Convergent evolution of aerobic fermentation through divergent mechanisms acting on key shared glycolytic genes

Linda C. Horianopoulos[1,2]*, Antonis Rokas[3], Chris Todd Hittinger[1]*

[1] Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Center for Genomic Science Innovation, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI 53726, USA

[2] Department of Food Science, University of Guelph, Guelph, ON, Canada, N1G 2W1

[3] Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA; Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA

*corresponding authors: lhoriano@uoguelph.ca and cthittinger@wisc.edu

Keywords: transcriptional rewiring, cis-regulatory evolution, gene duplication, aerobic fermentation, glycolysis, Crabtree/Warburg Effect

ABSTRACT

As the tree of life becomes increasingly accessible to molecular investigations, describing mechanisms underlying evolutionary convergence and constraint will be crucial to understanding diversification. The lineage including the model yeast *Saccharomyces cerevisiae* evolved aerobic fermentation in part through an ancient whole genome duplication and retention of glycolytic genes. To evaluate the glycolytic rates across diverse yeasts, we developed and deployed an extracellular acidification rates (ECAR) assay on 299 species that span more than 400 million years of evolution and identified a clade in the genus *Saturnispora* that convergently evolved aerobic fermentation. Through comparative genomics and transcriptomics, we found that several glycolytic genes had higher expression and novel cis-regulatory elements in aerobically fermenting *Saturnispora* species. When the transcription factor required for their activation was deleted in *Saturnispora dispora,* the mutants had reduced glycolytic rates and increased respiration. Intriguingly, many of the upregulated genes are orthologous to duplicated glycolytic genes in *S. cerevisiae.* These divergent genetic mechanisms suggest that there are strong evolutionary constraints on how some traits like aerobic fermentation can arise convergently.

INTRODUCTION

The convergent evolution of complex traits provides insights into the mechanisms and constraints that govern their repeated emergence. Convergence can arise through distinct or common genetic mechanisms. One of the most compelling examples of convergent evolution in a complex metabolic trait is the evolution of C4 photosynthesis in more than 60 independent lineages (Sage *et al*, 2011). C4 plants have repeatedly and independently evolved mechanisms to increase photosynthesis efficiency by partitioning biochemical reactions and allowing the key photosynthesis enzyme, RuBisCO, to be in a high CO_2 environment (Brown *et al*, 2011; Williams *et al*, 2012; Lyu *et al*, 2025). Comparisons across C4 grasses have identified multiple underlying

53 evolutionary mechanisms, including gene family expansions of key enzymes, parallel amino
54 acid changes, and similar cis-regulatory elements in the promoters of C4 enzymes (Lyu *et al*,
55 2024). For example, a specific comparison between rice (C3) and sorghum (C4) revealed that
56 the promoters driving the C4 enzymes in sorghum co-opted an ancestral cis-regulatory element
57 for cell type specificity, thereby promoting this physical partitioning of biochemical reactions
58 (Swift *et al*, 2024). These innovations highlight how improved metabolic efficiency has been
59 rediscovered in many independent lineages through different mechanisms altering the
60 abundance, expression patterns, or activity of the same key C4 enzymes.

61 Within the yeasts of the subphylum Saccharomycotina, the Crabtree/Warburg Effect, in
62 which yeasts ferment even in the presence of oxygen, is a complex metabolic trait that has
63 evolved in at least two independent lineages, in *Saccharomyces* and relatives and some
64 *Brettanomyces* species (Rozpedowska *et al*, 2011). Comparative phenotyping of up to forty
65 species demonstrated that the Crabtree-positive phenotype likely evolved in a stepwise fashion
66 and that overflow metabolism, wherein cells maximize energy production rates through
67 glycolysis but produce fermentation products, is the fundamental mechanism driving aerobic
68 fermentation (Hagman & Piškur, 2015; Hagman *et al*, 2013). Some yeasts, including *S.*
69 *cerevisiae*, have undergone dramatic genetic changes, including preferential retention of
70 duplicated glycolytic genes from an ancient whole genome duplication (WGD), further
71 increasing their fermentative capacity (Conant & Wolfe, 2007; Escalera-Fanjul *et al*, 2019).
72 *Brettanomyces* species, which are also Crabtree-positive, do not have the high glucose
73 consumption rates of *S. cerevisiae*, but they do repress respiration when glucose is present
74 (Hagman *et al*, 2013; Rozpedowska *et al*, 2011). These independent occurrences of aerobic
75 fermentation suggest that there may be additional lineages of yeasts with independently evolved
76 preferences for fermentation.

77 Complex traits, such as the Crabtree/Warburg Effect, are difficult to assess using high
78 throughput assays. However, since a high glycolytic rate and overflow metabolism are thought

79 to precede the evolution of the Crabtree/Warburg Effect, we developed an assay to evaluate the
80 glycolytic rate of diverse species based on glucose-induced extracellular acidification rates
81 (ECAR). ECAR has been used to study mammalian cells, primarily cancer cells, which display
82 the Crabtree/Warburg Effect wherein they accumulate lactate as a byproduct of rapid aerobic
83 glycolysis (Schmidt *et al*, 2021; Zhang & Zhang, 2019). Acidification measurements have also
84 been applied in *S. cerevisiae* to assess fermentative capacity in industrial settings (Sigler, 2013;
85 M & K, 1982; Yamashoji et al, 2020; Sigler et al, 1981), and more recently, ECAR was
86 measured with the Seahorse Extracellular Flux analyzer to assess glycolytic rate in pathogenic
87 yeasts (Tucey *et al*, 2018; Yoo *et al*, 2024; Zhang *et al*, 2018). Assessing glycolytic rates across
88 a wide range of species could potentially uncover novel instances of high glycolytic rates and
89 provide insight into the genetic mechanisms underlying high overflow metabolism. Identification
90 of species with high glycolytic rates may also have important implications as increasing efforts
91 are being made to use non-conventional yeasts for biotechnological applications (Geijer *et al*,
92 2022; Wang *et al*, 2025; Radecka *et al*, 2015; Pyne *et al*, 2023; Patra *et al*, 2021).
93 Herein, we developed a microtiter plate-based assay to screen diverse yeast species for
94 ECAR. We confirmed that the species in the order Saccharomycetales, including the model
95 yeast *S. cerevisiae*, display rapid ECAR phenotypes, which are consistent with their rapid
96 glycolytic rates. However, we also identify multiple species in the genus *Saturnispora*, including
97 *Saturnispora dispora*, that have rapid ECAR compared to other species in the order Pichiales.
98 This order already contains two yeasts of considerable biotechnological interest: *Komagataella*
99 *phaffii* (previously known as *Pichia pastoris*), which is a leading chassis for protein production
100 (Heistinger *et al*, 2020; Gasser & Mattanovich, 2018), and *Pichia kudriavzevii* (also known as
101 *Issatchenkia orientails*, *Candida glycerinogenes*, and *Candida krusei*), which is emerging as an
102 important microbial cell factory for the production of organic acids (Tan *et al*, 2025; Suthers &
103 Maranas, 2022). Furthermore, additional species within the genus *Pichia* are actively being
104 developed for organic acid production (Pyne *et al*, 2023). This emerging interest in the industrial

105 applications of Pichiales warrants further work to understand the metabolic capabilities of this
106 order to guide their development as microbial cell factories. Moreover, characterizing the genetic
107 mechanism driving the strong fermentative capacity in *Saturnispora* can reveal the evolutionary
108 constraints (if any) on rapid fermentation and guide rational strategies to maximize rates of
109 organic acid production in Pichiales.

110 To understand how the unusually high glycolytic rate of *Sat. dispora* and its relatives
111 evolved, we used comparative genomics and transcriptomics and found that several genes
112 encoding glycolytic enzymes had increased expression at the transcript level in species with a
113 rapid ECAR phenotype. Intriguingly, genes orthologous to those whose duplications are key to
114 the high glycolytic rate of *S. cerevisiae* also had higher expression in the rapid ECAR
115 *Saturnispora*, which prompted us to investigate the promoter sequences of glycolytic genes for
116 binding sites of transcription factors (TFs) known to regulate carbon metabolism. We found that
117 there were more predicted Gal4p-binding sites ($CGGN_{11}CCG$) in promoters of some glycolytic
118 genes in the rapid ECAR species. Although Gal4p is named for its role in galactose metabolism
119 in *S. cerevisiae* (Johnston, 1987), this role is evolutionarily derived and specific to
120 Saccharomycetales (Harrison *et al*, 2022). Gal4p homologs play important roles regulating
121 carbon metabolism in *Candida albicans* (Askew *et al*, 2009) and glycolysis in *K. phaffii*, although
122 these species do not natively exhibit aerobic fermentation (Ata *et al*, 2018). Using novel genetic
123 tools that we developed for *Sat. dispora*, we showed that *GAL4* was necessary for the induction
124 of the most highly expressed glycolytic gene, *TDH*. Altogether, we found an independently
125 evolved group of yeasts with a high glycolytic rate phenotype achieved through the recruitment
126 of cis-regulatory elements driving the direct transcriptional upregulation of orthologs of the same
127 glycolytic genes that retain duplicates in *S. cerevisiae*. This convergence suggests that the
128 evolution of increased glycolytic rate is constrained and requires high activity of key glycolytic
129 enzymes, which can be achieved by distinct molecular mechanisms.
130

131 **RESULTS**

132 **Rapid ECAR in *Saccharomyces* and relatives and in the distantly related genus**
133 ***Saturnispora***

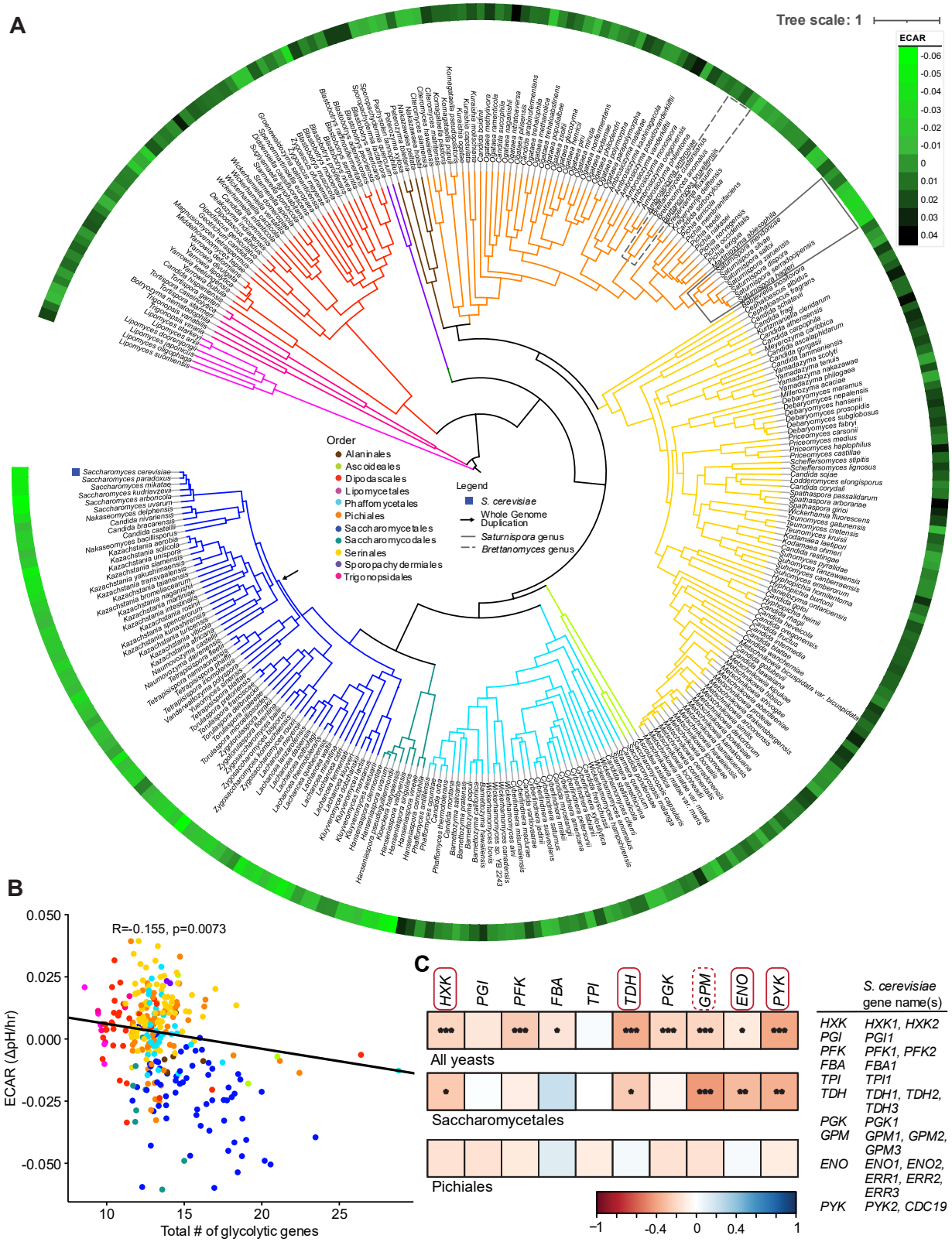
134 We developed a high-throughput method to measure the glucose-dependent ECAR of
135 yeast species as a proxy for glycolytic rate using a 96-well plate and a standard microplate
136 reader (Fig. S1). We plotted the mean of three biological replicates per species as a colour
137 gradient across the yeast phylogeny to identify clades with similar ECAR profiles. Accordingly,
138 we found that many species in the order Saccharomycetales, particularly those yeasts that
139 underwent an ancestral whole genome duplication (WGD), such as *S. cerevisiae*, had rapid
140 acidification (Fig. 1A). Although this result was expected and validated the ability of our method
141 to capture glycolytic rate, we were particularly interested in other clades that showed a rapid
142 ECAR phenotype. For example, we were surprised to find a group in the order Pichiales in the
143 genus *Saturnispora* that showed a rapid ECAR phenotype (as low as -0.032 pH/hr in *Sat.*
144 *dispora*). These species have not been thoroughly characterized but are in the same order as
145 multiple biotechnologically relevant species, including the Crabtree-negative species *K. pastoris*
146 (low ECAR=0.022 pH/hr). The growing biotechnological interest in Pichiales yeasts prompted
147 our specific interest in the metabolism of these species and in the evolution of the rapid ECAR
148 trait in some *Saturnispora* species.

149

150 **High glycolytic rate correlates with gene duplication in *S. cerevisiae* and relatives**

151 The high glycolytic rate in *S. cerevisiae* is hypothesized to be, at least in part, due to an
152 ancestral WGD and subsequent preferential retention of glycolytic genes (Conant & Wolfe,
153 2007; Hagman *et al*, 2013; Conant & Wolfe, 2008). To further validate the results of our screen
154 and test the generality of the hypothesis that having a higher number of glycolytic genes is
155 associated with rapid ECAR across yeasts, we compared the number of glycolytic genes and
156 the measured ECAR from each yeast species in our dataset. We found a significant correlation

157 between the total number of glycolytic genes and ECAR across all yeast species after correcting
158 for phylogeny (Fig. 1B). To interrogate this pattern further, we examined the correlations
159 between the size of the gene families encoding each enzymatic step in glycolysis and ECAR.
160 Again, we found strong correlations when looking at all species (Fig. 1C). However, when we
161 focused on the taxonomic orders containing rapid ECAR species, we noticed that several of
162 these correlations held true in the order Saccharomycetales but were not significant within the
163 order Pichiales (Fig. 1C). When the correlations of individual genes were corrected for
164 phylogeny, we found that the significance values of their correlations were lost in the
165 Saccharomycetales, likely due to the limited power. Thus, the correlation between ECAR and
166 glycolytic gene family size may be driven mainly by the well-known WGD in the
167 Saccharomycetales (Fig. S2). Altogether, these results suggest that the *Saturnispora* species
168 with the rapid ECAR phenotype in the order Pichiales likely evolved independently from the
169 duplication of glycolytic genes observed in *S. cerevisiae*.


A
Tree scale: 1
ECAR
-0.06
-0.05
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
0.03
0.04
Order
Alaninales
Ascoideales
Dipodascales
Lipomycetales
Phaffomycetales
Pichiales
Saccharomycetales
Saccharomycodales
Serinales
Sporopachydermiales
Trigonopsidales
Legend
S. cerevisiae
Whole Genome Duplication
Saturnispora genus
Brettanomyces genus
B
R=-0.155, p=0.0073
ECAR (ΔpH/hr)
0.050
0.025
0.000
-0.025
-0.050
10
15
20
25
Total # of glycolytic genes
C
HXK
PGI
PFK
FBA
TPI
TDH
PGK
GPM
ENO
PYK
All yeasts
Saccharomycetales
Pichiales
−1
-0.4
0
0.4
1
S. cerevisiae gene name(s)
HXK HXK1, HXK2
PGI PGI1
PFK PFK1, PFK2
FBA FBA1
TPI TPI1
TDH TDH1, TDH2, TDH3
PGK PGK1
GPM GPM1, GPM2, GPM3
ENO ENO1, ENO2, ERR1, ERR2, ERR3
PYK PYK2, CDC19

170

171 **Figure 1. Variation of extracellular acidification rates (ECAR) across the yeast subphylum**
172 **is associated with glycolytic genes.** (A) The mean of the measured ECAR for each species is
173 visualized as a circle around the yeast phylogeny with the lighter green blocks corresponding to
174 more rapid ECAR. The model yeast *S. cerevisiae*, the whole genome duplication (WGD) in the
175 Saccharomycetales (arrow), the genus *Saturnispora* (box), and the genus *Brettanomyces*
176 (dashed box) are highlighted on the phylogeny. (B) A scatter plot showing the correlation
177 between ECAR and the total number of glycolytic genes where each dot represents a single
178 species. The correlation and p values are based on a phylogenetic generalized least squares
179 model, which accounts for correlation stemming from shared phylogenetic history. (C) Pearson
180 correlations between the number of copies of each glycolytic gene and the measured ECAR.
181 The gene names in boxes have retained duplicates in *S. cerevisiae*; solid boxes indicate
182 duplicates retained in all post-WGD yeasts, whereas the dashed boxes indicate duplicates
183 retained in some but not all post-WGD yeasts (Conant & Wolfe, 2007). Significance of the
184 Pearson correlation is indicated as * $p<0.05$, ** $p<0.01$, *** $p<0.005$. The yeast phylogeny was
185 pruned from Shen et al. (Shen *et al*, 2018).

186

187 **Divergent phenotypes of rapid and low ECAR species in *Saturnispora***

188 To interrogate the Pichiales-specific rapid ECAR phenotype, we sought to identify closely
189 related species with diverging phenotypes. Within the genus *Saturnispora*, we found that
190 several species consistently showed a rapid ECAR phenotype, namely *Sat. dispora, Sat.*
191 *hagleri, Sat. serradocipensis, Sat. zaruensis,* and *Sat. saitoi.* Conversely, *Sat. silvae* and *Sat.*
192 *mendoncae* consistently displayed a low ECAR phenotype (Fig. 2A). Moving forward, we
193 focused on four species, two rapid ECAR and the two low ECAR with the hypothesis that the
194 rapid ECAR species would have metabolic phenotypes more similar to known Crabtree-positive
195 yeasts. To confirm that the acidification reflected rapid glucose consumption, we measured
196 glucose depletion in the media in a typical growth experiment. The rapid ECAR species
197 consumed glucose more rapidly than the low ECAR species (Fig. 2B). The rapid ECAR species
198 even produced ethanol and acetate from overflow metabolism, despite being grown under
199 highly aerobic conditions in baffled flasks (Fig 2B). Accordingly, the rapid ECAR species also
200 grew to a lower final OD compared to the low ECAR species (Fig. 2C). These divergent
201 metabolic phenotypes show that these rapid ECAR yeasts are highly fermentative and indeed
202 exhibit the Crabtree/Warburg Effect under aerobic conditions, which prompted us to interrogate
203 potential genetic differences between the rapid ECAR and low ECAR species.

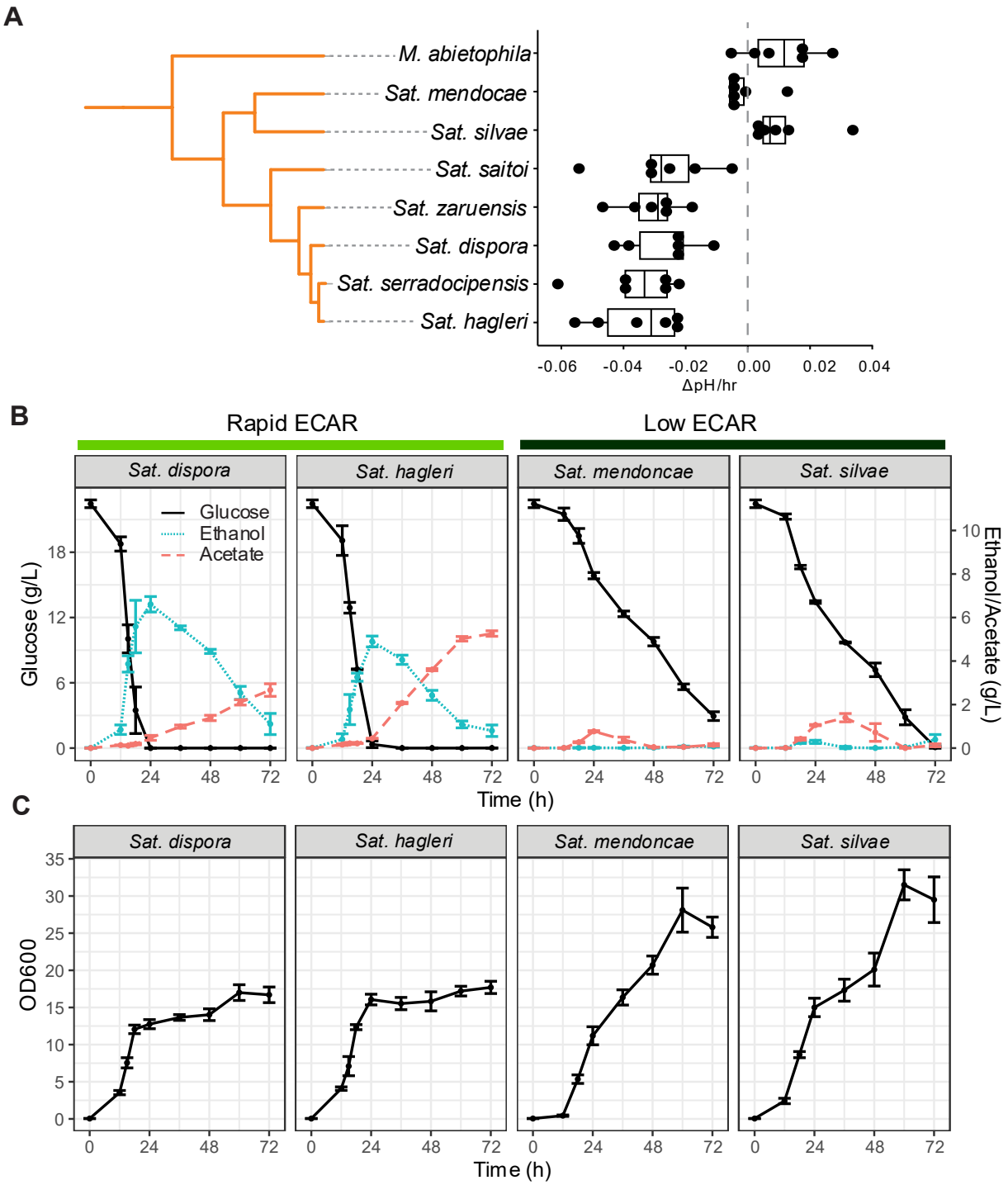
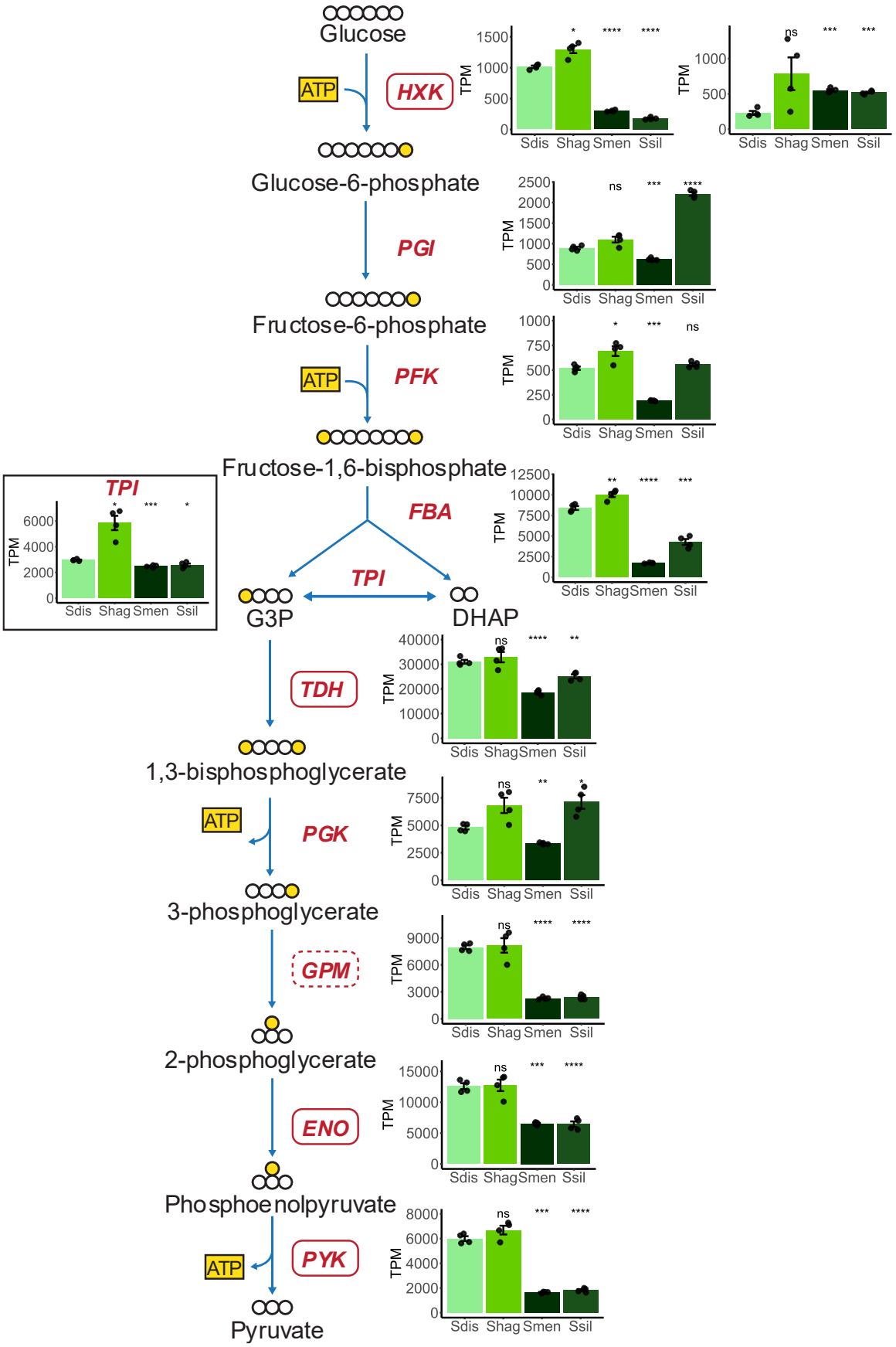


Figure 2. Species within the genus *Saturnispora* display divergent metabolic phenotypes. (A) ECAR measurements identified species with consistently low or consistently rapid ECAR within the genus *Saturnispora*. ECAR was also measured for *Martiniozyma abietophila* as a closely related outgroup. The (B) extracellular metabolites in the supernatant and (C) growth of two rapid ECAR and two low ECAR species were measured during growth in a 250-ml baffled flask, which was highly aerobic. The error bars represent the standard deviation of four biological replicates.

212 **Transcriptional profiling of rapid and low ECAR species**

213 Since there were no duplications in glycolytic genes in the rapid ECAR *Saturnispora*
214 species (Fig1C, Table S1), we used comparative transcriptomics to evaluate differences in gene
215 expression between the rapid and low ECAR species. To compare across species, we mapped
216 the reads to genes in each species, generated counts, summed counts across genes in the
217 same orthogroup, and calculated transcripts per million (TPM) for each orthogroup. Since our
218 primary interest was the glycolytic rate trait, we specifically retrieved the orthogroups encoding
219 central carbon metabolism genes. These data revealed that the rapid ECAR species had
220 significantly higher expression of several glycolytic genes, especially genes encoding
221 hexokinases and the enzymes of lower glycolysis (*TDH*, *GPM*, *ENO*, *PYK*) (Fig. 3). These
222 transcriptional differences are consistent with the rapid glucose consumption and aerobic
223 fermentation phenotype. Accordingly, we also found that the low ECAR species had higher
224 expression of genes encoding citric acid cycle enzymes, further supporting the conclusion that
225 the metabolisms of these closely related yeasts are highly divergent and favour aerobic
226 respiration (Fig. S3).


Glucose
ATP
HXK
Glucose-6-phosphate
PGI
Fructose-6-phosphate
ATP
PFK
Fructose-1,6-bisphosphate
FBA
TPI
G3P
DHAP
TDH
1,3-bisphosphoglycerate
ATP
PGK
3-phosphoglycerate
GPM
2-phosphoglycerate
ENO
Phosphoenolpyruvate
ATP
PYK
Pyruvate
TPM
Sdis
Shag
Smen
Ssil

227

228 **Figure 3. Divergent expression patterns of glycolytic genes between rapid ECAR and low**
229 **ECAR *Saturnispora* species.** The gene expression for orthologous glycolytic genes across two
230 rapid ECAR (*Sat. dispora* = Sdis, *Sat. hagleri* = Shag) and two low ECAR (*Sat. mendoncae* =
231 Smen, *Sat. silvae* = Ssil) are shown as transcripts per million (TPM). The bar height represents
232 the average of four biological replicates, each replicate is shown as a black dot, and the error
233 bars represent the standard error of the mean. The significance is relative to the TPM in *Sat.*
234 *dispora* (ns = not significant, * $p<0.05$, ** $p<0.01$, *** $p<0.005$, **** $p<0.001$). In the schematic of
235 glycolysis, carbon backbones are white, and phosphates are yellow. Boxes around gene names
236 are orthologous to genes that have retained duplicates in *S. cerevisiae* as in Fig. 1C, solid
237 boxes indicate duplicates retained in all post-WGD yeasts, whereas the dashed boxes indicate
238 duplicates retained in some but not all post-WGD yeasts (Conant & Wolfe, 2007).
239

240 **The Gal4p transcription factor contributes to rapid glycolytic rate in rapid ECAR**

241 ***Saturnispora***

242 Since we observed such a striking difference in the expression of glycolytic genes
243 between rapid and low ECAR yeasts, we hypothesized that this difference was due to genetic
244 differences in the regulatory regions of their glycolytic genes. To test this, we retrieved the
245 nucleotide sequences 1 kb upstream of each glycolytic gene as the putative promoter
246 sequences of these species, as well as a few other Pichiales for comparison. We predicted the
247 conserved transcription factor (TF) binding sites in these putative promoters to identify TFs that
248 had more binding sites in our rapid ECAR species compared to the low ECAR species. We
249 found that the promoters of some glycolytic genes had additional Gal4p-binding sites in the
250 rapid ECAR species (Fig 4A, Table S2). In rapid ECAR *Saturnispora*, there was an extra
251 predicted Gal4p-binding site in the promoter of *TDH*, the most highly expressed glycolytic gene.
252 There were also uniquely predicted Gal4p-binding sites in the promoters of *HXK* and *ENO* in the
253 rapid ECAR species, which were absent from the low ECAR species (Fig. 4A, Table S2).
254 Interestingly, when we compared these promoters to several other Pichiales species, we found
255 that there was considerable variation, suggesting that they may be quite variable across
256 species.
257 To investigate the importance of Gal4p in the rapid ECAR phenotype, we performed a
258 phylogenetic analysis and found that, despite some lineage-specific duplications in this

259 orthogroup, *GAL4* is likely a 1:1 ortholog in *S. cerevisiae*, *K. phaffii*, *Sat. dispora*, and most
260 yeasts (Fig. S4). We then developed a system to genetically modify a rapid ECAR yeast, *Sat.*
261 *dispora,* based on methods previously described for *K. phaffii* (De Schutter & Callewaert, 2012)
262 and generated a deletion mutant by replacing this *GAL4* ortholog with a *kanMX* or *natMX*
263 resistance cassette using long homology to direct homologous recombination. Two independent
264 deletion mutants lacking *GAL4* had decreased glucose consumption, ethanol production, and a
265 low ECAR phenotype (Fig. 4). We also found that mutants lacking *GAL4* had a more respiratory
266 phenotype with a higher oxygen consumption rate (Fig. 4E, Fig. S5), further supporting the
267 hypothesis that the Gal4p transcription factor is required for the rapid ECAR metabolic
268 phenotype and aerobic fermentation in *Sat. dispora*.

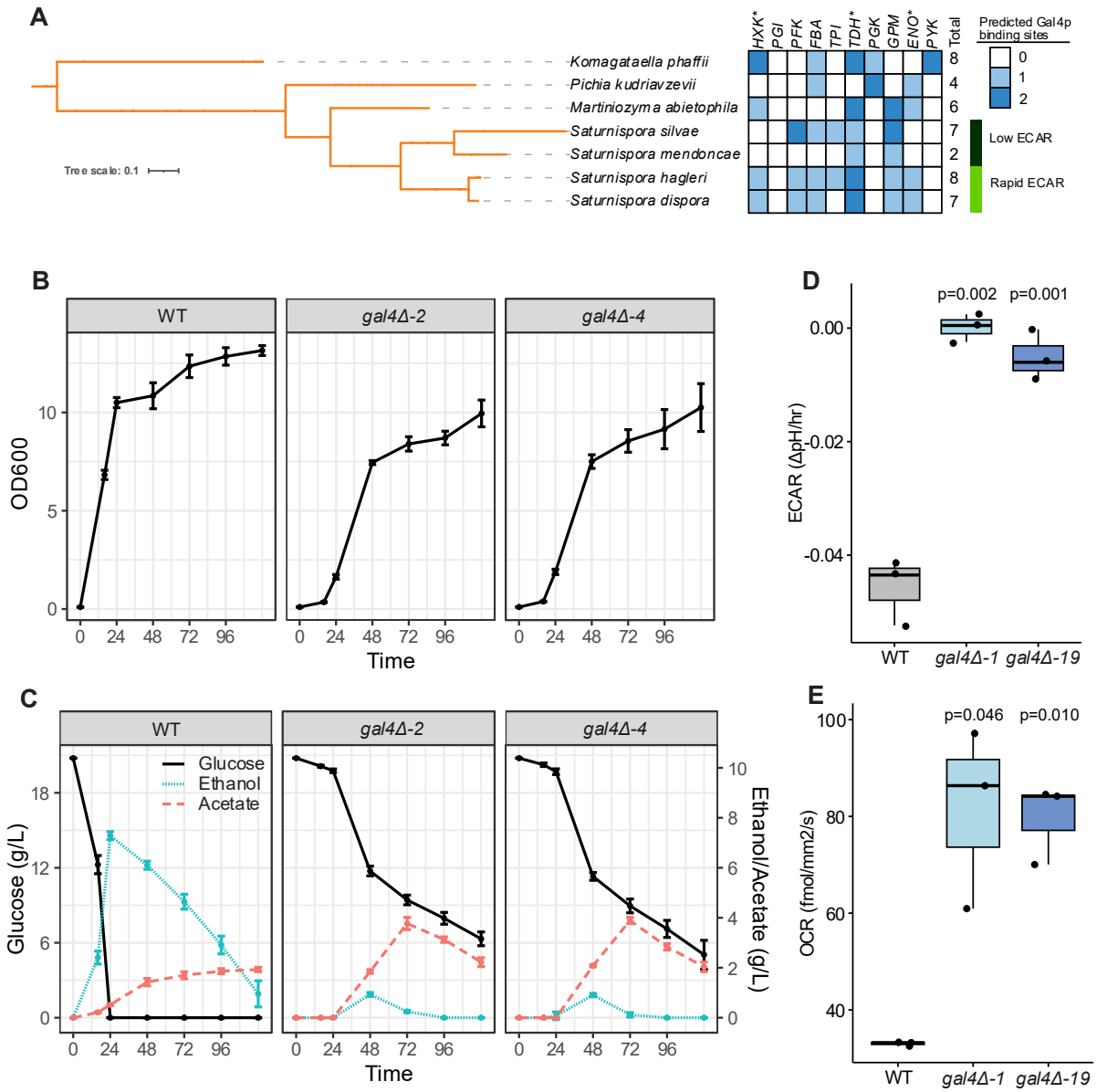


Figure 4. The Gal4p transcription factor contributes to the rapid ECAR metabolic phenotype in *Sat. dispora*. (A) The number of predicted Gal4p-binding sites in the putative promoters 1kb upstream of each glycolytic gene in several Pichiales species. The genes marked with an asterisk have an additional binding site in the rapid ECAR species. The (B) growth and (C) extracellular metabolites for the *Sat. dispora* wild-type (WT) control and two independent *gal4Δ* deletion mutants. Two high ECAR and two low ECAR species were measured during growth in a highly aerobic 250-ml baffled flask. The error bars represent the standard deviation of four biological replicates. (D) The *gal4Δ* deletion mutants showed a low ECAR phenotype compared to the parent (WT) control. (E) The *gal4Δ* deletion mutants displayed higher oxygen consumption rates (OCR) at 90 minutes post inoculation compared to the parent (WT) control. The centre lines of boxplots represent the median, the bounds of the boxes represent the interquartile range, the whiskers represent the spread of the data, and each dot represents one of three biological replicates. The p values shown are the results of two-sided t-tests compared to the wild-type control.

284 To test whether there was a direct cis-regulatory effect of Gal4p, we also made a GFP
285 reporter construct driven by the *Sat. dispora TDH* promoter. Furthermore, we made a version of
286 this reporter in which the predicted Gal4p-binding sites were mutated using site-directed
287 mutagenesis. We inserted these reporters into a genomic safe haven, in which there were no
288 predicted genes, in both the wild-type background and the *gal4Δ::natMX* deletion mutants. The
289 expression of GFP was significantly reduced in the wild-type background when the predicted
290 binding sites were mutated. The expression was also significantly reduced when the reporter
291 construct with intact Gal4p-binding sites was transformed into the *gal4Δ::natMX* mutant.
292 Together these results show that, in the absence of either the trans-acting transcription factor or
293 the cis-regulatory binding sites, the promoter activity is greatly reduced (Fig. 5). Although the
294 GFP expression was very low in the *gal4Δ::natMX* deletion background, there was a small but
295 significant difference between the wild-type and mutant reporter constructs, suggesting that the
296 mutations may also have interfered with the ability of other transcription factors to bind.
297 Altogether, this result directly shows that Gal4p is necessary to activate the most highly
298 expressed glycolytic gene.

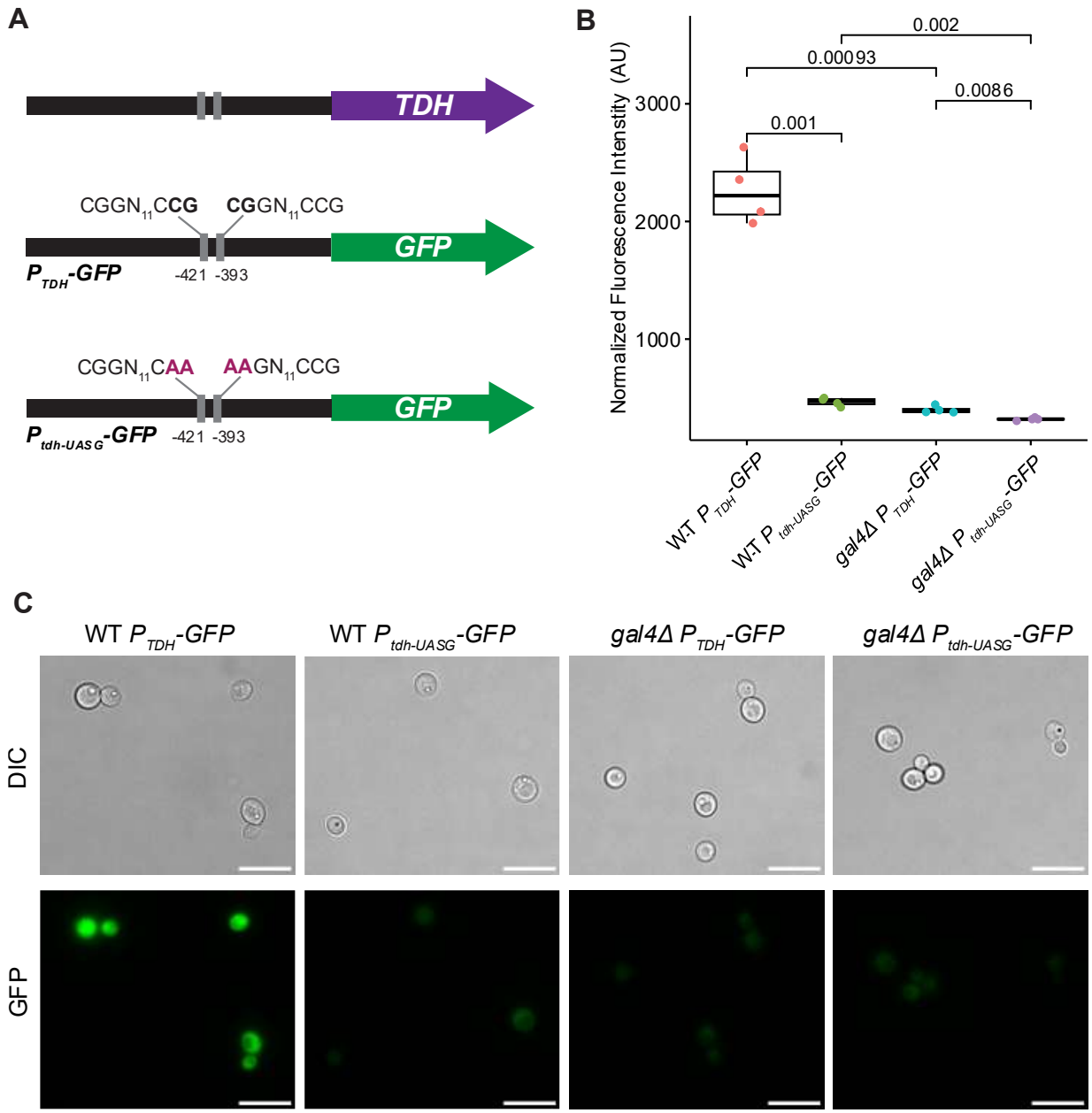


Figure 5. Gal4p is required for the induction of the *TDH* promoter. (A) A schematic diagram showing the location of the predicted Gal4p-binding sites in the *TDH* promoter, the reporter construct, and the mutations made in the predicted Gal4p-binding sites in the reporter construct promoter. (B) The fluorescence intensity of the wild-type and *gal4Δ* deletion mutants containing the reporter with the wild-type *TDH* promoter (P_{TDH}-*GFP*) or the mutated *TDH* promoter ($P_{tdh\text{-}UASG}$-*GFP*). The fluorescence of each reporter-expressing strain was normalized to its respective control strain (i.e. wild-type and *gal4Δ* mutants) lacking any reporter. The centre lines of boxplots represent the median, the bounds of the boxes represent the interquartile range, the whiskers represent the spread of the data, and each dot represents one of four biological replicates. The p values shown are the results of two-sided t-tests comparing the strains as indicated. (C) Representative images of the GFP fluorescence from the strains carrying the reporter constructs. The scale bar represents 10 µm. Uncropped images are available as Supplemental Data 1.

313 **DISCUSSION**

314 The discovery of shared or novel mechanisms to increase glycolytic rate can help us
315 understand the evolutionary constraints and guide manipulation of this trait across diverse yeast
316 species. In this study, we developed an assay to assess variation in glycolytic rate across the
317 yeast subphylum. Comparative genomics corroborated previous reports that duplications of
318 glycolytic genes were associated with increased glycolytic rates in yeasts, particularly in the
319 Saccharomycetales where a WGD occurred (Conant & Wolfe, 2007; Hagman & Piškur, 2015;
320 Dashko *et al*, 2014; Hagman *et al*, 2013; Lin & Li, 2011). However, there were no duplications or
321 expansions in glycolytic genes associated with the rapid ECAR phenotype in the Pichiales. This
322 contrast suggested that the rapid ECAR phenotype we observed in several *Saturnispora*
323 species had an independent mechanism. Using transcriptomics, we observed that several
324 glycolytic genes were upregulated in species with rapid ECAR. While these results were
325 consistent with the phenotypic observations, we were surprised to see how closely they
326 mirrored the gene duplications in *S. cerevisiae* (Conant & Wolfe, 2007). In particular, the genes
327 encoding enzymes in lower glycolysis, which were duplicated and retained in *S. cerevisiae*, had
328 much higher expression in the rapid ECAR species. The association of elevated expression of
329 lower glycolysis genes with aerobic fermentation is also consistent with previous evidence that
330 pull through lower glycolysis is key to fermentative capacity (Smits *et al*, 2000). Furthermore,
331 lower glycolysis, carries more flux than any biochemical alternative (Court *et al*, 2015),
332 suggesting that upregulating these existing genes may be the most effective strategy to
333 increase glycolytic rate.

334 Since the driving force between this phenotypic difference seemed to be at the
335 transcriptional level, we retrieved the putative promoter sequences upstream of glycolytic genes
336 to identify transcriptional activators that had more binding sites in the promoters in rapid ECAR
337 species. Several of the glycolytic gene promoters had more conserved binding sites for the
338 Gal4p transcription factor. Although Gal4p is well characterized for its role in activating the Leloir

339 pathway for galactose catabolism in *S. cerevisiae* (Johnston, 1987), it is also known to have
340 other roles in metabolism in different yeasts (Askew *et al*, 2009; Kim *et al*, 2024; Ata *et al*,
341 2018). Indeed, in another Pichiales yeast, the protein production chassis, *K. phaffii*,
342 overexpression of Gal4p has been shown to induce overflow metabolism (Ata *et al*, 2018). In
343 the case of *Sat. dispora,* this yeast natively has a rapid glycolytic rate and produces ethanol and
344 acetate as overflow metabolites. The promoters with additional Gal4p-binding sites in rapid
345 ECAR species compared to the low ECAR species in *Saturnispora* were upstream of *HXK,*
346 *TDH,* and *ENO.* This configuration again bore a striking parallel to modelling in *S. cerevisiae*
347 that predicted these three enzymes have a large impact on flux when their abundance
348 decreases (Conant & Wolfe, 2007). These results provide mechanistic insights into how the
349 transcriptional activation of glycolysis in *Saturnispora* evolved and suggest that the modelling
350 predicting the importance of these three key glycolytic enzymes extends to other yeasts beyond
351 *S. cerevisiae.*
352 To confirm the direct influence of Gal4p on the rapid ECAR phenotype, we developed a
353 transformation protocol for *Sat. dispora*. The deletion of *GAL4* decreases the glucose
354 consumption rate and renders the mutants more respiratory than fermentative. This result
355 confirms the importance of the Gal4p transcription factor in the rapid ECAR phenotype. To
356 ensure that *GAL4* deletion had a specific impact on the cis-regulation of glycolytic genes, we
357 developed a reporter system using the promoter of *TDH*, which was the most highly expressed
358 gene in both *Sat. dispora* and *Sat. hagleri*. Through monitoring the expression of GFP as a
359 reporter under the *TDH* promoter, we found that, if either the *GAL4* gene were missing or if the
360 Gal4p-binding sites were mutated, the GFP expression was greatly reduced, showing that both
361 the transcription factor Gal4p and its cis-regulatory binding sites are directly required for
362 induction. Although Gal4p is required for the high glucose consumption rate of *Sat. dispora*, we
363 can also infer that there are likely additional factors that contribute to the divergence in
364 phenotypes of the rapid and low ECAR species. For example, the *gal4Δ* mutants still

365 accumulated high levels of acetate and did not grow to the high biomass seen in the low ECAR
366 species.

367 The yeast phenotype of overflow metabolism is one that, on the surface, seems wasteful
368 as the organism produces fewer ATP molecules per glucose consumed. However, this strategy
369 has been proposed to deplete a limited resource and convert it into another product, which
370 would limit the growth of other microbes and could later be consumed. This “make-accumulate-
371 consume” phenotype has already been observed in two distinct yeast lineages (Rozpedowska
372 *et al*, 2011), and we now have evidence for a third independent lineage with a novel mechanism
373 driven by increased transcriptional activation of glycolysis. There are also interesting parallels in
374 the ecology of the rapid ECAR *Saturnispora* species and *S. cerevisiae*. *Sat. dispora* has
375 primarily been isolated from tree exudates, and *Sat. hagleri* has been isolated from fruits and
376 insects (Kurtzman *et al*, 2011; Harrison *et al*, 2024). Similar to *S. cerevisiae*, these yeasts are
377 isolated from environments that have occasional periods of rich sugar availability, which may
378 have prompted the evolution of rapid glycolysis (Liti, 2015; Jouhten *et al*, 2016). Although the
379 rapid ECAR phenotype of *Sat. dispora* arose from a distinct cis-regulatory mechanism, many of
380 the changes appear to lead to similar functional outcomes to those observed in *S. cerevisiae*.
381 Even though the precise molecular mechanism of evolution differed between *Sat. dispora* and
382 *S. cerevisiae*, the end result was still increased abundance of lower glycolytic enzymes,
383 glyceraldehyde-3-phosphate dehydrogenase, and hexokinase. Like the evolution of C4
384 photosynthesis in plants, the aerobic fermentation phenotype in yeasts represents another case
385 of convergent evolution in metabolic efficiency.

386 In the case of C4 plants, the abundance, activity, or localization of the same key C4
387 enzymes has been optimized through convergent evolution, whereas we find that, in aerobically
388 fermenting yeasts, the abundance of the same key glycolytic enzymes has been convergently
389 increased through gene duplications or through the recruitment of cis-regulatory elements. This
390 study provides new evidence supporting recent findings that trait evolution often employs the

391 same genes across divergent lineages of yeasts (David *et al*, 2025b; Aranguiz *et al*, 2025).
392 While previous studies have focused on gene presence and family size, we were able to identify
393 a mechanism of gene upregulation through the recruitment of cis-regulatory elements in the
394 promoters of congeneric species. These convergent adaptations suggest that there are likely
395 strong constraints on how a yeast can evolve a rapid glycolytic rate. Indeed, high glycolytic rates
396 and retained glycolytic duplicates may be a predictable occurrence, as suggested by multiple
397 recently discovered novel yeast WGDs (David *et al*, 2025a). Whether by gene duplication or cis-
398 regulatory evolution, it seems that an increase in the abundance of the same set of glycolytic
399 enzymes is required for yeasts to achieve high glycolytic rates. While organisms may deploy
400 divergent molecular mechanisms due to chance and the historical contingencies of their
401 lineages, these mounting examples suggest that evolutionary convergence operates within the
402 constraints of the pathway being manipulated by natural selection.
403

404 **METHODS**

405 **Strain and culture conditions**

406 All strains screened were previously described (Shen *et al*, 2018; Opulente *et al*, 2024)
407 and are listed in Table S1. Yeasts were routinely maintained on yeast extract peptone dextrose
408 (YPD) agar at room temperature. Fermentation and extracellular acidification assays were
409 conducted using synthetic complete (SCD) media containing yeast nitrogen base, 5 g/L
410 ammonium sulfate, complete dropout mix, and 20 g/L D-glucose (US Biologicals). All growth
411 conditions were at room temperature (RT, 22 °C) unless otherwise specified.

412 **Extracellular acidification rate assay**

413 All yeasts were grown for 24 hrs in 2 ml YPD, sub-cultured 1:4 in fresh SCD pH 6.8, and
414 grown for an additional 6 hrs. These actively growing cells were harvested by centrifugation,
415 washed twice in sterile water to remove excess media, and normalized by optical density. Cells
416 were inoculated into a 96-well plate at OD_{600nm} = 0.15 in SCD supplemented with 0.002%

417 phenol red. For each strain, 2% dextrose or a water control was added immediately before
418 beginning the assay. The absorbance at 560 nm was measured every 3 minutes for 90 minutes,
419 and the change in extracellular pH was calculated based on a standard curve (Fig. S1). The
420 dextrose-dependent acidification rate was determined as the difference between the
421 acidification rate in the presence of 2% dextrose and the acidification rate in the water control.
422 For each yeast species, three biological replicates were conducted on different days. The
423 average ECAR was plotted as a colour strip outside the phylogeny using iTOL v7 (Letunic &
424 Bork, 2024)

425 **Comparative genomics**

426 The orthogroups corresponding to genes encoding glycolytic enzymes were retrieved
427 from the published OrthoFinder analysis (Opulente *et al*, 2024). A correlation test between the
428 number of orthologs and the dextrose dependent acidification of each species was conducted
429 for each gene as well as the total sum of glycolytic genes. This analysis was also repeated for
430 the orders Saccharomycetales and Pichiales. To account for phylogeny, we used phylogenetic
431 generalized least squares (PGLS) using *R* v4.2.2 and the package caper v1.0.3 (Orme, 2013).
432 Data were visualized using ggpubr v0.6.0 (Kassambara, 2023) and corrplot v0.95 (Wei & Simko,
433 2024).

434 **Growth and metabolite profiling**

435 Yeast strains were grown overnight in YPD in biological quadruplicate. Overnight
436 cultures were collected, washed twice in sterile water, and inoculated in 50 ml fresh SCD pH 6.8
437 in 250-ml baffled flasks at a density of OD_{600nm} = 0.05. The flasks were incubated at RT while
438 shaking at 225 rpm. At each time point, OD_{600nm} was measured, and the supernatant was
439 analysed for glucose, ethanol, and acetate concentrations using high performance liquid
440 chromatography with a refractive index detector (HPLC-RID), as previously described
441 (Schwalbach *et al*, 2012). The mean and standard deviation across replicates was calculated in
442 *R* v4.2.2 and plots were made using the package ggplot2 v3.5.2 (Wickham, 2016).

443 **RNA extraction and sequencing**

444 RNA was collected from log-phase cells using acidic phenol as previously described for
445 yeast (Collart & Oliviero, 1993). Briefly, cells were collected and mixed with 0.1 volumes of 5%
446 acid phenol in 95% ethanol, harvested by centrifugation, and flash frozen. Cell pellets were
447 lysed with glass beads, and RNA was extracted with acidic phenol twice, followed by a final
448 chloroform extraction. The RNA was treated with RQ1 DNase (Promega) and subsequently
449 cleaned up using an RNeasy column (Qiagen). Since the rapid and low ECAR yeasts grew at
450 different rates, to obtain mid-log phase for all strains, the rapid ECAR yeasts were collected at
451 15 hrs post-inoculation, and the low ECAR yeasts were collected at 18 hrs post-inoculation.

452 We enriched for mRNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module
453 (NEB), and libraries were prepared using the NEBNext Ultra II Directional RNA Library Kit for
454 Illumina (NEB) following the manufacturer's instructions. Paired-end libraries were sequenced
455 on a NovaSeq X Plus (Illumina) with 2x150 bp reads. Reads were trimmed using Trimmomatic
456 v0.30 (Bolger *et al*, 2014) to remove adapters and subsequently checked in FastQC (Andrews
457 S., 2010). Reads were mapped to the respective genome of each species using HiSat v2.1.0
458 (Kim *et al*, 2019). HTSeq-count v2.0.3 (Putri *et al*, 2022) was used to generate the counts of
459 annotated genes using the non-unique fraction setting to avoid overcounting or discarding highly
460 similar genes. To compare the transcriptome between species, we used OrthoFinder (Emms &
461 Kelly, 2019) to identify groups of orthologous genes within the *Saturnispora* species.
462 Subsequently, the sum of counts and gene lengths for each orthogroup was calculated using a
463 custom python script (https://github.com/Linda-Horianopoulos/Aerobic_fermentation_scripts).
464 The counts and total gene lengths were then used to determine the TPM for each orthogroup in
465 each species using EdgeR (Robinson *et al*, 2009). The orthogroups encoding glycolytic and
466 citric acid cycle enzymes were retrieved, and plots were made to visualize the expression in
467 each species in *R* v4.2.2 using ggplot2 v3.5.2 (Wickham, 2016).

468 **Identification of putative transcription factor binding sites**

469 Putative promoter sequences for each of the glycolytic genes were retrieved from the
470 annotation files using gffread v0.12.7 (Pertea & Pertea, 2020). Yeastract (Teixeira *et al*, 2023)
471 was used to identify all conserved *S. cerevisiae* transcription factor binding sites in the 1000 bp
472 upstream of the coding sequence of each glycolytic gene from *Sat. dispora, Sat. hagleri, Sat.*
473 *mendoncae, Sat. silvae*, *Martiniozyma abietophilia*, *P. kudriavzevii*, and *K. phaffii.* The number of
474 predicted Gal4p-binding sites (exact consensus matches to $CGGN_{11}CCG$ (Johnston, 1987)) in
475 the promoter of each gene for each species was visualized using iTOL v7 (Letunic & Bork,
476 2024).

477 **Phylogenetic analysis of *GAL4* orthologs**

478 The genes in the same orthogroup as the canonical *S. cerevisiae GAL4* (YPL248C)
479 were retrieved for all species included in our ECAR screen. The translated amino acid
480 sequences were retrieved, aligned using MAFFT v7.222 (Katoh & Standley, 2013), and the
481 alignment was trimmed using trimal v1.4.1 applying the gappyout option (Capella-Gutiérrez *et*
482 *al*, 2009). The phylogeny was inferred using IQtree v1.6.8 with automatic substitution model
483 selection and 1000 bootstraps (Nguyen *et al*, 2015). The phylogeny was visualized in iTOL v7
484 (Letunic & Bork, 2024) and manually pruned to remove an outgroup on a disproportionately long
485 internal branch, which encode divergent C6 zinc transcription factors.

486 **Gene deletion**

487 The coding sequence of *GAL4* in *Sat. dispora* was replaced with a *kanMX* marker using
488 an electroporation protocol adapted from *K. phaffii* (De Schutter & Callewaert, 2012). Briefly, ~1-
489 kb regions upstream and downstream of the *GAL4* coding sequence were amplified from
490 genomic DNA using primers with homology to the *kanMX* marker and the pRS426 backbone
491 (Table S3). The *kanMX* marker was amplified from pUG6 (Goldstein & McCusker, 1999) and the
492 pRS426 backbone (Christianson *et al*, 1992) was also amplified using primers specified in Table
493 S3. These fragments were assembled using HiFi assembly (NEB) and transformed into *E. cloni*
494 (Lucigen). The full construct was subsequently amplified from the plasmid, and 1 µg was

495 transformed into *Sat. dispora*. Electroporation was performed in 80 µL of ice cold 1 M sorbitol
496 using 1.5 kV. After electroporation, cells were recovered in 1 mL of YPD + 1 M sorbitol and
497 plated on YPD plates supplemented with 1 M sorbitol and 300 µg/ml G418. Transformants were
498 screened for stable resistance against G418, and correct replacement of the coding sequence
499 was confirmed with colony PCR. To allow the *kanMX* marker to be used for additional
500 modifications, the *kanMX* cassette was replaced with a *natMX* cassette, which was amplified
501 from pAG25, using the homology of the promoter and terminator. For selection of colonies
502 integrating the *natMX* marker, colonies were plated on YPD plates supplemented with 1 M
503 sorbitol and 25 µg/ml nourseothricin. To ensure that the phenotype was not due to any spurious
504 background mutations, two independent colonies were used in all experiments. The strains
505 constructed for this study are listed in Table S4.

506 **Oxygen consumption measurements**

507 The Resipher system (Lucid Scientific) was used to monitor the oxygen levels and
508 calculate the oxygen consumption rates (OCR). Three biological replicates of overnight cultures
509 were washed twice in sterile water and normalized to OD600 = 0.1 in 95 µl in a 96-well plate in
510 SC medium lacking a carbon source. The plate was centrifuged at 300 *x g* for 2 minutes to
511 pellet the cells to the bottom of the plate, and immediately before beginning reading, 5 ul of 40%
512 dextrose was added for a final concentration of 2% dextrose. The plate was incubated in the
513 dark at room temperature while the oxygen concentration was measured. The oxygen
514 consumption rates were retrieved at the 90-minute timepoint for consistency with the ECAR
515 assays. The data was visualized in *R* using ggpubr v0.6.0 (Kassambara, 2023) and the
516 significance of differences between strains were determined using a two-sided t-test.

517 **Evaluating the cis-regulatory elements in the promoter of *TDH***

518 An insertion construct was designed in a genomic safe haven to allow heterologous
519 proteins to be expressed from a region of the genome with no risk of gene disruption. A plasmid
520 containing ~1kb regions to direct homologous recombination to the genomic safe haven site

521 flanking the sequence for GFP driven by the *TDH* promoter from *Sat. dispora* was synthesized
522 by the Joint Genome Institute (JGI). Since this promoter had two predicted Gal4p-binding sites,
523 we mutated them both using site directed mutagenesis as previously described (Zheng *et al*,
524 2004) with the modification that Phusion polymerase was used to amplify the construct. Binding
525 site mutants were confirmed by whole plasmid sequencing (Plasmidsaurus). Both versions of
526 the reporter construct were transformed into the *Sat. dispora* wild-type strain and *gal4Δ::natMX*
527 deletion mutant using electroporation. Transformants were screened using colony PCR. Four
528 biological replicates of each strain, including controls lacking any reporter were grown overnight
529 in YPD, washed twice in sterile water, and inoculated in fresh SC+D pH 6.8 at OD_{600nm} = 0.1.
530 After growing for 40 hrs, cells were normalized to OD_{600nm} = 1 and the fluorescence intensity
531 was measured using a BMG Labtech FLUOstar plate reader. The background from control
532 strains without a reporter was subtracted and the fluorescence intensity was plotted in *R* using
533 the package ggpubr v0.6.0 (Kassambara, 2023). The statistical significance was determined
534 using two-sided t-tests.
535 Images of cells expressing GFP in the WT and *gal4Δ* background with the intact and
536 mutated promoter grown under the same conditions (SC+D pH 6.8 for 40 hours) were captured
537 with an EVOS FL microscope with 100x oil immersion plan apochromat objective lens. Images
538 were uniformly processed and cropped in FIJI (Schindelin *et al*, 2012). Unedited images are
539 available as Supplemental Data 1.
540
541 **Data availability**
542 The RNA-Sequencing data has been deposited to NCBI's GEO and the accession number will
543 be available upon publication. All other data are provided as supplemental tables. Custom code
544 is available on Github (https://github.com/Linda-Horianopoulos/Aerobic_fermentation_scripts) or
545 as a release on Zenodo (https://doi.org/10.5281/zenodo.17185390).
546

547 **Acknowledgements**

548 The authors thank members of the Hittinger Lab and Y1000+ Project (http://y1000plus.org) team
549 members for helpful discussions. This project was supported by the National Science
550 Foundation under Grants No. DEB-2110403 (to C.T.H.); DEB-2110404 (to A.R.); in part by the
551 Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science,
552 Biological and Environmental Research Program under Award Number DE-SC0018409 (of
553 which C.T.H. is a co-investigator); and the National Institute of Food and Agriculture, United
554 States Department of Agriculture, Hatch project 7005101 (to C.T.H.). Research in A.R.'s lab is
555 also supported by the National Institutes of Health/National Institute of Allergy and Infectious
556 Diseases (R01 AI153356). The work (proposal: 10.46936/10.25585/60008615) conducted by
557 the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE
558 Office of Science User Facility, is supported by the Office of Science of the U.S. Department of
559 Energy operated under Contract No. DE-AC02-05CH11231. We specifically thank Ian Blaby for
560 helping design and Robert Evans for synthesizing the safe haven insertion plasmid as part of
561 Yasuo Yoshikuni's team. L.C.H. was supported by a Natural Sciences and Engineering
562 Research Council of Canada (NSERC) postdoctoral fellowship.

563 **Author contributions**

564 L.C.H. and C.T.H. conceptualized this project. L.C.H. developed methodology, performed
565 experiments, and analysed data with the input and guidance of A.R. and C.T.H. L.C.H. and
566 C.T.H. wrote the manuscript, and A.R. reviewed, edited, and approved the manuscript. L.C.H.,
567 C.T.H., and A.R. acquired funding for this project.

568 **Conflict of Interest**

569 A.R. is a scientific consultant of LifeMine Therapeutics, Inc. All other authors declare no
570 competing interests.

571

572

573 **REFERENCES**

574 Andrews S. (2010) FastQC: a quality control tool for high throughput sequence data.
575 *http://www.bioinformatics.babraham.ac.uk/projects/fastqc*

576 Aranguiz K, Horianopoulos LC, Elkin L, Abá KS, Jordahl D, Overmyer KA, Wrobel RL, Coon JJ,
577 Shiu SH, Rokas A, *et al* (2025) Machine learning reveals genes impacting oxidative stress
578 resistance across yeasts. *Nature Communications 2025 16:1* 16: 1–15

579 Askew C, Sellam A, Epp E, Hogues H, Mullick A, Nantel A & Whiteway M (2009) Transcriptional
580 Regulation of Carbohydrate Metabolism in the Human Pathogen *Candida albicans*. *PLoS*
581 *Pathog* 5: e1000612

582 Ata Ö, Rebnegger C, Tatto NE, Valli M, Mairinger T, Hann S, Steiger MG, Çalık P & Mattanovich
583 D (2018) A single Gal4-like transcription factor activates the Crabtree effect in
584 *Komagataella phaffii*. *Nat Commun* 9: 1–10

585 Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence
586 data. *Bioinformatics* 30: 2114

587 Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K & Hibberd JM (2011)
588 Independent and parallel recruitment of preexisting mechanisms underlying C4
589 photosynthesis. *Science (1979)* 331

590 Capella-Gutiérrez S, Silla-Martínez JM & Gabaldón T (2009) trimAl: a tool for automated
591 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972

592 Christianson TW, Sikorski RS, Dante M, Shero JH & Hieter P (1992) Multifunctional yeast high-
593 copy-number shuttle vectors. *Gene* 110: 119–122

594 Collart MA & Oliviero S (1993) Preparation of Yeast RNA. *Curr Protoc Mol Biol* 23: 13.12.1-
595 13.12.5

596 Conant GC & Wolfe KH (2007) Increased glycolytic flux as an outcome of whole-genome
597 duplication in yeast. *Mol Syst Biol* 3

598 Conant GC & Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new
599 functions. *Nature Reviews Genetics 2008 9:12* 9: 938–950

600 Court SJ, Waclaw B & Allen RJ (2015) Lower glycolysis carries a higher flux than any
601 biochemically possible alternative. *Nature Communications 2015 6:1* 6: 1–8

602 Dashko S, Zhou N, Compagno C & Piškur J (2014) Why, when, and how did yeast evolve
603 alcoholic fermentation? *FEMS Yeast Res* 14: 826–832

604 David KT, Horianopoulos L, Gonçalves C, Steenwyk JL, Pontes A, Gonçalves P, Hittinger CT,
605 Pennell M & Rokas A (2025a) Discovery of additional ancient genome duplications in
606 yeasts. *bioRxiv*: 2025.08.31.673279

607 David KT, Schraiber JG, Crandall JG, Labella AL, Opulente DA, Harrison MC, Wolters JF, Zhou
608 X, Shen XX, Groenewald M, *et al* (2025b) Convergent expansions of keystone gene
609 families drive metabolic innovation in Saccharomycotina yeasts. *Proc Natl Acad Sci U S A*
610 122: e2500165122

611 Emms DM & Kelly S (2019) OrthoFinder: Phylogenetic orthology inference for comparative
612 genomics. *Genome Biol* 20

613 Escalera-Fanjul X, Quezada H, Riego-Ruiz L & González A (2019) Whole-Genome Duplication
614 and Yeast's Fruitful Way of Life. *Trends in Genetics* 35

615 Gasser B & Mattanovich D (2018) A yeast for all seasons – Is *Pichia pastoris* a suitable chassis
616 organism for future bioproduction? *FEMS Microbiol Lett* 365: 181

617 Geijer C, Ledesma-Amaro R & Tomas-Pejo E (2022) Unraveling the potential of non-
618 conventional yeasts in biotechnology. *FEMS Yeast Res* 22

619 Goldstein AL & McCusker JH (1999) Three new dominant drug resistance cassettes for gene
620 disruption in *Saccharomyces cerevisiae*. *Yeast* 15

621 Hagman A & Piškur J (2015) A Study on the Fundamental Mechanism and the Evolutionary
622 Driving Forces behind Aerobic Fermentation in Yeast. *PLoS One* 10: e0116942

623 Hagman A, Säll T, Compagno C & Piskur J (2013) Yeast '"Make-Accumulate-Consume"' Life
624 Strategy Evolved as a Multi-Step Process That Predates the Whole Genome Duplication.

625 Harrison MC, LaBella AL, Hittinger CT & Rokas A (2022) The evolution of the GALactose
626 utilization pathway in budding yeasts. *Trends in Genetics* 38: 97–106

627 Harrison MC, Opulente DA, Wolters JF, Shen XX, Zhou X, Groenewald M, Hittinger CT, Rokas A
628 & LaBella AL (2024) Exploring Saccharomycotina Yeast Ecology Through an Ecological
629 Ontology Framework. *Yeast* 41: 615–628

630 Heistinger L, Gasser B & Mattanovich D (2020) Microbe Profile: *Komagataella phaffii*: a
631 methanol devouring biotech yeast formerly known as *Pichia pastoris*. *Microbiology (N Y)*
632 166: 614–616

633 Johnston M (1987) A model fungal gene regulatory mechanism: the *GAL* genes of
634 *Saccharomyces cerevisiae*. *Microbiol Rev* 51

635 Jouhten P, Ponomarova O, Gonzalez R & Patil KR (2016) *Saccharomyces cerevisiae*
636 metabolism in ecological context. *FEMS Yeast Res* 16: 80

637 Kassambara A (2023) ggpubr: 'ggplot2' Based Publication Ready Plots. *R package version 060*

638 Katoh K & Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7:
639 Improvements in Performance and Usability. *Mol Biol Evol* 30: 772–780

640 Kim D, Paggi JM, Park C, Bennett C & Salzberg SL (2019) Graph-based genome alignment and
641 genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37: 907–915

642 Kim M-J, White AM & Mitchell AP (2024) Strain variation in *Candida albicans* glycolytic gene
643 regulation. *mSphere* 9

644 Kurtzman CP, Fell JW & Boekhout T (2011) The Yeasts: A Taxonomic Study, Volume 1 Elsevier

645 Letunic I & Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic
646 tree display and annotation tool. *Nucleic Acids Res* 52

647 Lin Z & Li WH (2011) Expansion of hexose transporter genes was associated with the evolution
648 of aerobic fermentation in yeasts. *Mol Biol Evol* 28: 131–142

649 Liti G (2015) The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *Elife* 4

650 Lyu H, Yim WC & Yu Q (2024) Genomic and Transcriptomic Insights into the Evolution of C4
651 Photosynthesis in Grasses. *Genome Biol Evol* 16

652 Lyu MJA, Du H, Yao H, Zhang Z, Chen G, Huang Y, Ni X, Chen F, Zhao YY, Tang Q, *et al* (2025)
653 A dominant role of transcriptional regulation during the evolution of C4 photosynthesis in
654 *Flaveria* species. *Nature Communications 2025 16:1* 16: 1–17

655 Opekarova M & Sigler K (1982) Acidification power: indicator of metabolic activity and autolytic
656 changes in *Saccharomyces cerevisiae*. *Folia Microbiol (Praha)* 27: 395–403

657 Nguyen LT, Schmidt HA, Von Haeseler A & Minh BQ (2015) IQ-TREE: A Fast and Effective
658 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 32:
659 268–274

660 Opulente DA, LaBella AL, Harrison MC, Wolters JF, Liu C, Li Y, Kominek J, Steenwyk JL,
661 Stoneman HR, VanDenAvond J, *et al* (2024) Genomic factors shape carbon and nitrogen
662 metabolic niche breadth across Saccharomycotina yeasts. *Science* 384: eadj4503

663 Orme D (2013) The caper package : comparative analysis of phylogenetics and evolution in R.
664 *R package version 05, 2*

665 Patra P, Das M, Kundu P & Ghosh A (2021) Recent advances in systems and synthetic biology
666 approaches for developing novel cell-factories in non-conventional yeasts. *Biotechnol Adv*
667 47 doi:10.1016/j.biotechadv.2021.107695

668 Pertea G & Pertea M (2020) GFF Utilities: GffRead and GffCompare. *F1000Res* 9

669 Putri GH, Anders S, Pyl PT, Pimanda JE & Zanini F (2022) Analysing high-throughput
670 sequencing data in Python with HTSeq 2.0. *Bioinformatics* 38: 2943–2945

671 Pyne ME, Bagley JA, Narcross L, Kevvai K, Exley K, Davies M, Wang Q, Whiteway M & Martin
672 VJJ (2023) Screening non-conventional yeasts for acid tolerance and engineering *Pichia*
673 *occidentalis* for production of muconic acid. *Nature Communications 2023 14:1* 14: 1–13

674 Radecka D, Mukherjee V, Mateo RQ, Stojiljkovic M, Foulquié-Moreno MR & Thevelein JM
675 (2015) Looking beyond *Saccharomyces*: The potential of non-conventional yeast species
676 for desirable traits in bioethanol fermentation. *FEMS Yeast Res* 15
677 doi:10.1093/femsyr/fov053

678 Robinson MD, McCarthy DJ & Smyth GK (2009) edgeR: a Bioconductor package for differential
679 expression analysis of digital gene expression data. *Bioinformatics* 26: 139

680 Rozpedowska E, Hellborg L, Ishchuk OP, Orhan F, Galafassi S, Merico A, Woolfit M, Compagno
681 C & Piškur J (2011) Parallel evolution of the make–accumulate–consume strategy in
682 *Saccharomyces* and *Dekkera* yeasts. *Nature Communications 2011 2:1* 2: 1–7

683 Sage RF, Christin PA & Edwards EJ (2011) The C4 plant lineages of planet Earth. *J Exp Bot* 62:
684 3155–3169

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al* (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9: 676–682

Schmidt CA, Fisher-Wellman KH & Darrell Neufer P (2021) From OCR and ECAR to energy: Perspectives on the design and interpretation of bioenergetics studies. *Journal of Biological Chemistry* 297: 101140–101141

De Schutter K & Callewaert N (2012) *Pichia* surface display: A tool for screening single domain antibodies. *Methods in Molecular Biology* 911: 125–134

Schwalbach MS, Keating DH, Tremaine M, Marner WD, Zhang Y, Bothfeld W, Higbee A, Grass JA, Cotten C, Reed JL, *et al* (2012) Complex physiology and compound stress responses during fermentation of alkali-pretreated corn stover hydrolysate by an *Escherichia coli* ethanologen. *Appl Environ Microbiol* 78: 3442–3457

Shen XX, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh K V., Haase MAB, Wisecaver JH, Wang M, Doering DT, *et al* (2018) Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* 175: 1533-1545.e20

Sigler K (2013) Acidification power (AP) test and similar methods for assessment and prediction of fermentation activity of industrial microorganisms. *Kvasny Prumysl* 59: 204–208

Sigler K, Knotková A & Kotyk A (1981) Factors governing substrate-induced generation and extrusion of protons in the yeast *Saccharomyces cerevisiae*. *BBA - Biomembranes* 643: 572–582

Smits HP, Hauf J, Müller S, Hobley TJ, Zimmermann FK, Hahn-Hägerdal B, Nielsen J & Olsson L (2000) Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of *Saccharomyces cerevisiae*. *Yeast* 16

Suthers PF & Maranas CD (2022) Examining organic acid production potential and growth-coupled strategies in *Issatchenkia orientalis* using constraint-based modeling. *Biotechnol Prog* 38

Swift J, Luginbuehl LH, Hua L, Schreier TB, Donald RM, Stanley S, Wang N, Lee TA, Nery JR, Ecker JR, *et al* (2024) Exaptation of ancestral cell-identity networks enables C4 photosynthesis. *Nature 2024 636:8041* 636: 143–150

Tan SI, Liu Z, Tran VG, Martin TA & Zhao H (2025) *Issatchenkia orientalis* as a platform organism for cost-effective production of organic acids. *Metab Eng* 89: 12–21

Teixeira MC, Viana R, Palma M, Oliveira J, Galocha M, Mota MN, Couceiro D, Pereira MG, Antunes M, Costa I V., *et al* (2023) YEASTRACT+: a portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis. *Nucleic Acids Res* 51: D785–D791

Tucey TM, Verma J, Harrison PF, Snelgrove SL, Lo TL, Scherer AK, Barugahare AA, Powell DR, Wheeler RT, Hickey MJ, *et al* (2018) Glucose Homeostasis Is Important for Immune Cell Viability during *Candida* Challenge and Host Survival of Systemic Fungal Infection. *Cell Metab* 27: 988-1006.e7

Wang Y, Wang Y, Cui J, Wu C, Yu B & Wang L (2025) Non-conventional yeasts: promising cell factories for organic acid bioproduction. *Trends Biotechnol* 43: 1609–1626

Wei T & Simko V (2024) R package ‘corrplot’: Visualization of a Correlation Matrix. *(Version 095)*

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Williams BP, Aubry S & Hibberd JM (2012) Molecular evolution of genes recruited into C4 photosynthesis. *Trends Plant Sci* 17: 213–220

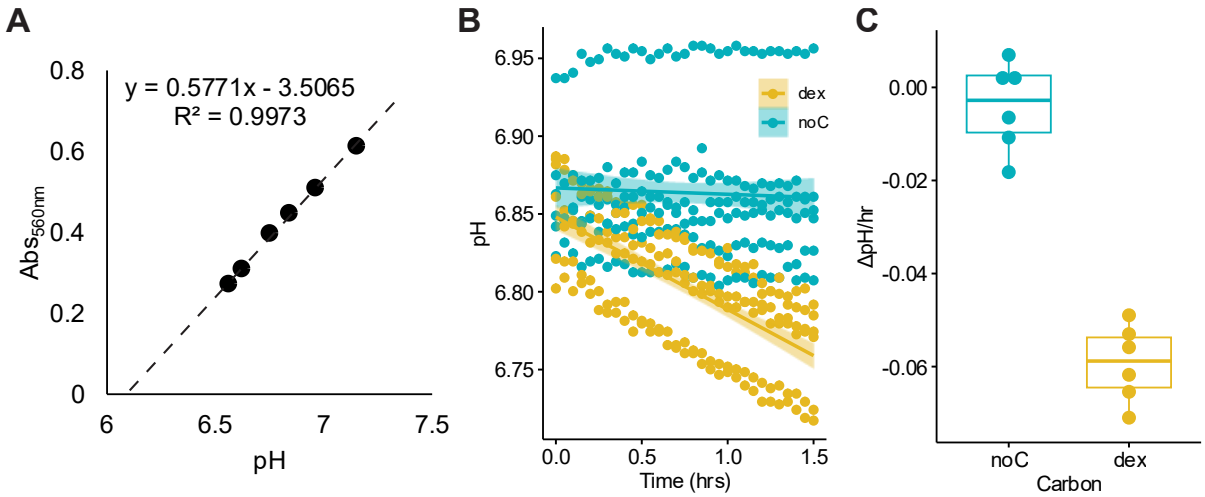
Yamashoji S, Al Mamun A & Bari L (2020) Visual and simple determination of glucose-induced acidification by yeast cells: application to rapid cytotoxicity test. *Heliyon* 6: e03924

Yoo K, Bhattacharya S, Oliveira NK, Pereira de Sa N, Matos GS, Del Poeta M & Fries BC (2024) With age comes resilience: how mitochondrial modulation drives age-associated fluconazole tolerance in *Cryptococcus neoformans*. *mBio* 15

Zhang J & Zhang Q (2019) Using Seahorse machine to measure OCR and ECAR in cancer cells. *Methods in Molecular Biology* 1928: 353–363

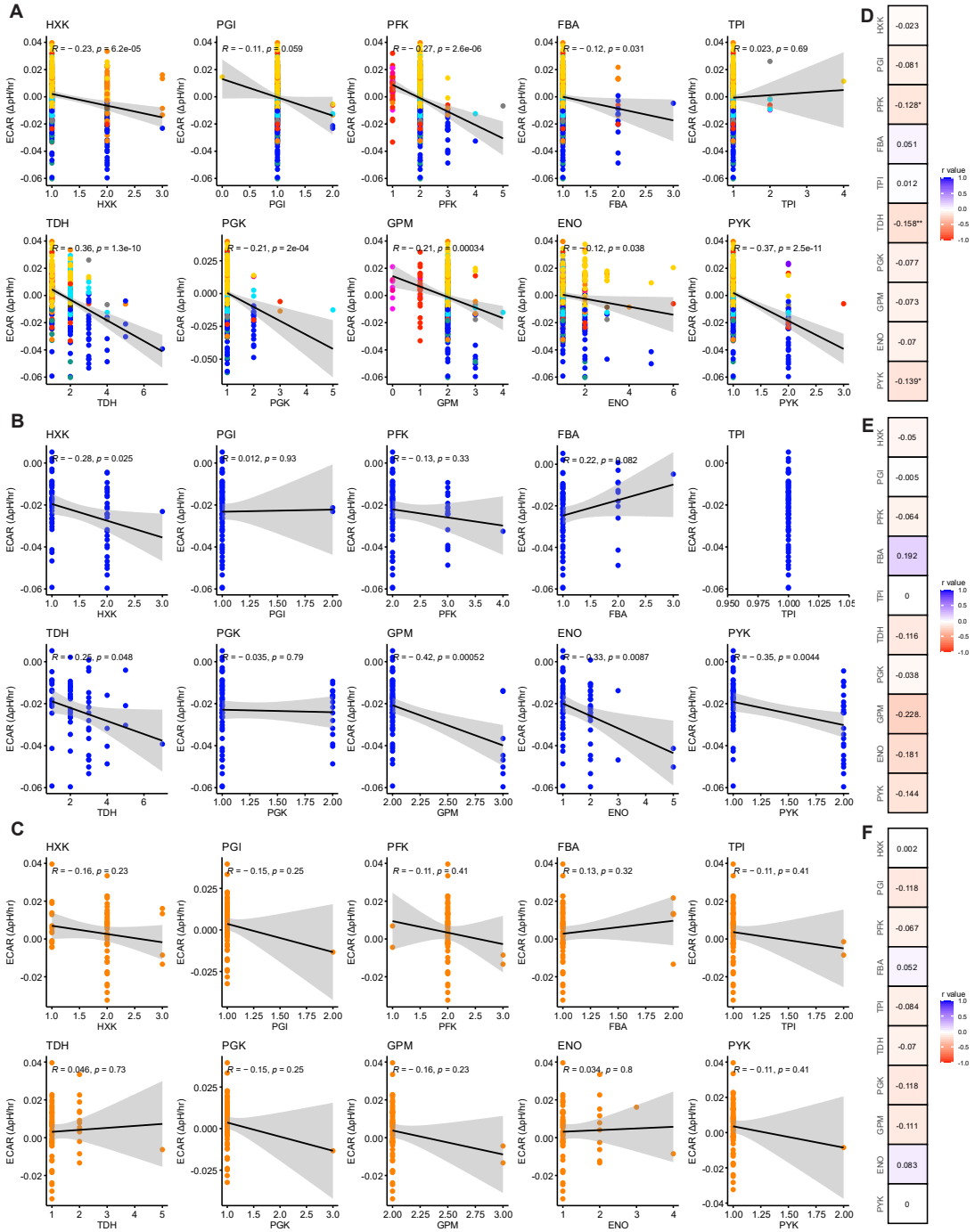
Zhang P, Li H, Cheng J, Sun AY, Wang L, Mirchevska G, Calderone R & Li D (2018) Respiratory stress in mitochondrial electron transport chain complex mutants of *Candida albicans* activates Snf1 kinase response. *Fungal Genetics and Biology* 111: 73–84

Zheng L, Baumann U & Reymond JL (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* 32: e115–e115

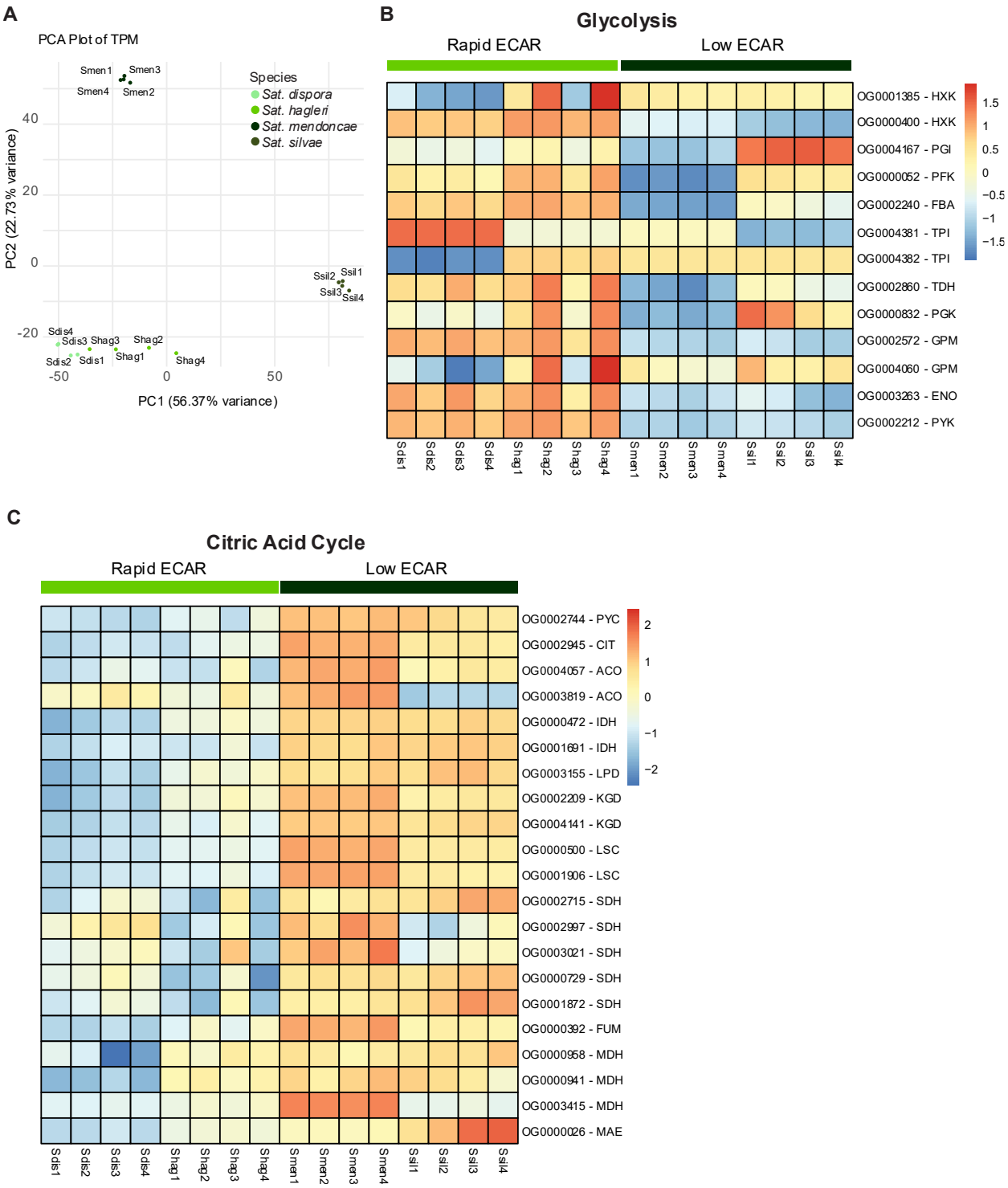


749

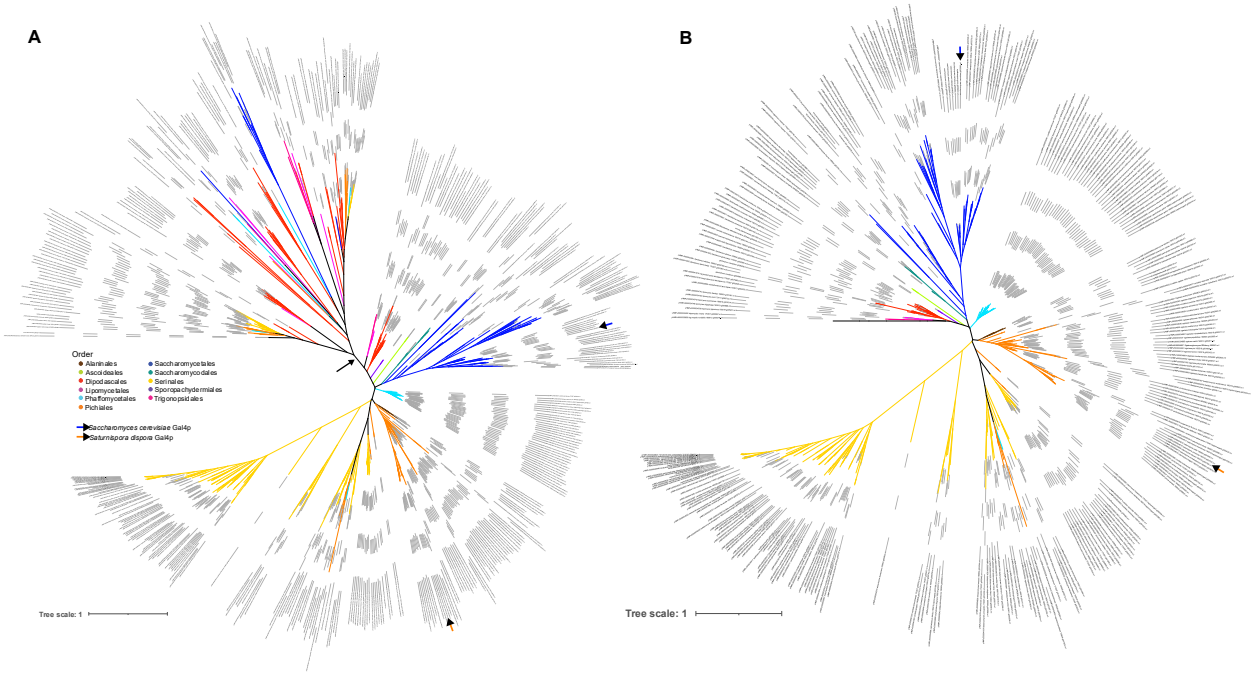
750 **Supplemental Figure 1.Optimization of the Extracellular Acidification Rate (ECAR) assay.**
751 (A) A standard curve of the absorbance at 560nm generated from media between pH 6.5 and
752 7.5. (B) Optimization assay using *S. cerevisiae* showing that the control with no carbon source
753 (noC) did not acidify the media, whereas the cells with 2% glucose added (dex) consistently
754 acidified the media across six biological replicates. (C) The slopes calculated from the six
755 biological replicates showing the dextrose-dependent pH change.
756



Supplemental Figure 2. The gene family size of glycolytic genes correlates with ECAR across all yeasts. (A-C) The correlation between ECAR and the number of homologs of each glycolytic gene showing the Pearson's correlation and p-value for (A) all phenotyped yeasts, (B) the yeasts in the order Saccharomycetales, and (C) the yeasts in the order Pichiales. (D-F) The phylogenetically corrected correlations between ECAR and the number of homologs of each glycolytic gene for (D) all phenotyped yeasts, (E) the yeasts in the order Saccharomycetales, and (F) the yeasts in the order Pichiales. The R-value from the phylogenetic generalized least squares is in the box and the p-value is indicated as * p<0.05 and ** p<0.01. Data points are colour coded according to taxonomic order as in Figure 1.

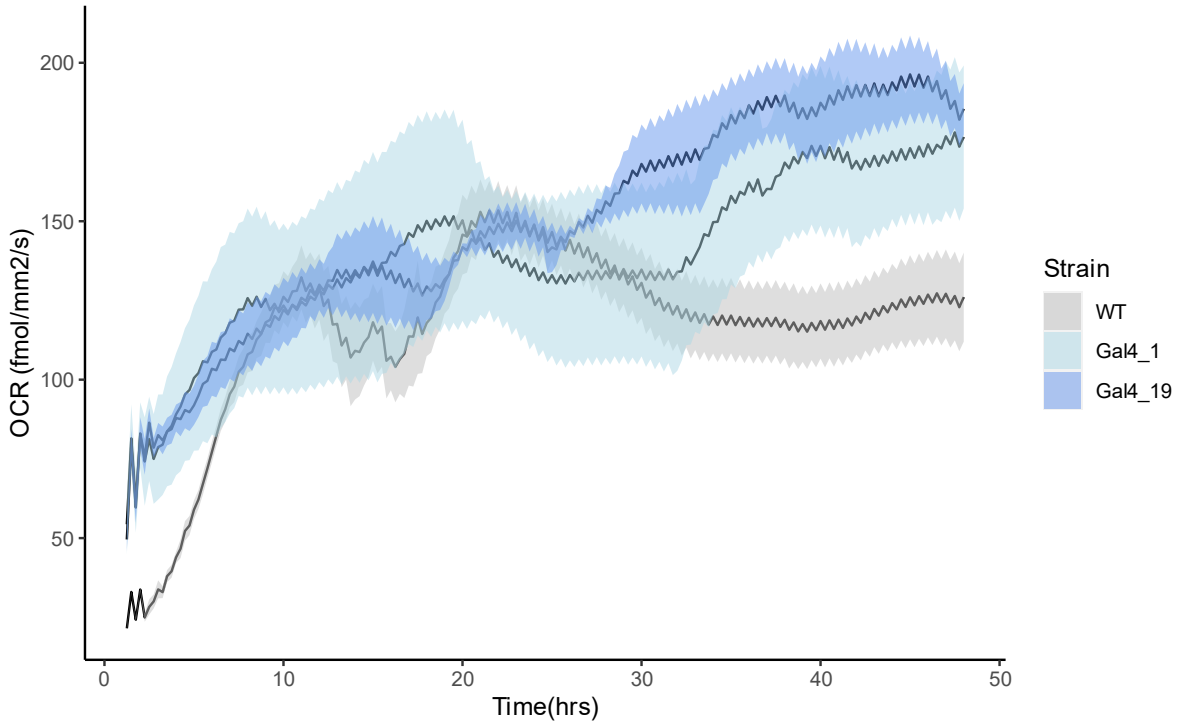


Supplemental Figure 3. Overview of comparative transcriptomics in rapid and low ECAR *Saturnispora* species. (A) A Principal Component Analysis (PCA) showing the similarity of expression profiles for the four species evaluated. The relative expression of (B) glycolytic genes and (C) citric acid cycle encoding genes across all four replicates for each species (Sdis = *Sat. dispora,* Shag = *Sat. hagleri*, Smen = *Sat. mendoncae,* Ssil = *Ssilvae*).



774

775 **Supplemental Figure 4. Unrooted phylogeny of the Gal4p orthologs in the budding yeasts**
776 **used in this study.** (A) A phylogeny including all translated amino acid sequences of genes in
777 the orthogroup containing canonical Gal4p. A black arrow indicates the location of a branch,
778 which was pruned in (B) due to the length of the branch and the topological incongruence with
779 the species phylogeny in this part of the tree, suggesting that these are likely other related C6
780 zinc transcription factors. The Newick files for these trees are available as Supplemental Data 2
781 & 3. The location of the Gal4p orthologs from *S. cerevisiae* and *Sat. dispora* are highlighted on
782 both trees and the branches are coloured according to the phylogenetic order.
783
784



Supplemental Figure 5. The respiration profile of *Sat. dispora* and *gal4Δ* deletion mutants over 48 hours of growth. For each strain, the black line represents the average oxygen consumption rate (OCR), and the shaded area represents the standard error of the mean. Note that both mutants (Gal4_1, Gal4_19) have considerably higher OCR at several phases of the experiment, including the beginning and end.

Supplemental Tables:

Table S1. The species used in this study. For each species used, information about the strain, the phylogenetic order, and naming conventions relating this to the Y1000+ Project genome information (Opulente *et al*, 2024) are provided to improve usability. The data underlying Figure 1 including measured extracellular acidification rates, the number of orthologs of each glycolytic gene, and the sum of all of the glycolytic genes are also given for each species.

Table S2. The systematic identification of transcription factor binding sites in the putative promoters of glycolytic genes for the four *Saturnispora* species used for transcriptomics. For each gene, the transcription factor, binding site, and position are noted.

Table S3. The primers used in this study.

Table S4. Genetically modified strains made and used for this study.

Supplemental Data:

Supplemental data 1. Uncropped microscopy images of GFP expressing *Sat. dispora*.

Supplemental data 2. Newick tree file of the full unrooted Gal4p tree.

Supplemental data 3. Newick tree file of the pruned unrooted Gal4p tree.