; CITm, mitochondrial citrate synthase; CITc; cytoplasmic citrate synthase; MDHm, mitochondrial malate dehydrogenase; MDHc, cytoplasmic malate dehydrogenase; CAT, carnitine o-acetyltransferase. **(C)** FBA predictions for protein and phospholipid yield from glucose with the PDH bypass (via NADP-ALD), PDH bypass (via NAD-ALD), and ACL using a genome-scale metabolic reconstruction for *S. cerevisiae* . Highest yields of protein and phospholipids occur with the PDH bypass via ALD6, which is absent in Proto-Yeast.

bioRxiv preprint doi: https://doi.org/10.1101/237974; this version posted December 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Electron sink	*S. cerevisiae*
	Biomass yield
NDI1	0.50 g/g
Complex I	0.63 g/g

Figure 5. **(A)** Phylogenetic reconstruction of the Type II membrane-bound NADH dehydrogenase family. Grey stars indicate the origin of duplications. Major ortholog groups are highlighted in different colors and designated as internal NADH dehydrogenase with NDI and external NADH dehydrogenase with NDE. **(B)** Organization of the ETC in fungi. Complex I, external and internal alternative NADH dehydrogenase and alternative oxidase are enzymes not conserved across all fungi in this study and are designated with red text with underlines. **(C)** Transmembrane posterior probability for aligned *S. cerevisiae* internal NADH dehydrogenase (NDI1_YEAST) and external NADH dehydrogenase (NDE1_YEAST). NDI orthologs have C-terminal transmembrane predictions; NDE orthologs have N-terminal transmembrane predictions and sometimes lower C-terminal transmembrane predictions. **(D)** Biomass yield as a function of non-proton-translocating NDI1 and proton-translocating Complex I in *S. cerevisiae* genome-scale metabolic reconstruction.