



OPEN

## Multi-omics network model reveals key genes associated with p-coumaric acid stress response in an industrial yeast strain

F. E. Ciamponi<sup>1,4</sup>, D. P. Procópio<sup>2,4</sup>, N. F. Murad<sup>1</sup>, T. T. Franco<sup>3</sup>, T. O. Basso<sup>2✉</sup> & M. M. Brandão<sup>1✉</sup>

The production of ethanol from lignocellulosic sources presents increasingly difficult issues for the global biofuel scenario, leading to increased production costs of current second-generation (2G) ethanol when compared to first-generation (1G) plants. Among the setbacks encountered in industrial processes, the presence of chemical inhibitors from pre-treatment processes severely hinders the potential of yeasts in producing ethanol at peak efficiency. However, some industrial yeast strains have, either naturally or artificially, higher tolerance levels to these compounds. Such is the case of *S. cerevisiae* SA-1, a Brazilian fuel ethanol industrial strain that has shown high resistance to inhibitors produced by the pre-treatment of cellulosic complexes. Our study focuses on the characterization of the transcriptomic and physiological impact of an inhibitor of this type, *p*-coumaric acid (pCA), on this strain under chemostat cultivation via RNAseq and quantitative physiological data. It was found that strain SA-1 tend to increase ethanol yield and production rate while decreasing biomass yield when exposed to pCA, in contrast to pCA-susceptible strains, which tend to decrease their ethanol yield and fermentation efficiency when exposed to this substance. This suggests increased metabolic activity linked to mitochondrial and peroxisomal processes. The transcriptomic analysis also revealed a plethora of differentially expressed genes located in co-expressed clusters that are associated with changes in biological pathways linked to biosynthetic and energetical processes. Furthermore, it was also identified 20 genes that act as interaction hubs for these clusters, while also having association with altered pathways and changes in metabolic outputs, potentially leading to the discovery of novel targets for metabolic engineering toward a more robust industrial yeast strain.

Expansion of the global lignocellulosic ethanol production is heading towards second-generation ethanol (2G) biorefineries. Although 2G ethanol is still more expensive than first-generation ethanol (1G), current production costs for 2G are up to 50% higher than 1G<sup>1,2</sup>, recent advances in biofuel technology suggest that 2G ethanol will be more cost-efficient in the long run, with some of the more optimistic scenario placing the turning point for this technology in the year 2025<sup>1,3,4</sup>. For commodities like ethanol, even small changes in production costs can have a big impact on the supply chain. Reducing operating costs by a few cents can result in savings of millions of dollars per year<sup>5-7</sup>.

During 2G ethanol production, before the biomass being placed in fermentation vats, the lignocellulosic feedstock undergoes pre-processing to unleash less complex sugars located in the cell wall to make these molecules available for the yeasts<sup>8</sup>. However, this process also releases several toxic compounds in the medium, and, as a result, it requires microorganisms with increasing resistance to inhibitors generated during pretreatment

<sup>1</sup>Center for Molecular Biology and Genetic Engineering (CBMEG), State University of Campinas (Unicamp), Av. Cândido Rondon, 400, Campinas, SP 13083-875, Brazil. <sup>2</sup>Department of Chemical Engineering, University of São Paulo (USP), Av. Prof. Luciano Gualberto, 380, São Paulo, SP 05508-010, Brazil. <sup>3</sup>School of Chemical Engineering (FEQ), State University of Campinas (Unicamp), Av. Albert Einstein, 500, Campinas, SP 13083-852, Brazil. <sup>4</sup>These authors contributed equally: F. E. Ciamponi and D. P. Procópio. ✉email: thiagobasso@usp.br; brandaom@unicamp.br

processes<sup>9,10</sup>. Therefore, understanding how these inhibitory molecules affect the fermentative performance of *Saccharomyces cerevisiae* is essential to implementing strategies to increase its robustness against adverse conditions in industrial fermentation and contributing to its implementation as a stable platform for biofuel production<sup>11–13</sup>.

Among these compounds, two classes of molecules—furans and organic acids—have an active physiological impact on the growth rate and overall fermentation metabolism of *S. cerevisiae*<sup>11,13–18</sup>. Additionally, phenolic compounds also inhibit the production of ethanol in anaerobic fermentation<sup>13</sup>. Although certain *S. cerevisiae* strains are resistant to these molecules, the molecular mechanism used by these yeasts to metabolize such inhibitors into less toxic compounds is complex, involving multiple regulatory processes and pathways<sup>19</sup>. One of the major byproducts resulting from the sugarcane bagasse pretreatment in the production of 2G bioethanol is *p*-coumaric acid (pCA)<sup>20,21</sup>, which was found to be in concentrations of up to 2.0 g/kg (dry weight) of bagasse after pretreatment<sup>22,23</sup>. This chemical usually inhibits the growth of *S. cerevisiae* and disrupts the production of ethanol<sup>24–27</sup>. Although some reports show that certain *S. cerevisiae* strains are capable of surviving high concentrations of pCA, or even being engineered to serve as templates for the production of this compound<sup>28,29</sup>, the same cannot be said for strains currently in use in the 2G bioethanol industry. These strains have a significantly different genomic makeup, in the form of nucleotide variation and structural rearrangements, caused by the intense selection process that this particular group of *S. cerevisiae* was subjected to<sup>30–33</sup>, and appear to be more susceptible to the inhibitory effects of these compounds<sup>26,34,35</sup>. Moreover, pCA is insoluble in water, but easily reacts and solubilizes in ethanol, posing a significant challenge for any type of industrial-scale fermentation process<sup>36</sup>.

Recently, a Brazilian industrial strain used in the bioethanol industry called SA-1 was shown to be highly resistant to several lignocellulosic inhibitors, being capable of maintaining 70% of its normal growth rate even when exposed to 7 mM of pCA, a feature that was not observed even in other industrial strains, such as the case of JAY270, a haploid derivative of PE-2 which is one of the most widespread strains currently in use in the Brazilian bioethanol industry<sup>26,37,38</sup>. However, the molecular characterization of specific survival mechanisms used by SA-1 to survive in such conditions has not yet been described. Considering that the response of *S. cerevisiae* gene expression to environmental conditions is a powerful tool for identifying targets associated with increased ethanol production and survivability and has been used to direct bioengineering efforts toward the desired phenotype by altering the transcriptional machinery of these organisms<sup>39–41</sup>. The use of data obtained for differentially expressed genes in combination with systems biology approaches also allowed researchers to identify the metabolic pathways that are affected (activated or repressed) under specific conditions<sup>42–45</sup>.

Our study focuses on characterizing the SA-1 strain profile when exposed to high concentrations of pCA under continuous fermentation conditions in a controlled bioreactor environment, using a combination of metabolite analysis, transcriptomics, and genomics in an integrative and systemic multi-omics analysis to elucidate the underlying mechanisms by which this particular strain is capable of thriving even when exposed to such inhibitors.

Our main objectives are not only to characterize the molecular aspects of the response but also to identify key genes associated with the response to such inhibitors, allowing a deeper understanding of these functions. By using these findings as a framework for future bioengineering efforts, be it in the form of gene models or pathways of interest, it is expected our work to provide valuable insights into the development of more robust industrial strains that are capable of increased survival rates when exposed to the adversary conditions present in industrial fermentation vats, assisting in the reduction overall production costs of 2G ethanol production and establishing this platform as a stable source for biofuel production.

## Results and discussion

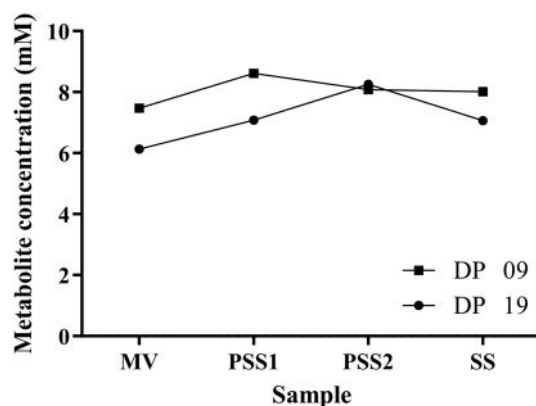
**Metabolic changes in SA-1 strain upon pCA exposure.** A thorough understanding of the effects of pCA on yeast metabolism is required to generate potential metabolic engineering strategies that can improve strain robustness. Since lignocellulosic hydrolysates contain a high number of phenolic compounds of which 80% represents pCA, this compost was added to the feed-medium of carbon-limited chemostats. Cultivation without inhibitors served as control. Data collected along each batch phase was linearized by applying the natural logarithm to exit of CO<sub>2</sub> values as a function of time. Specific consumption rates of glucose and specific production rates of selected extracellular metabolites are shown in Table 1 and Table S1.

It has been reported that in aerobic cultures containing pCA, the growth rate is significantly reduced in a dose-dependent manner and inhibiting the efficient bioconversion of lignocellulose biomass by fermentative organisms<sup>46,47</sup>. Although the *S. cerevisiae* CEN.PK113-7D strain was shown to be capable of slow in situ catabolic conversion of 9.7 mM pCA in aerobic batch cultivations, performing a complete conversion of pCA into other phenolic compounds over a period of 72 h<sup>46,47</sup>. Our results indicate that under anaerobic conditions, the same catabolic effect does not occur with the pCA concentrations remaining relatively stable across steady-state measurements. This highlights the importance of analyzing industrial *S. cerevisiae* strains under anaerobic chemostat conditions, especially considering that mitochondrial respiration has been shown to be of significant importance for controlling yeast growth rate processes and resistance to phenolic compounds<sup>48–50</sup>.

Exploring the differences in the physiological parameters resulting from the addition of the phenolic compound, multiple alterations were observed. Some specific consumption and production rates increased, such as for glucose (26%), CO<sub>2</sub> (12%), and ethanol (53%). On the other hand, decrease in biomass yield (22%), and the glycerol production rate (19%) was observed (Table 1). In anaerobic glucose-limited chemostat cultures of the *S. cerevisiae* strains, carbon is mainly diverted to ethanol and CO<sub>2</sub>, and minor amounts of glycerol, lactic and acetic acids, with a concomitant formation of yeast biomass. The ethanol yield of SA-1 in the control condition was 21% lower than in the presence of pCA. Moreover, under anaerobic glucose-limited chemostat cultivations pCA is not metabolized by SA-1 (Fig. 1).

Conditions/parameters	Control	7 mM pCA
$\mu$ (batch phase)	0.38 ± 0.02	0.37 ± 0.03
Residual glucose (g/L)	0.69 ± 0.05	0.69 ± 0.22
$q$ glucose	- 5.80 ± 0.05	- 7.34 ± 0.50
$q$ CO <sub>2</sub>	9.82 ± 1.01	11.04 ± 1.03
$q$ ethanol	8.66 ± 0.34	13.27 ± 1.17
$q$ glycerol	1.04 ± 0.07	0.84 ± 0.24
$q$ lactate	0.08 ± 0.00	0.06 ± 0.03
$q$ pyruvate	0.02 ± 0.00	0.03 ± 0.02
$q$ acetate	0.00 ± 0.00	0.00 ± 0.00
X	2.64 ± 0.01	2.14 ± 0.02
$Y_{X/S}$	0.13 ± 0.00	0.09 ± 0.00
$Y_{Eth/S}$	0.38 ± 0.02	0.46 ± 0.01
C recovery	98.5 ± 3.3	100.7 ± 0.1

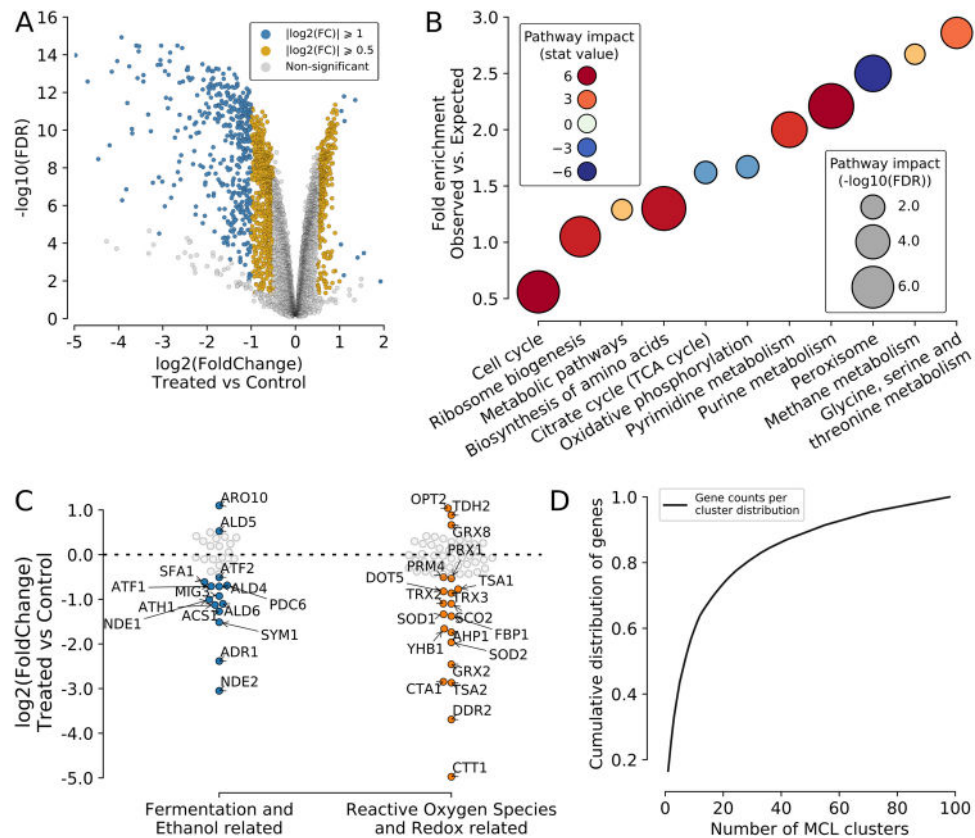
**Table 1.** Physiology of *S. cerevisiae* strains in glucose-limited anaerobic chemostats at a dilution rate of 0.1 h<sup>-1</sup>. Specific rates ( $q$ ) are given in mmol g<sup>-1</sup> h<sup>-1</sup>,  $\mu$  in h<sup>-1</sup>, yeast biomass (X) in g DW L<sup>-1</sup>, conversion factor of substrate into biomass ( $Y_{X/S}$ ) in g DW g glucose<sup>-1</sup>, ethanol yield ( $Y_{Eth/S}$ ) in g ethanol<sup>-1</sup> g glucose<sup>-1</sup>, and C recovery in (%). The  $\mu$  value represents the growth rate measured during the batch phase before the steady-state is achieved. Data is the average values of duplicate experiments ± deviation of the mean.



**Figure 1.** pCA concentration during anaerobic glucose-limited chemostat cultivations of *S. cerevisiae* SA-1 strain. MV medium vessel, PSS1 first pre-steady-state, PSS2 second pre-steady state; SS, steady-state. DP09 and DP19 code for the duplicate runs. The pre-steady-state samples (PSS1 and PSS2) represent samplings at 24 and 48 h after starting the feeding, respectively.

Overall, physiological data collected from the anaerobic chemostat cultures showed that pCA exposure may increase ethanol production in SA-1 strain ( $\log_2(\text{foldChange}) = 0.58$ ; t-test p-value < 0.05). These changes are accompanied by an increased glucose uptake rate ( $\log_2(\text{fC}) = 0.33$ , pval = 0.10), which might suggest alterations in the metabolic state. In addition, these changes were followed by a decrease in the overall dry-weight cell biomass ( $\log_2(\text{fC}) = -0.27$ , pval < 0.05) and biomass yield ( $\log_2(\text{fC}) = -0.25$ , pval = 0.10). This is different from previous studies that showed that *S. cerevisiae* Ethanol Red<sup>+</sup>, when exposed to pCA during aerobic batch cultivations, increases biomass yield and decreases ethanol yield in comparison to control conditions (without pCA)<sup>47</sup>. Strains resistant to lignocellulosic inhibitors have lower growth rates when exposed to such compounds<sup>27</sup>; however, SA-1 strain showed no signs of slower growth during the batch phase, with both control and treated samples showing an average growth rate around 0.37 h<sup>-1</sup>.

**Differential gene expression between treated and control samples.** Gene expression analysis based on RNASeq data revealed that both conditions (treated and control) have a high correlation between biological duplicates (Pearson's  $R^2 > 0.99$ , Fig. S1). These values are within the established parameters for chemostat cultures<sup>51</sup>, which generate replicates with low biological variability. Additionally, the principal component analysis showed that 79.57% of the explained variance observed in the samples can be associated with the axis that represents separation based on experimental conditions, with only 10.28% of explained variance being associated with alterations from samples under the same conditions (Fig. S1B). These analyses indicate significant changes in the transcriptomic landscape of SA-1 strain when exposed to pCA.



**Figure 2.** Differentially expressed genes under *p*-coumaric stress are skewed towards downregulation and can be grouped into functional clusters. **(A)** Volcano plot showing the relation between  $\log_2(\text{Fold Change})$  (X-axis) and  $-\log_{10}(\text{FDR})$  (Y-axis) for differentially expressed genes (DEGs in blue and gold) after pCA treatment. **(B)** Point plot showing the predicted perturbations in KEGG pathways (X-axis) based on RNAseq data. The Y-axis shows the fold-enrichment score identified for each pathway of DEGs; the size of the point shows the significant value of the prediction (in  $-\log_{10}(\text{FDR})$  scale), and the color represents the direction of perturbation: downregulated (brown) or upregulated (blue). **(C)** Swarmplot showing the  $\log_2(\text{FoldChange})$  profile (Y-axis) of the 13 DEGs associated with fermentation and/or ethanol (blue) and the 17 DEGs associated with reactive oxygen species response and redox processes (orange). **(D)** Cumulative distribution function (CDF) plot showing the percentage of DEGs (Y-axis) that are in clusters (X-axis), starting from the biggest clusters (in the number of genes).

In order to assess the quality of our differential gene expression data, we measured the observed relationships between average gene expression ratios and their deviation from the mean (Fig. S1C), the observed p-values and variation and expression quantiles (Fig. S1D) as well as the correlation between observed gene expression standard deviation and median expression value (Fig. S1E). Not only does our quality control data show that there is little-to-no bias in our dataset, it has a similar behavior to what is expected from a well controlled differential gene expression experiment when using RNA samples collected in the second steady-state of chemostat cultures<sup>52</sup>.

It was identified a total of 1472 differentially expressed genes between conditions (404 up- and 1068 down-regulated); however, only 448 (10 up- and 438 down-regulated) of the genes was statistically significant, with  $p\text{-value} < 0.05$ ,  $\text{FDR} \leq 0.01$ , and had  $|\log_2(\text{FoldChange})| \geq 1$  (Fig. 2A, Table S2). It is known that the presence of insoluble lignocellulosic inhibitors in the medium can promote generalized downregulation of gene expression in another industrial *S. cerevisiae* strain<sup>34</sup>. With that in mind, the possible transcriptomic alterations induced by pCA in our chemostat experiments were explored. By using gene expression data obtained from our RNA-seq data, we observed a tendency towards global downregulation of the transcriptional machinery. This same expression pattern was already observed by Moreno et al.<sup>34</sup> as a response to the presence of insoluble inhibitory compounds in the fermentation medium. Our findings not only reflect the same pattern reported, but highlights the importance of understanding the biological impacts of different types of inhibitory conditions on fermentative microorganisms for directing bioengineering efforts towards more robust strains for the ethanol industry.

The DEGs identified in our study were mapped in the KEGG database<sup>53</sup> to identify perturbations in pathways containing SA-1 differentially expressed gene sets (either up- or down-regulated), which affect the overall pathway activity (Fig. 2B, Table S3). Using this approach, a total of 11 pathways that had statistically significant gene set alterations by pCA stress were identified, and three were negatively altered (repressed): oxidative phosphorylation, citrate cycle, and peroxisome. The other eight had perturbations with positive effects on the pathway

(activated): cell cycle, ribosome biogenesis, metabolic pathways, biosynthesis of amino acids, pyrimidine metabolism, purine metabolism, methane metabolism and glycine, serine, and threonine metabolism. Additionally, a total of 20 differentially expressed genes (17 downregulated and 3 upregulated) were identified, associated with redox regulation and/or response to reactive oxygen species and another set of 15 DEGs (13 downregulated and 2 upregulated) that can be linked to ethanol metabolism and/or fermentation (Fig. 2C). No overlapping genes between these two sets (ROS/Redox and Ethanol/Fermentation) were observed.

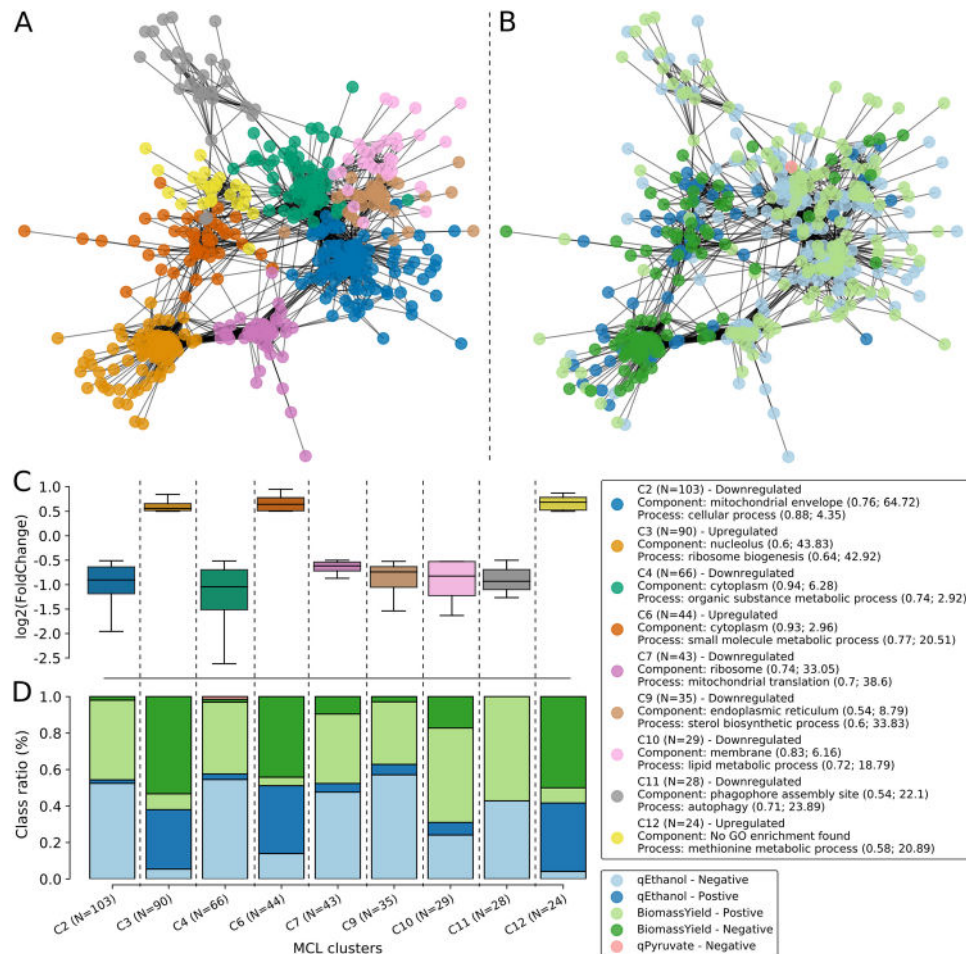
Amongst the impacted pathways, purine metabolism was one of the most prominent, this pathway is involved in the formation of adenine and guanine. The former is an essential part of the overall cell metabolism, in the form of ATP/ADP/AMP, by providing energy for cellular processes, and the latter plays a distinct role in cell response to stress conditions in the form of GTP/GDP, a molecule often used in signaling processes during stress response for transcriptional regulation<sup>54</sup> and glucose signaling via GPCR<sup>55</sup>. The increase in the metabolism of these compounds could also be related to an increased rate of metabolic processes during the stress response. Furthermore, several differentially expressed genes that have known associations, via Gene Ontology<sup>56,57</sup>, to reactive oxygen species and ethanol fermentative processes are identified. Both of these processes are known to be intricately related to the mitochondria and peroxisome organelles<sup>48,58,59</sup>, and are known hallmarks of yeast response to stress induced by lignocellulosic inhibitors<sup>59–62</sup>. In addition to the aforementioned processes, it was also identified several other upregulated pathways that are linked to mitochondrial activity, such as biosynthesis and metabolism of amino acids<sup>63</sup>. These processes are intrinsically associated with the TCA cycle, which is downregulated in our dataset, and of paramount importance in maintaining amino-acid homeostasis, which, in turn, is vital for promoting long-term viability in yeasts<sup>64</sup>.

In addition to several of our results suggesting alterations surrounding pathways related to the mitochondria in response to pCA stress, multiple studies already pointed out the importance of this organelle towards the resistance to this particular compound<sup>65–67</sup>. However, the exact mechanisms by which the pCA affects this *S. cerevisiae* organelle under anaerobic conditions are still largely unexplored. When exploring the literature available for other organisms, some studies conducted in rat liver and human cells showed that pCA may cause damage to the mitochondria: by inhibiting the pyruvate transport mechanism<sup>68,69</sup>, inducing reactive oxygen species (ROS) damage<sup>70</sup> and mitochondrial membrane depolarization<sup>71</sup>. Another study demonstrated that PAD1, a mitochondrial protein that is downregulated ( $\log_2(\text{FC}) = -0.5$ ) in SA-1, is essential for the decarboxylation of phenylacrylic acids<sup>65</sup>. In all the three cases described in the literature, mitochondrial damage ultimately leads to a signaling cascade that starts cell autophagy.

Graph network clustering using known protein–protein interaction data associated with fold changes derived from RNASeq was applied to identify co-expressed gene clusters. This approach is advantageous for inferring networks from data with a low number of samples, as is the case of most experiments in biology. This is because it brings precision from a prior knowledge of a manually curated database (STRINGdb) combined with fold change values that include the particularity of the experiment. A total of 98 clusters were found within the differentially expressed genes, with 50% of DEGs being located in the 7 biggest clusters (Fig. 2D, Table S4). These clusters were then characterized according to their expression profile (Table S5), based on the distribution of  $\log_2(\text{fold change})$  for genes within each cluster, enrichment of significant Gene Ontology classes and KEGG pathways (Table S5). Additionally, no correlation ( $R^2 = 0.033$ ) was found between the number of genes and the overall standard deviation of fold changes observed within each cluster.

**Functional characterization of co-expressed gene clusters.** To further explore the clustered gene sets, the clusters were filtered by the fold change values, located in a distance of  $1.5 \times \text{IQR}$  (interquartile range), were in the same quadrant, either above or below zero, and had more than 20 genes. A total of 9 clusters (C2, C3, C4, C6, C7, C9, C10, C11, and C12), with a total of 462 DEGs, were selected and characterized according to biological functional enrichment and association (Fig. 3A, Table S6) with phenotypic alterations (Fig. 3B, Table S7). Six of these clusters (C2, C4, C7, C9, C10, and C11) comprised downregulated genes, while 3 (C3, C6, and C12) were mostly from upregulated genes (Fig. 3C). However, each cluster had distinct associations with biomass yield and ethanol production (Fig. 3D). Downregulated clusters tended to share positive associations with biomass and negative relations with ethanol production. Upregulated clusters, on the other hand, showed an inverse pattern, with negative regulation of biomass and positive association with ethanol.

To delve deeper into the transcriptomic landscape alterations induced by pCA stress on SA-1 strain, we shifted our focus to the co-expressed gene clusters extracted from our network in order to characterize which functional groups of genes were being up/down-regulated and if those clusters could be related to the pathways changes observed previously. Amongst the groups of genes that stood out from the background, it was noticed that one of the co-expressed clusters associated with autophagy (C11) was actively repressed in SA-1 under p-coumaric stress. Our data also suggests a potential disruption of the mitochondria, with 70% of the genes located in two downregulated clusters being directly associated with mitochondrial cellular processes (C2) and translation (C7). This correlates with the negative impact observed for KEGG pathways associated with the citrate cycle and oxidative phosphorylation. Furthermore, 20 of the 27 differentially expressed genes associated with the peroxisomal pathway (AGX1, CAT2, CTA1, DCI1, ECI1, FAA2, IDP2, IDP3, PEX1, PEX11, PEX14, PEX2, PEX5, POT1, POX1, PXA1, PXA2, SPS19, YAT1, and YAT2) are also located in a downregulated cluster associated with the cytoplasmic organic substance metabolic process (C4). Peroxisomes are involved with long fatty acid degradation and biosynthesis in yeasts<sup>72,73</sup>, with lipid metabolism/biogenesis being one of the downregulated gene clusters identified (C10). In contrast, increased expression was detected in genes linked with nuclear ribogenesis (C3) and metabolic activities (C6 and C12). When compared with the predicted pathway impact, C3 has 19 of the 23 DEGs associated with the ribosome biogenesis pathway (AFG2, DIP2, EMG1, KRE33, POP1, POP7, PWP2, RIO2, RIX7, RNT1, SDO1, UTP13, UTP18, UTP22, UTP4, UTP5, UTP6, UTP8, UTP9), C6 contains 14 of the 31

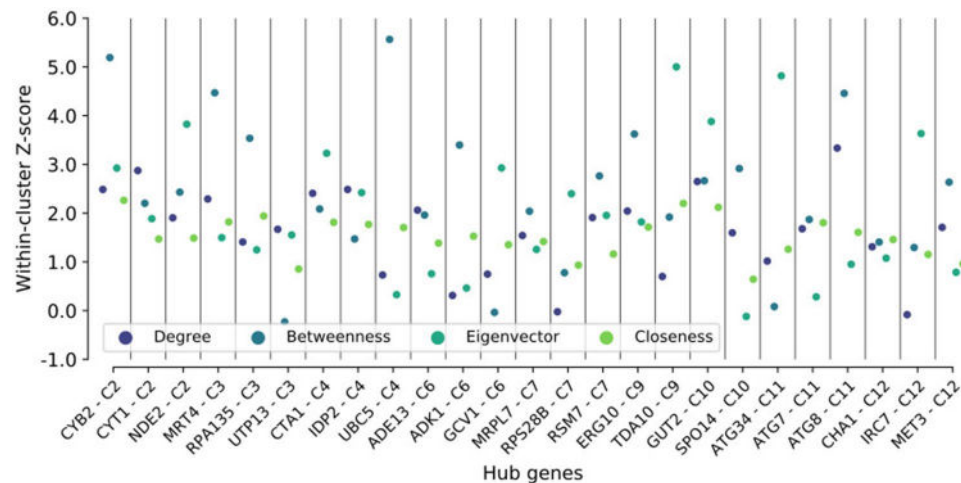


**Figure 3.** Co-expressed gene clusters can be classified according to functional enrichment and association with phenotype alterations. **(A, B)** Network representation of the 9 co-expressed gene clusters selected for further exploration. On the right **(A)**, the clusters were classified according to their gene ontology functional enrichment; on the left **(B)** each gene in the network was classified according to their probability of having a positive or negative association with either qEthanol or biomass yield metabolic phenotypes. **(C)** Boxplot showing the expression profile (Y-axis) of each of the selected clusters (X-axis). The color of each cluster matches those of the network. **(D)** Stacked barplot showing the ratio of genes (Y-axis) associated with the metabolic classes (positive/negative relation to biomass yield/qEthanol) for each of the selected clusters.

DEGs associated with purine metabolism (ADE1, ADE12, ADE13, ADE17, ADE2, ADE4, ADE5,7, ADE6, ADE8, ADK1, GUD1, HPT1, IMD2, and IMD4) and C12 contains 13 of the 31 DEGs associated with the biosynthesis of amino acids (ARG1, ASN1, GLY1, HIS1, HIS4, HIS6, HIS7, HOM3, LYS2, LYS4, SER1, SER2, and SHM2). When coupled with the alterations observed in gene sets associated with the mitochondria, our data imply that multiple biological pathways are being regulated to compensate for the stress induced by pCA exposure. Such regulation might be associated with processes used by SA-1 to survive under adverse conditions.

To further explore the characterized clusters, we also evaluated each of the genes found within the clusters to establish which of them acted as “network hubs” in their respective clusters, based on their values for eigenvector centrality, betweenness, degree, and closeness. In total, 25 genes (Fig. 4, Table S8) that act as the main points of interaction for the genes in the network were identified.

**Prediction of genomic short-variants based on RNA-seq.** Using data collected from high-throughput RNA sequencing, we reconstructed short variants (SNPs and INDELs) that occur within the transcripts and predicted the impact that they might have on their associated coding sequence<sup>74</sup>. A total of 38,420 short variations were identified (compared to the R64-1-1 reference annotation) that can be subdivided into three major categories: homozygous (both alleles carry the variant), heterozygous with reference (one allele carries the variant, while the other is equal to the reference) and heterozygous without reference (both alleles carry different variants, and neither is equal to the reference). These short variants were then classified according to their predicted impact on their associated coding sequence (Fig. 5A): 683 modifier variants (non-coding, e.g., intronic or UTR variants); 23,388 low impact variants (e.g., synonymous mutations); 13,655 moderate impact variants (e.g., missense mutations that preserve overall protein length/structure) and 1093 high-impact variants



**Figure 4.** Hub genes are located in co-expressed gene clusters. Swarmplot showing the centrality scores (Y-axis) measured for the hub-genes selected from each cluster (X-axis). Each metric used is shown in a different color.

(e.g., frameshift INDELS or stop-gaining SNPs). However, the sum of the subclasses exceeds the total of variants (Table S9). This occurs because a variant can impact multiple genes, and current limitations make it difficult to solve such conflicts using HTS data alone<sup>74,75</sup>.

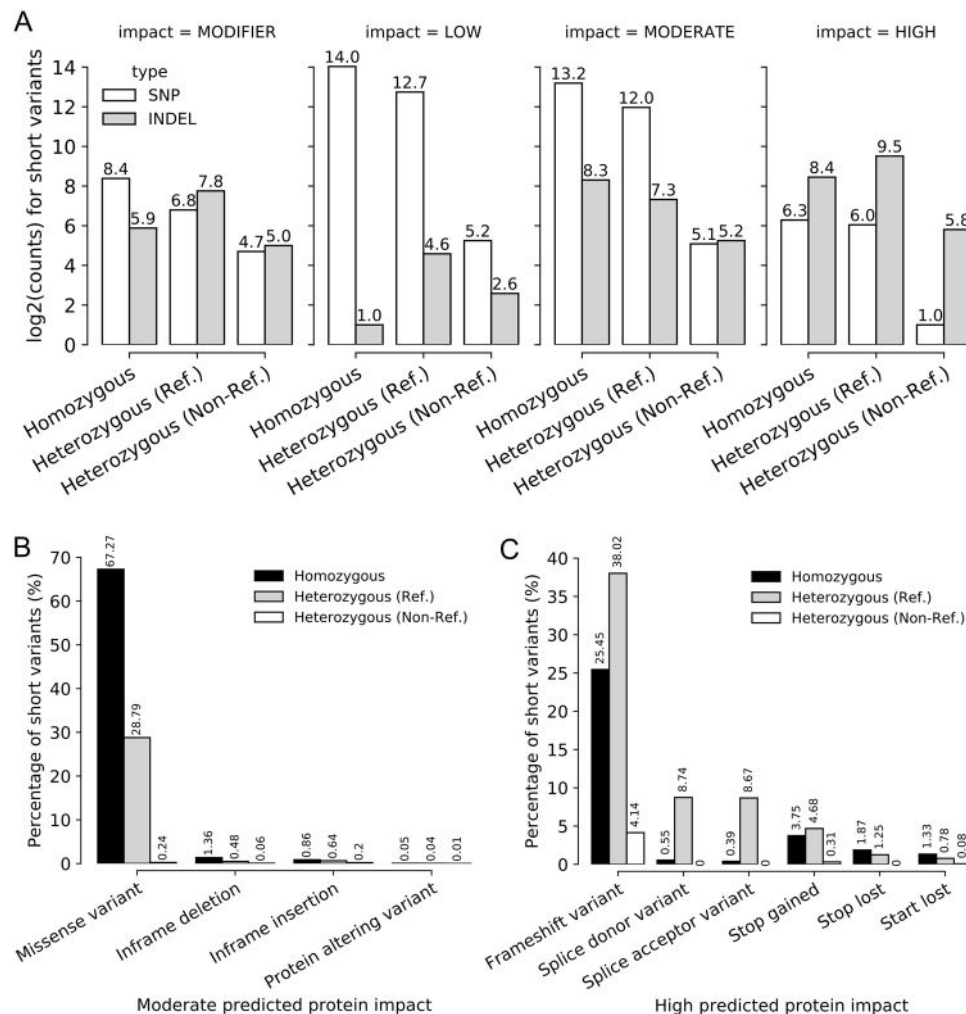
The major classes of high-impact variants involved frameshift variants, and most of them (~38%) were heterozygous (concerning the reference) in nature. Homozygous frameshift variants also comprised the second biggest class, with approximately 25% of variants falling in that class, and non-reference heterozygous frameshifts comprising ~4% of high-impact variants. Heterozygous variants located in splicing sites were also a major class of high-impact predictions, with a combined total of ~17%. Lastly, stop-gaining mutations represented ~8% of all high-impact variants, with approximately half (3.75%) being homozygous in nature and the remaining (4.68%) being heterozygous.

Then only the variants that had predicted moderate and high protein impacts were filtered for further exploration. Our analysis showed that the major class of moderate impact alterations was the class of missense variants, with ~67% of them being homozygous in nature and ~29% being heterozygous with the reference (Fig. 5B). However, when comparing the same results for high-impact variants, a much broader distribution of CDS consequences ( $\chi^2(3, N = 15,166) = 7096.5, p < 0.001$ , Fig. 5C) was observed.

In addition to gene expression changes, mutations are also a major player in the process of generating resistant strains for biofuel production, be them artificially generated or naturally selected<sup>76–78</sup>. These structural alterations can promote changes in the function of genes and proteins in multiple ways<sup>79</sup>, with even small mutations possibly having far-reaching effects<sup>80,81</sup>. By characterizing the profile of single nucleotide variants when compared to the S288C reference strain (which has extensive functional gene annotation data), it was possible to predict the impacts of mutations found in SA-1's genes<sup>82</sup>. Although we recognize that a comprehensive analysis based on comparative genomics of industrial yeasts, which is outside the scope of the present article, would be ideal to characterize the genomic complexity of the SA-1 industrial strain, it's expect that, the extraction of the information on the genetic diversity encountered in SA-1 will provide an important layer of information on the mechanisms associated to pCA response on this strain.

**Assembly of a multi-omics network model for pCA response.** To generate a single model that represents the association between differentially expressed genes located in perturbed pathways, the association with fermentation/ethanol and ROS/Redox, the presence of high-impact short variants, and phenotype impact prediction, all the information described in the previous sections was converted into a multi-omics graph-based network model. Integrative approaches, such as this, are especially relevant to “Big Data” datasets, such as ours, to extract comprehensive models that capture subtleties involved in biological regulation that otherwise would be lost if each “omic” was only analyzed independently<sup>83</sup>. This type of analysis has already been successfully used in multiple fields, from biomedical research<sup>84,85</sup> to biotechnology<sup>86,87</sup>. Furthermore, it has been demonstrated that in-silico modeling of complex biological networks can be a powerful tool for driving the improvement of commercially relevant organisms, such as the development of an improved model of *Aspergillus nidulans* metabolic network models, which is important for the construction and optimization of glucoamylase-producing strains<sup>88–90</sup>.

When applied specifically to *S. cerevisiae*, this strategy has also been proven to be crucial in unraveling novel molecular mechanisms associated with gene regulation<sup>91</sup>, stress tolerance<sup>92</sup> and selection of targets for bioengineering<sup>93</sup>. From the complete network, all the edges in which at least one vertex was either a hub gene (as shown in Fig. 4) or a gene associated with fermentation/ethanol or ROS/Redox (as shown in Fig. 2C) were extracted. A total of 16 genes (Fig. 6, Table 2) were selected based on the aforementioned criteria for constructing the model, while these targets were clustered into two major groups: those associated with ethanol production (IDP2, ERG10, CYT1, ARO10, GCV1, TDA10, and CHA1) and those related to biomass yield (SOD1, CTA1,



**Figure 5.** Homozygous missense and heterozygous frameshift variants are the major classes of short-variants with predicted moderate-to-high protein impact. **(A)** Barplots showing the overall number of variants (Y-axis) identified from RNASeq data for the SA-1 strain when compared to the R64-1-1 reference annotation for *S. cerevisiae*. The X-axis shows the ploidy identified for each variant: homozygous, heterozygous with reference, or heterozygous without reference. The columns are separated according to the overall level of predicted protein impact, from lowest to highest. **(B, C)** Barplot showing the breakdown of the predicted protein impacts (X-axis) and their associated proportions (Y-axis) for variants with moderate **(B)** and high **(C)** predicted protein impact.

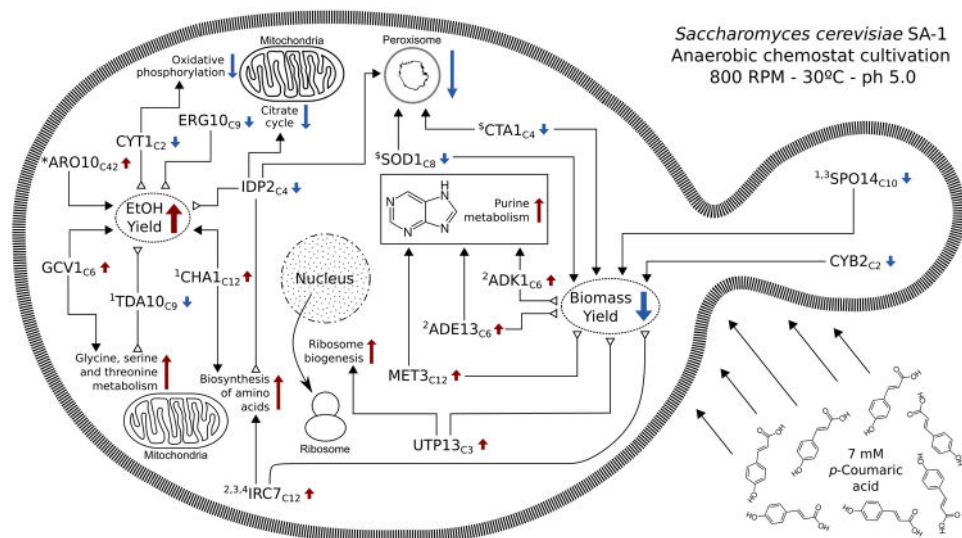
IRC7, SPO14, UTP13, CYB2, MET3, ADK1, and ADE13). In order to facilitate the discussion, we will be focusing on the hub genes mentioned previously, since they are the most representative targets for their respective co-expressed clusters. However, we strongly encourage the exploration of our entire network model made available in the supplementary files (Table S10) by the readers.

This multi-level network also showed the type of interaction between genes and their targeted phenotype and associated pathway, which can be either a direct relationship between the changes in gene expression and the pathway/phenotype alteration (e.g. both upregulated) or an inverse relationship (where one is upregulated and the other is downregulated). In total, 19 positive interactions (where the gene foldChange occurs in the same direction as the change in pathway activity or metabolite measurement) and 11 negative interactions (where the gene foldChange occurs in the opposite direction as the change in pathway activity or metabolite measurement) were found. The proposed network showed 6 genes with predicted mutations (TDA10, CHA1, SPO14, IRC7, ADK1 and ADE13), 2 genes associated with ROS/Redox processes (CTA1 and SOD1) and 1 gene directly related to fermentation/ethanol (ARO10).

Genes identified as anchor nodes in the multi-omics model were annotated using keywords assigned by the UniProt database that reflect their functional and structural characteristics. The additional columns also show their predicted phenotypic association (Phenotype), any associated pathways (Pathway) and if the gene contains an SNV site (SNV).

By anchoring our network on genes and using their associations (positive, negative or neutral) to pathways, phenotypes and genomic variants, it was possible to identify two main groups of targets: those associated with





**Figure 6.** Multi-omics integrated model for alterations induced by pCA stress in the *S. cerevisiae* SA-1 strain. Network representation of the relation between hub genes, short variants (SVs), perturbed pathways, and phenotypic alterations. Red arrows indicate positive regulation (upregulated genes and pathways), blue arrows show negative relations (downregulated genes and pathways) and black arrows represent additional information sources (presence of SVs or gene ontology categories), with their connector showing the relationship as direct (solid arrowhead, both elements moving in the same direction) or inverted (hollow triangle, elements moving in opposite directions). Genes labeled with numbers indicate the presence of a short variant (1: missense homozygous mutation; 2: missense heterozygous; 3: frameshift heterozygous and 4: stop-gained heterozygous). Genes labeled with symbols represent known associations to either fermentation processes (\*) or response to reactive oxygen species (\$).

ethanol production and those associated with biomass yield. Each of these groups has unique features, particularly in terms of their target pathways, which will be explored separately.

When analyzing the first group of genes (ethanol related) of the integrated response model, the TDA10 gene, an ATP-binding protein with unknown function that resembles *E. coli* kinases<sup>94,95</sup>, was downregulated ( $\log_2\text{FoldChange} - 0.72$ ) and had an inverse relation with glycine, serine, and threonine metabolism and ethanol production. In addition, the TDA10 gene was also the target of a homozygous missense variant in position 343 of the CDS, which changes the corresponding amino acid from phenylalanine to leucine, causing structural changes to the overall protein. A negative correlation with ethanol production for the ERG10 and IDP2 genes was observed. The former (ERG10) may act in the oxidative stress response<sup>95</sup> and its deletion was associated with slower doubling times and susceptibility to high NaCl concentrations<sup>96</sup>, being a major target for genetic engineering approaches<sup>29,97,98</sup>. The latter (IDP2) is an isocitrate dehydrogenase that was downregulated in our dataset ( $\log_2\text{FC} - 3.57$ ) and has been linked to small reductions in yeast lifespan<sup>99</sup>—this gene was also downregulated in mutants susceptible to thermosensitive autolysis and associated with mitochondrial dysfunction<sup>100</sup>.

It was also identified 3 genes that had a positive correlation with ethanol production: ARO10, GCV1, and CHA1 (Fig. 2C). Besides its regulatory role in fermentation<sup>101</sup>, ARO10 acts in the detoxification of damaged amino acids and resistance to lignocellulosic compounds, such as HMF and furfural<sup>102</sup>. This gene was upregulated ( $\log_2\text{FoldChange} 1.09$ ) upon exposure to 7 mM of pCA, with a positive correlation to ethanol production. The GCV1 gene, upregulated in our dataset ( $\log_2\text{FC} 0.82$ ), encodes the T subunit of the mitochondrial glycine decarboxylase system and increases in expression under multiple types of stress responses in *S. cerevisiae*<sup>103–106</sup>; however, the exact role of GCV1 in these scenarios is still not fully understood. Lastly, the CHA1 was slightly upregulated in our dataset ( $\log_2\text{FC} 0.51$ ) and had a positive correlation with the biosynthesis of amino acids and metabolism of glycine, serine, and threonine. This gene catalyzes the degradation of L-serine and L-threonine to use them as nitrogen sources and is upregulated in the response to ethanol stress<sup>107</sup> and congo red<sup>108</sup>.

Both CYB2 and CYT1 are mitochondrial genes that are regulated during changes in the anaerobic metabolic processes and fermentation of glucose<sup>109–111</sup>. In our dataset, both genes showed a negative correlation with metabolic pathways (that is, they were downregulated while the pathway was activated). Moreover, they appear to have inverse relations with phenotypic changes: CYB2 has a positive (direct) relation with biomass yield, and CYT1 has a negative (inverse) relation with the ethanol output.

In the second group of genes (biomass related) a total of four genes (SOD1, CTA1, CYB2, and SPO14) with a positive influence on the biomass yield and five with a negative correlation (IRC7, UTP13, MET3, ADK1 and ADE13) were identified. One of the most interesting targets in this group is SOD1, a downregulated gene ( $\log_2\text{FC} - 1.34$ ): it encodes a Cu–Zn superoxide dismutase that has the main role of catalyzing the breakdown of toxic superoxides in the cell<sup>112</sup> and is also involved in signaling processes involving oxygen and glucose stimuli<sup>113</sup>. However, recent studies suggested that the main biological role of these proteins in yeasts is the peroxide signaling and activation of peroxisomes and multiple cell homeostasis pathways<sup>114</sup>. This is in accordance with our findings

Gene	Cluster	Phenotype	Pathway	SNV	Description (UniProt keyword)
UTP13	C3	Biomass (-)	Ribosome biogenesis (+)	No	Nucleus, repeat, ribonucleoprotein, ribosome biogenesis, WD repeat, rRNA processing
ADE13	C6	Biomass (-)	Purine metabolism (+)	Yes	Isopeptide bond, lyase, purine biosynthesis, Ubl conjugation
ADK1	C6	Biomass (-)	Purine metabolism (+)	Yes	ATP-binding, acetylation, cytoplasm, kinase, mitochondrion, nucleotide-binding, transferase
IRC7	C12	Biomass (-)	Biosynthesis of amino acids (+)	Yes	Amino-acid biosynthesis, lyase, methionine biosynthesis, pyridoxal phosphate
MET3	C12	Biomass (-)	Purine metabolism (+)	No	ATP-binding, amino-acid biosynthesis, cysteine biosynthesis, cytoplasm, methionine biosynthesis, nucleotide-binding, nucleotidyltransferase, transferase
CYB2	C2	Biomass (+)	None	No	Electron transport, FMN, flavoprotein, heme, iron, metal-binding, mitochondrion, oxidoreductase, respiratory chain, transit peptide, transport
CTA1	C4	Biomass (+)	Peroxisome (+)	No	Acetylation, heme, hydrogen peroxide, iron, metal-binding, oxidoreductase, peroxidase, peroxisome
SOD1	C8	Biomass (+)	Peroxisome (+)	No	Antioxidant, copper, cytoplasm, disulfide bond, isopeptide bond, metal-binding, mitochondrion, oxidoreductase, phosphoprotein, ubl conjugation, zinc
SPO14	C10	Biomass (+)	None	Yes	Acetylation, hydrolase, lipid degradation, lipid metabolism, meiosis, phosphoprotein, repeat, sporulation
CYT1	C2	Ethanol (-)	Oxidative phosphorylation (+)	No	Electron transport, heme, iron, membrane, metal-binding, mitochondrion, mitochondrion inner membrane, respiratory chain, transit peptide, translocase, transmembrane, transmembrane helix, transport
IDP2	C4	Ethanol (-)	Peroxisome (+), Citrate cycle (+), Biosynthesis of amino acids (-)	No	Cytoplasm, glyoxylate bypass, magnesium, manganese, metal-binding, NADP, oxidoreductase, tricarboxylic acid cycle
ERG10	C9	Ethanol (-)	None	No	Acetylation, acyltransferase, cytoplasm, metal-binding, potassium, transferase
TDA10	C9	Ethanol (-)	Glycine, serine and threonine metabolism (-)	Yes	ATP-binding, cytoplasm, kinase, nucleotide-binding, nucleus, transferase
GCV1	C6	Ethanol (+)	Glycine, serine and threonine metabolism (+)	No	Aminotransferase, mitochondrion, transferase, transit peptide
CHA1	C12	Ethanol (+)	Biosynthesis of amino acids (+)	yes	Acetylation, lyase, mitochondrion, pyridoxal phosphate
ARO10	C42	Ethanol (+)	None	No	Branched-chain amino acid catabolism, cytoplasm, decarboxylase, isopeptide bond, lyase, magnesium, metal-binding, phenylalanine catabolism, thiamine pyrophosphate, tryptophan catabolism, tyrosine catabolism, ubl conjugation

**Table 2.** List of genes that were used as nodes in the multi-omics model.

for both gene expression, pathway impact and co-expressed gene clusters enrichment. The other gene associated with response to reactive oxygen species was CTA1, a downregulated gene ( $\log_2\text{FC} - 2.84$ ) in our dataset. This gene encodes a catalase associated with ROS detoxification in peroxisome and in the mitochondria<sup>115</sup>, and its activity is relevant in oxidative<sup>116</sup>, acetic acid<sup>117</sup> and heat<sup>118</sup> stress responses. These genes showed positive associations with the activity of the peroxisomal pathway, which is one of the most affected by the stress induced by pCA in SA-1 strain. While SOD1 did not appear as a hub in the selected gene clusters, it did act as a major interactor for cluster 8, which is associated with regulatory and cell homeostasis pathways (Table S5). However, CTA1 is a major interaction hub for C4, being enriched for genes related to the metabolism of organic substances in the cytoplasm. Additionally, the SPO14 ( $\log_2\text{FC} - 0.82$ ) gene was associated with changes in the cell cycle regulation (Honigberg et al. 1992) and regulation of lipid metabolism<sup>119,120</sup> in *S. cerevisiae*. In addition to its role in the cell cycle, SPO14 was also a hub gene for cluster 10, which is enriched in genes for the lipid metabolic process.

As for the genes that had negative correlations with the biomass yield, the IRC7 gene seems to be a target with multiple associated conditions. Besides its transcriptional behavior ( $\log_2\text{FC} 0.60$ ) in SA-1 strain, when exposed to pCA, this gene is involved in the production of thiol compounds<sup>121</sup> and yeast survivability using cysteine as nitrogen source<sup>122</sup>. This gene was the most affected by our analysis of variants, accumulating a large amount of moderate-to-high-impact impact variants in heterozygosity, but none in homozygosity, suggesting that one of the alleles might be severely impaired. This corroborates an analysis of wine fermenting yeasts, which showed that several *S. cerevisiae* strains carried inactivating mutations for one or both alleles of IRC7<sup>123</sup>, reducing the overall enzymatic activity of this protein. Another study showed that the over-expression of IRC7 also resulted in the increased production of hydrogen sulfide<sup>122</sup>, a volatile sulfur compound that has been linked to increased longevity in *S. cerevisiae*<sup>124</sup>. Lastly, 3 genes that showed a positive correlation with purine metabolism: MET3 ( $\log_2\text{FC} 0.73$ ), ADK1 ( $\log_2\text{FC} 0.50$ ) and ADE13 ( $\log_2\text{FC} 0.82$ ) were also identified. Upregulated in our dataset ( $\log_2\text{FoldChange} 0.73$ ), MET3 is an ATP sulfurylase involved in sulfate and methionine metabolism<sup>125</sup>, which was upregulated during hypoxia<sup>48</sup>. Moreover, the over-expression of ADE13 may increase fermentation efficiency under acetic acid stress<sup>126</sup>, while ADK1 appears to be activated in response to sulphuric acid<sup>127</sup> and to heat stress<sup>128</sup>. These three genes were also associated with clusters enriched in genes linked to the regulation of metabolic processes.

## Conclusion

Our results suggest that *p*-coumaric acid (pCA) stress may induce higher cellular activity in SA-1 strain under anaerobic conditions, with increased glucose uptake rate, and CO<sub>2</sub> and ethanol production rates, being the major indicators obtained from quantitative physiological data. In accordance, it was also observed a decrease in biomass yield and overall dry-weight cell biomass, which implicates the existence of some type of disturbance in the cell homeostasis. It was demonstrated that pCA stress can cause an overall activation of metabolic and biosynthesis pathways, which are also followed by increased rRNA biogenesis. Downregulation of several mitochondrial and peroxisomal-associated pathways may also be an indicator of cellular damage caused by the exposure to pCA; our data suggest that SA-1 strain has yet-to-be-explored molecular mechanisms that allow them to circumvent triggers that lead to programmed cell death. At the gene level, multiple genes that could be novel and/or interesting targets for bioengineering were identified. Our results highlight the importance of an integrated approach for target identification and association with phenotypes of interest for industrial applications. By using network-enhanced gene cluster detection, genes that could be the most influential in their biological vicinity were identified. These “hub genes” are prime targets for genetic engineering approaches, as they are the ones with the highest impact on their sphere of influence and are most-likely to produce deep alterations in the associated biological process within the gene community. Although exploratory in nature, the data presented in this study contributes to understanding the characteristics of pCA-induced stress in *S. cerevisiae* and deepening the knowledge of mechanisms used by industrial yeast strains that can thrive under high-stress conditions, such as exposure to lignocellulosic inhibitors.

Taken together, our results show that the biological mechanisms used by *S. cerevisiae* SA-1 to survive under the influence of lignocellulosic inhibitors are much more intricate than previously understood. Multiple biological pathways, which sometimes have opposite effects when analyzed individually, are intertwined in a complex balance that allows these yeasts to thrive even when exposed to high levels of stress. Systemic analysis is essential to understand the nuances involved in such interactions, with several information sources and analyses being integrated into a single model that can reflect multiple levels of biological data. This is especially relevant for researches in economically-driven or similar fields, such as bioethanol production and other industrial capacities, where the ability to select targets for bioengineering approaches that maximize the desired effect (e.g. improving ethanol production) while minimizing undesired side-effects (e.g. affecting unrelated pathways and/or other phenotypes) can be of paramount importance to gain a competitive edge. By using our network model as a frame of reference to develop strains that are more robust to the effects of inhibitory compounds, such as pCA, we hope to drive innovation towards a more robust yeast strain that is capable of improved efficiency under the strenuous conditions imposed by industrial fermentation vats.

## Methods

**Yeast strain and cultivation conditions.** The strain investigated in this study, *S. cerevisiae* SA-1, is an industrial strain obtained from Fermentec (Piracicaba, Brazil). Inoculum cultures were prepared from glycerol stocks stored at – 80 °C on a defined medium<sup>129,130</sup>, whose composition (in g L<sup>-1</sup>) is described as follows: (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5.0; KH<sub>2</sub>PO<sub>4</sub>, 3.0; MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.5; 1 mL L<sup>-1</sup> trace element solution, 1 mL L<sup>-1</sup> vitamin solution, and 25 g L<sup>-1</sup> glucose. Cultures were grown overnight at 30 °C in a rotary shaker at 200 rpm. Chemostat cultivation with pCA was performed in a 2.0 L water jacket model Labfors 5 (Infors AG, Switzerland) with 1.0 L working volume, which was kept constant by a mechanical drain and a peristaltic pump. Throughout cultivation, both the culture vessel (0.5 L min<sup>-1</sup>) and the medium vessel (flow rate not measured) were purged with nitrogen gas to maintain anaerobic conditions. The agitation frequency was set at 800 rpm, the temperature was controlled at 30 °C, and the pH was adjusted to 5.0 using a controlled 2 M KOH solution. Pre-cultures for batch bioreactor cultivations were grown overnight in an orbital shaker at 30 °C and 200 rpm in 500 mL shake flasks containing 100 mL of the defined medium with 20 g L<sup>-1</sup> starting glucose. The medium had the same composition as the preculture, except that Tween 80 and ergosterol were added at a final concentration of 0.01 g L<sup>-1</sup> and 0.42 g L<sup>-1</sup>, respectively, to allow anaerobic growth. The batch phase was terminated after glucose depletion (monitored by a sharp drop in CO<sub>2</sub> concentration in the exhaust gas), whereupon cultivation switched to continuous mode with the addition of fresh medium supplemented or not supplemented with 7 mM pCA. The dilution rate was set at 0.1 h<sup>-1</sup> and the cultivation was assumed to be in a steady state when the dry weight of the culture and the specific carbon dioxide production rate varied by less than 2% for two volume changes during at least five residence times<sup>131</sup>.

The chemostat system was chosen due to its characteristics of maintaining physiological conditions in constant values among experiments, which is important when trying to isolate transcriptomic alterations that arise in response to a singular input (in our case, the presence of pCA), eliminating the effects of growth rates and other stochastic perturbations which may arise due to environmental conditions<sup>132</sup>.

**Analysis of extracellular metabolites.** Cell dry mass concentration was determined by the gravimetric method<sup>133</sup>. Extracellular metabolite samples from the chemostat cultures were filtered through 0.2 μm syringe filters. Concentrations of residual carbon, ethanol, glycerol, and organic acids were quantified by high-performance liquid chromatography (HPLC)<sup>134</sup>, using a Prominence HPLC model (Shimadzu Corporation, Japan) and an HPX-87H analytical column (Bio-Rad Laboratories, USA) at 60 °C with 5 mM H<sub>2</sub>SO<sub>4</sub> as mobile phase at 0.6 mL min<sup>-1</sup>. Ethanol concentrations were corrected for evaporation<sup>135</sup> and pCA was quantified<sup>136</sup> by using an HPLC and a C18 analytical column (Supelco Inc. model Waters Spherisorb ODS—25 μm, 250 mm × 4.6 mm) at 30 °C with 2% (v/v) acetic acid in ionized water (eluent A) and acetic acid 0.5% in ionized water and acetonitrile (50:50, v/v; eluent B) as mobile phase at 1.0 mL min<sup>-1</sup> using a gradient program: from 10 to 15% B (10 min), 15% B isocratic (3 min), 15 to 25% B (7 min), 25 to 55% B (30 min), 55 to 100% B (1 min), 100% B isocratic

(5 min), from 100 to 10% B (0.1 min). The total run time was 60 min, with a flow rate of 1.0 mL min<sup>-1</sup> and an oven temperature of 30 °C. The injection volume for all samples was 10 µL. Monitoring was performed using a Shimadzu UV detector at wavelengths of 280 nm and 320 nm. Concentrations of compounds were calculated from calibration curves obtained from standard solutions<sup>131</sup>.

**RNA extraction and sequencing.** RNA extraction was performed using the Direct-zol™ RNA MiniPrep kit (Zymo Research catalog no. R2051) following the manufacturer's instructions. RNA samples were sequenced using BGISEQ-500, with each library generating approximately 24 M paired-end reads of 100 bp. Raw RNA-seq reads were filtered to remove adapter contamination and low-quality reads, with Trimmomatic<sup>137</sup>. Each sample was aligned against the R64-1-1 version of the *S. cerevisiae* reference genome, which is based on the S288C strain, with STAR v2.7.0<sup>138</sup>, using “—sjdbGTFfile,” “—quantMode GeneCounts,” “—twopassMode Basic” and the ENCODE guidelines for best practices of eukaryotic RNASeq<sup>139</sup> as additional parameters. The corresponding gene annotation files and variant call files were also obtained for the same assembly. All genome data was obtained from Ensembl Fungi release 48<sup>140</sup>.

**Gene expression analysis and functional characterization.** Differential gene expression was assessed by edgeR v.3.3<sup>141</sup>, using FDR ≤ 0.01 and |log<sub>2</sub>(FoldChange)| ≥ 0.5 as cutoffs for statistical significance. Gene expression in log<sub>2</sub>(CPM) scale was used to perform principal component analysis (PCA) and replicate similarity assessment to check the significance of biological duplicates. All downstream functional enrichment analyses were done using STRINGdb v.11<sup>142</sup>, and pathway perturbation analysis with Pathview API and GAGE v.2.38<sup>143,144</sup>. In both cases, an FDR cutoff of 0.01 was applied using the coding genome as background, and the Fold enrichment value was calculated based on the number of observed genes in comparison to the number of expected hits.

**Co-expressed gene cluster detection and hub gene identification.** Co-expressed gene clusters were identified using an adaptation of the kNN-enhance method, which intensifies an existing network with node attributes<sup>145</sup>. The total protein–protein interaction (ppi) network from STRINGdb (v11) was converted into an undirected graph, where each node is a protein and the edges represent known interactions between them—only interactions with a total ppi\_score ≥ 0.7 (high confidence) were considered for downstream analysis. Each node metadata information was enhanced with an extra attribute corresponding to the log<sub>2</sub>(foldChange) value of that protein, and for each pair of vertices connected by an edge, a second score was used (called foldChange\_score). This metric was calculated by  $1 - \text{norm}((X_i - X_j)_2)$  where  $X_i$  is the foldChange for vertex 1, and  $X_j$  is the foldChange for vertex 2, and  $\text{norm}((X_i - X_j)_2)$  is the normalized Euclidean distance between  $X_i$  and  $X_j$ . Thus, values for foldChange\_score varied from 1 (identical foldChange scores) to 0 (the largest foldChange difference between two nodes in the network).

Final edge weight scores were calculated by combining foldChange\_score and ppi\_score in a new “enhanced\_score.” Co-expressed gene clusters were identified using MCL clustering<sup>146,147</sup>, applied to the attribute-enhanced network, with inflation hyper-parameter tuned to maximize modularity score (Q). For each of the identified clusters, it was also extracted genes that could serve as “local hubs” based on four different metrics: degree, betweenness, eigenvector and closeness.

BNFinder<sup>148</sup> was used for the association between genes and phenotypical data, combining per-sample normalized gene expression (from RNASeq) with physiological data (derived from HPLC), and converted the observations into classes with the Sturges' rule<sup>149</sup>.

A two-fold strategy was applied to generate cluster functional labels: the first was based on gene ontology enrichment classes, with the most significant enriched class (lowest FDR) that represented at least 50% of genes within the cluster; the second strategy was based on Bayesian inference of association with physiological data, with genes being able to be associated either positively or negatively with the changes in each measured metabolite.

**Short variant discovery.** Short variants (SNPs and Indels) were identified using the GATK4 pipeline, in accordance with the best practices for RNASeq short variant discovery<sup>74,75</sup>. Aligned RNA/seq reads in BAM format were used as input, as well as GFF and VCF files for R64-1-1 annotation obtained from Ensembl. Short variant impact was estimated using Ensembl Variant Effect Predictor<sup>150</sup>.

**Multi-omics network assembly.** A graph-based approach was used to integrate all data layers (gene expression, co-expressed cluster hubs, pathway impact and nucleotide variants) into a unified network model using NetworkX<sup>151</sup>. The information for each layer was re-structured and merged with the others in a way that for every pair of vertices  $u$  and  $v$ , the first vertex ( $u$ ) represents a gene and the second vertex ( $v$ ) represents the characteristic associated to that gene (a pathway, phenotype or mutation). The weight of the edge defined by  $(u,v)$  was assigned according to the relationship between the expression change of the gene and alteration on the pathway/phenotype: +1 for direct relationships (the direction of the gene fold change is in the same direction of the altered pathway/phenotype, i.e. both upregulated), -1 for inverse relationships (the direction of the gene fold change is in the opposed direction of the altered pathway/phenotype, i.e. one upregulated while the other is downregulated) or 0 for neutral relationships (in the case of nucleotide variants). Directionality of the network was established in accordance with the following structure: variant → gene → pathway/phenotype, to reflect the idea that: “a nucleotide variant may affect the gene function, leading downstream alterations”.

**Ethical approval.** This article does not contain any studies with human participants or animals performed by any of the authors.

### Data availability

Raw sequencing files, scripts and all supplementary data are available at <https://labis.cbmeg.unicamp.br/labis/publicacoes/71-sal-pcoumaric>, fastq reads were deposited on Sequence Read Archive with the following accession codes: SRR15944188 (DP06); SRR15944187 (DP07); SRR15944185 (DP09); SRR15944186 (DP19). Additionally, all files can also be accessed via Unicamp's Research Data Repository via DOI code <https://doi.org/10.25824/redu>.

Received: 28 September 2022; Accepted: 21 December 2022

Published online: 28 December 2022

### References

1. Junqueira, T. L. *et al.* Techno-economic analysis and climate change impacts of sugarcane biorefineries considering different time horizons. *Biotechnol. Biofuels* **10**, 50 (2017).
2. Petersen, A. M., Okoro, O. V., Chireshe, F., Moonsamy, T. & Görgens, J. F. Systematic cost evaluations of biological and thermochemical processes for ethanol production from biomass residues and industrial off-gases. *Energy Convers. Manag.* **243**, 114398 (2021).
3. Raj, T. *et al.* Recent advances in commercial biorefineries for lignocellulosic ethanol production: Current status, challenges and future perspectives. *Bioresour. Technol.* **344**, 126292 (2022).
4. Tapia Carpio, L. G. & de Souza, F. S. Competition between second-generation ethanol and bioelectricity using the residual biomass of sugarcane: Effects of uncertainty on the production mix. *Molecules* **24**, 369 (2019).
5. Kohler, M. Chapter 19: Economic assessment of ethanol production. in *Ethanol* (eds. Basile, A., Iulianelli, A., Dalena, F. & Veziroglu, T. N.) 505–521 (Elsevier, 2019). <https://doi.org/10.1016/B978-0-12-811458-2.00019-5>.
6. McAloon, A., Taylor, F. & Yee, W. Determining the cost of producing ethanol from corn starch and lignocellulosic feedstocks. *Molecules* **44**, 1–10 (2000).
7. Mizik, T. Impacts of international commodity trade on conventional biofuels production. *Sustainability* **12**, 2626 (2020).
8. Tumuluru, J. S. *Biomass Preprocessing and Pretreatments for Production of Biofuels: Mechanical, Chemical and Thermal Methods* (CRC Press, 2018).
9. Maurya, D. P., Singla, A. & Negi, S. An overview of key pretreatment processes for biological conversion of lignocellulosic biomass to bioethanol. *3 Biotech* **5**, 597–609 (2015).
10. Sindhu, R., Binod, P. & Pandey, A. Biological pretreatment of lignocellulosic biomass: An overview. *Bioresour. Technol.* **199**, 76–82 (2016).
11. Almeida, J. R. M. M. *et al.* Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *J. Chem. Technol. Biotechnol.* **82**, 340–349 (2007).
12. Hamelinck, C. N., Van Hooijdonk, G. & Faaij, A. P. C. Ethanol from lignocellulosic biomass: Techno-economic performance in short-, middle- and long-term. *Biomass Bioenergy* **28**, 384–410 (2005).
13. Sambusiti, C., Monlau, F., Ficara, E., Carrère, H. & Malpei, F. A comparison of different pre-treatments to increase methane production from two agricultural substrates. *Appl. Energy* **104**, 62–70 (2013).
14. Larsson, S. *et al.* The generation of fermentation inhibitors during dilute acid hydrolysis of softwood. *Enzyme Microb. Technol.* **24**, 151–159 (1999).
15. Larsson, S., Quintana-Sáinz, A., Reimann, A., Nilvebrant, N.-O. & Jönsson, L. J. Influence of lignocellulose-derived aromatic compounds on oxygen-limited growth and ethanolic fermentation by *Saccharomyces cerevisiae*. *Appl. Biochem. Biotechnol.* **84–86**, 617–632 (2000).
16. Nilsson, A., Gorwa-Grauslund, M. F., Hahn-Hägerdal, B. & Lidén, G. Cofactor dependence in furan reduction by *Saccharomyces cerevisiae* in fermentation of acid-hydrolyzed lignocellulose. *Appl. Environ. Microbiol.* **71**, 7866–7871 (2005).
17. Russell, J. B. Another explanation for the toxicity of fermentation acids at low pH: Anion accumulation versus uncoupling. *J. Appl. Bacteriol.* **73**, 363–370 (1992).
18. Taherzadeh, M. J., Gustafsson, L., Niklasson, C. & Lidén, G. Physiological effects of 5-hydroxymethylfurfural on *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **53**, 701–708 (2000).
19. Favaro, L., Jansen, T. & van Zyl, W. H. Exploring industrial and natural *Saccharomyces cerevisiae* strains for the bio-based economy from biomass: The case of bioethanol. *Crit. Rev. Biotechnol.* <https://doi.org/10.1080/07388551.2019.1619157> (2019).
20. Jönsson, L. J. & Martín, C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour. Technol.* <https://doi.org/10.1016/j.biortech.2015.10.009> (2016).
21. Reinoso, F. A. M. *et al.* Fate of p-hydroxycinnamates and structural characteristics of residual hemicelluloses and lignin during alkaline-sulfite chemithermomechanical pretreatment of sugarcane bagasse. *Biotechnol. Biofuels* <https://doi.org/10.1186/s13068-018-1155-3> (2018).
22. Biazzi, L. E. *et al.* Adaptation strategy to increase the tolerance of *Scheffersomyces stipitidis* NRRL Y-7124 to inhibitors of sugarcane bagasse hemicellulosic hydrolysate through comparative studies of proteomics and fermentation. *BioEnergy Res.* **15**, 479–492 (2022).
23. van der Pol, E. *et al.* Analysis of by-product formation and sugar monomerization in sugarcane bagasse pretreated at pilot plant scale: differences between autohydrolysis, alkaline and acid pretreatment. *Bioresour. Technol.* **181**, 114–123 (2015).
24. Adeboye, P. T., Bettiga, M. & Olsson, L. ALD5, PAD1, ATF1 and ATF2 facilitate the catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid in *Saccharomyces cerevisiae*. *Sci. Rep.* **7**, 1–13 (2017).
25. Baranowski, J. D., Davidson, P. M., Nagel, C. W. & Branan, A. L. Inhibition of *Saccharomyces cerevisiae* by naturally occurring hydroxycinnamates. *J. Food Sci.* <https://doi.org/10.1111/j.1365-2621.1980.tb04107.x> (1980).
26. Cola, P. *et al.* Differential effects of major inhibitory compounds from sugarcane-based lignocellulosic hydrolysates on the physiology of yeast strains and lactic acid bacteria. *Biotechnol. Lett.* <https://doi.org/10.1007/s10529-020-02803-6> (2020).
27. Gu, H., Zhang, J. & Bao, J. Inhibitor analysis and adaptive evolution of *Saccharomyces cerevisiae* for simultaneous saccharification and ethanol fermentation from industrial waste corncob residues. *Bioresour. Technol.* <https://doi.org/10.1016/j.biortech.2014.01.060> (2014).
28. Borja, G. M. *et al.* Metabolic engineering and transcriptomic analysis of *Saccharomyces cerevisiae* producing p-coumaric acid from xylose. *Microb. Cell Factories* **18**, 191 (2019).
29. Liu, J. F., Xia, J. J., Nie, K. L., Wang, F. & Deng, L. Outline of the biosynthesis and regulation of ergosterol in yeast. *World J. Microbiol. Biotechnol.* **35**, 7 (2019).
30. Jacobus, A. P. *et al.* Comparative genomics supports that Brazilian bioethanol *Saccharomyces cerevisiae* comprise a unified group of domesticated strains related to cachaça spirit yeasts. *Front. Microbiol.* **12**, 687 (2021).

31. Wohlbach, D. J. *et al.* Comparative genomics of *Saccharomyces cerevisiae* natural isolates for bioenergy production. *Genome Biol. Evol.* **6**, 2557–2566 (2014).
32. Zhang, K. *et al.* Genomic reconstruction to improve bioethanol and ergosterol production of industrial yeast *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* **42**, 207–218 (2015).
33. Zhang, K. *et al.* Genomic structural variation contributes to phenotypic change of industrial bioethanol yeast *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **16**, 118 (2016).
34. Moreno, A. D., González-Fernández, C., Ballesteros, M. & Tomás-Pejó, E. Insoluble solids at high concentrations repress yeast's response against stress and increase intracellular ROS levels. *Sci. Rep.* <https://doi.org/10.1038/s41598-019-48733-w> (2019).
35. van der Pol, E. C., Bakker, R. R., Baets, P. & Eggink, G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. *Appl. Microbiol. Biotechnol.* **98**, 9579–9593 (2014).
36. Salameh, D., Brandam, C., Medawar, W., Lteif, R. & Strehaiano, P. Highlight on the problems generated by p-coumaric acid analysis in wine fermentations. *Food Chem.* <https://doi.org/10.1016/j.foodchem.2007.09.052> (2008).
37. de Mello, F. *et al.* Static microplate fermentation and automated growth analysis approaches identified a highly-aldehyde resistant *Saccharomyces cerevisiae* strain. *Biomass Bioenergy* <https://doi.org/10.1016/j.biombioe.2018.10.019> (2019).
38. Nagamatsu, S. T. *et al.* Genome assembly of a highly aldehyde-resistant *Saccharomyces cerevisiae* SA1-derived industrial strain. *Microbiol. Resour. Announc.* **8**, e00071–e119 (2019).
39. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568 (2006).
40. Lin, Z., Zhang, Y. & Wang, J. Engineering of transcriptional regulators enhances microbial stress tolerance. *Biotechnol. Adv.* **31**, 986–991 (2013).
41. Techaparin, A., Thanonkeo, P. & Klanrit, P. Gene expression profiles of the thermotolerant yeast *Saccharomyces cerevisiae* strain KCU-VN8 during high-temperature ethanol fermentation using sweet sorghum juice. *Biotechnol. Lett.* **39**, 1521–1527 (2017).
42. Feng, X. & Zhao, H. Investigating host dependence of xylose utilization in recombinant *Saccharomyces cerevisiae* strains using RNA-seq analysis. *Biotechnol. Biofuels* **6**, 1–17 (2013).
43. Hancock, T., Takigawa, I. & Mamitsuka, H. Mining metabolic pathways through gene expression. *Bioinformatics* **26**, 2128–2135 (2010).
44. Hernández-Elvira, M. & Sunnerhagen, P. Post-transcriptional regulation during stress. *FEMS Yeast Res.* <https://doi.org/10.1093/femsyr/foac025> (2022).
45. Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* **25**, 75–82 (2009).
46. Adeboye, P. T., Bettiga, M. & Olsson, L. The chemical nature of phenolic compounds determines their toxicity and induces distinct physiological responses in *Saccharomyces cerevisiae* in lignocellulose hydrolysates. *AMB Express* **4**, 1–10 (2014).
47. Adeboye, P. T., Bettiga, M., Aldaeus, F., Larsson, P. T. & Olsson, L. Catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid by *Saccharomyces cerevisiae* yields less toxic products. *Microb. Cell Factories* <https://doi.org/10.1186/s12934-015-0338-x> (2015).
48. Kitagaki, H. & Takagi, H. Mitochondrial metabolism and stress response of yeast: Applications in fermentation technologies. *J. Biosci. Bioeng.* **117**, 383–393 (2014).
49. Fletcher, E. & Baetz, K. Multi-faceted systems biology approaches present a cellular landscape of phenolic compound inhibition in *Saccharomyces cerevisiae*. *Front. Bioeng. Biotechnol.* **8**, 539902 (2020).
50. Malecki, M., Kamrad, S., Ralsler, M. & Bähler, J. Mitochondrial respiration is required to provide amino acids during fermentative proliferation of fission yeast. *EMBO Rep.* **21**, e50845 (2020).
51. Nookaew, I. *et al.* A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: A case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **40**, 10084–10097 (2012).
52. Hesketh, A., Vergnano, M. & Oliver, S. G. Determination of the global pattern of gene expression in yeast cells by intracellular levels of guanine nucleotides. *MBio* **10**, e02500–e2518 (2019).
53. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
54. Brauer, M. J. *et al.* Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol. Biol. Cell* **19**, 352–367 (2008).
55. Fuchs, B. B. & Mylonakis, E. Our paths might cross: The role of the fungal cell wall integrity pathway in stress response and cross talk with other stress response pathways. *Eukaryot. Cell* <https://doi.org/10.1128/EC.00193-09> (2009).
56. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
57. Gene Ontology Consortium. The gene ontology resource: Enriching a GOLD mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
58. Ayer, A., Gourlay, C. W. & Dawes, I. W. Cellular redox homeostasis, reactive oxygen species and replicative ageing in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **14**, 60–72 (2014).
59. Sibirny, A. A. Yeast peroxisomes: Structure, functions and biotechnological opportunities. *FEMS Yeast Res.* <https://doi.org/10.1093/femsyr/fow038> (2016).
60. Jayakody, L. N., Horie, K., Hayashi, N. & Kitagaki, H. Engineering redox cofactor utilization for detoxification of glycolaldehyde, a key inhibitor of bioethanol production, in yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **97**, 6589–6600 (2013).
61. Li, K. *et al.* Extracellular redox potential regulation improves yeast tolerance to furfural. *Chem. Eng. Sci.* **196**, 54–63 (2019).
62. Vall-llaura, N., Mir, N., Garrido, L., Vived, C. & Cabisco, E. Redox control of yeast Sir2 activity is involved in acetic acid resistance and longevity. *Redox Biol.* **24**, 101229 (2019).
63. Malina, C., Larsson, C. & Nielsen, J. Yeast mitochondria: An overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Res.* **18**, 040 (2018).
64. Baccolo, G., Stamerra, G., Coppola, D. P., Orlandi, I. & Vai, M. Mitochondrial metabolism and aging in yeast. In *International Review of Cell and Molecular Biology* Vol. 340 (eds López-Otín, C. & Galluzzi, L.) 1–33 (Academic Press, 2018).
65. Mukai, N., Masaki, K., Fujii, T., Kawamukai, M. & Iefuji, H. PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* <https://doi.org/10.1016/j.jbiosc.2009.11.011> (2010).
66. Richard, P., Viljanen, K. & Penttilä, M. Overexpression of PAD1 and FDC1 results in significant cinnamic acid decarboxylase activity in *Saccharomyces cerevisiae*. *AMB Express* <https://doi.org/10.1186/s13568-015-0103-x> (2015).
67. Morales, J., Mendoza, L. & Cotoras, M. Alteration of oxidative phosphorylation as a possible mechanism of the antifungal action of p-coumaric acid against *Botrytis cinerea*. *J. Appl. Microbiol.* **123**, 969–976 (2017).
68. Killeen, D. J., Boulton, R. & Knoesen, A. Advanced monitoring and control of redox potential in wine fermentation. *Am. J. Enol. Vitic.* **69**, 394–399 (2018).
69. Liu, C.-G., Qin, J.-C. & Lin, Y.-H. Fermentation and redox potential. *Ferment. Process.* <https://doi.org/10.5772/64640> (2017).
70. Lima, L. C. N. *et al.* Metabolic effects of p-coumaric acid in the perfused rat liver. *J. Biochem. Mol. Toxicol.* <https://doi.org/10.1002/jbt.20114> (2006).
71. Posadino, A. M. *et al.* Coumaric acid induces mitochondrial damage and oxidative-mediated cell death of human endothelial cells. *Cardiovasc. Toxicol.* <https://doi.org/10.1007/s12012-013-9205-3> (2013).

72. Jang, Y., Lim, Y. & Kim, K. *Saccharomyces cerevisiae* strain improvement using selection, mutation, and adaptation for the resistance to lignocellulose-derived fermentation inhibitor for ethanol production. *J. Microbiol. Biotechnol.* <https://doi.org/10.4014/jmb.1402.02040> (2014).
73. Zhou, Y. J. *et al.* Harnessing yeast peroxisomes for biosynthesis of fatty-acid-derived biofuels and chemicals with relieved side-pathway competition. *J. Am. Chem. Soc.* **138**, 15368–15377 (2016).
74. GATK. *RNAseq short variant discovery (SNPs + Indels)*. <https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels> (2020).
75. Van der Auwera, G. A. *et al.* From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* <https://doi.org/10.1002/0471250953.bi1110s43> (2013).
76. Qi, F. *et al.* Novel mutant strains of *Rhodospiridium toruloides* by plasma mutagenesis approach and their tolerance for inhibitors in lignocellulosic hydrolyzate. *J. Chem. Technol. Biotechnol.* <https://doi.org/10.1002/jctb.4180> (2014).
77. Reid, R. J. D. *et al.* Selective ploidy ablation, a high-throughput plasmid transfer protocol, identifies new genes affecting topoisomerase I-induced DNA damage. *Genome Res.* **21**, 477 (2011).
78. Yang, Y. *et al.* Progress and perspective on lignocellulosic hydrolyzate inhibitor tolerance improvement in *Zymomonas mobilis*. *Bioresour. Bioprocess.* <https://doi.org/10.1186/s40643-018-0193-9> (2018).
79. Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572–582 (2010).
80. Nagasundaram, N. *et al.* Analysing the effect of mutation on protein function and discovering potential inhibitors of CDK4: Molecular modelling and dynamics studies. *PLoS ONE* **10**, e0133969 (2015).
81. Zhang, M., Case, D. A. & Peng, J. W. propagated perturbations from a peripheral mutation show interactions supporting WW domain thermostability. *Structure* **26**, 1474–1485.e5 (2018).
82. Zhang, K. *et al.* Genetic characterization and modification of a bioethanol-producing yeast strain. *Appl. Microbiol. Biotechnol.* **102**, 2213–2223 (2018).
83. Cui, F., Cheng, L. & Zou, Q. Briefings in functional genomics special section editorial: analysis of integrated multiple omics data. *Brief. Funct. Genomics* **20**, 196–197 (2021).
84. Hoang, L. T. *et al.* Metabolomic, transcriptomic and genetic integrative analysis reveals important roles of adenosine diphosphate in haemostasis and platelet activation in non-small-cell lung cancer. *Mol. Oncol.* **13**, 2406–2421 (2019).
85. Khella, C. A., Mehta, G. A., Mehta, R. N. & Gatzka, M. L. Recent advances in integrative multi-omics research in breast and ovarian cancer. *J. Pers. Med.* **11**, 149 (2021).
86. Lu, H. *et al.* Multi-omics integrative analysis with genome-scale metabolic model simulation reveals global cellular adaptation of *Aspergillus niger* under industrial enzyme production condition. *Sci. Rep.* **8**, 14404 (2018).
87. Wang, R. *et al.* Integrative analyses of metabolome and genome-wide transcriptome reveal the regulatory network governing flavor formation in kiwifruit (*Actinidia chinensis*). *New Phytol.* **233**, 373–389 (2022).
88. Villena, G. K., Kitazono, A. A. & Hernández-Macedo, M. L. Bioengineering fungi and yeast for the production of enzymes, metabolites, and value-added compounds. In *Fungal Biotechnology and Bioengineering* (eds Hesham, A.E.-L. *et al.*) 209–237 (Springer, 2020). [https://doi.org/10.1007/978-3-030-41870-0\\_9](https://doi.org/10.1007/978-3-030-41870-0_9).
89. Lu, H. *et al.* Comprehensive reconstruction and in silico analysis of *Aspergillus niger* genome-scale metabolic network model that accounts for 1210 ORFs. *Biotechnol. Bioeng.* **114**, 685–695 (2017).
90. Fiedler, M. R. M., Barthel, L., Kubisch, C., Nai, C. & Meyer, V. Construction of an improved *Aspergillus niger* platform for enhanced glucoamylase secretion. *Microb. Cell Factories* **17**, 95 (2018).
91. Martínez-Matías, N. *et al.* Toward the discovery of biological functions associated with the mechanosensor Mtl1p of *Saccharomyces cerevisiae* via integrative multi-OMICs analysis. *Sci. Rep.* **11**, 7411 (2021).
92. Kang, K. *et al.* Linking genetic, metabolic, and phenotypic diversity among *Saccharomyces cerevisiae* strains using multi-omics associations. *GigaScience* **8**, 015 (2019).
93. Liu, J.-J. *et al.* Investigating the role of the transcriptional regulator Ure2 on the metabolism of *Saccharomyces cerevisiae*: A multi-omics approach. *Appl. Microbiol. Biotechnol.* **105**, 5103–5112 (2021).
94. De La Sierra-Gallay, I. L. *et al.* Crystal structure of the YGR205w protein from *Saccharomyces cerevisiae*: Close structural resemblance to *E. coli* pantothenate kinase. *Proteins Struct. Funct. Genet.* **54**, 776–783 (2004).
95. Higgins, V. J., Beckhouse, A. G., Oliver, A. D., Rogers, P. J. & Dawes, I. W. Yeast genome-wide expression analysis identifies a strong ergosterol and oxidative stress response during the initial stages of an industrial lager fermentation. *Appl. Environ. Microbiol.* **69**, 4777–4787 (2003).
96. Bhattacharya, S., Esquivel, B. D. & White, T. C. Overexpression or deletion of ergosterol biosynthesis genes alters doubling time, response to stress agents, and drug susceptibility in *Saccharomyces cerevisiae*. *MBio* **9**, e01291 (2018).
97. Jia, D. *et al.* *Yarrowia lipolytica* construction for heterologous synthesis of  $\alpha$ -santalene and fermentation optimization. *Appl. Microbiol. Biotechnol.* **103**, 3511–3520 (2019).
98. Kwak, S. *et al.* Enhanced isoprenoid production from xylose by engineered *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **114**, 2581–2591 (2017).
99. Laschober, G. T. *et al.* Identification of evolutionarily conserved genetic regulators of cellular aging. *Aging Cell* **9**, 1084–1097 (2010).
100. Zhang, M. *et al.* Screening of thermosensitive autolytic mutant brewer's yeast and transcriptomic analysis of heat stress response. *Can. J. Microbiol.* **66**, 631–640 (2020).
101. Deed, R. C., Hou, R., Kinzurik, M. I., Gardner, R. C. & Fedrizzi, B. The role of yeast ARO8, ARO9 and ARO10 genes in the biosynthesis of 3-(methylthio)-1-propanol from L-methionine during fermentation in synthetic grape medium. *FEMS Yeast Res.* **19**, 109 (2019).
102. Liu, Z. L. & Ma, M. Pathway-based signature transcriptional profiles as tolerance phenotypes for the adapted industrial yeast *Saccharomyces cerevisiae* resistant to furfural and HMF. *Appl. Microbiol. Biotechnol.* **104**, 3473–3492 (2020).
103. Alonso-Monge, R. *et al.* Hyperosmotic stress response and regulation of cell wall integrity in *Saccharomyces cerevisiae* share common functional aspects. *Mol. Microbiol.* **41**, 717–730 (2001).
104. Chandler, M., Stanley, G. A., Rogers, P. & Chambers, P. A genomic approach to defining the ethanol stress response in the yeast *Saccharomyces cerevisiae*. *Ann. Microbiol.* **54**, 427–454 (2004).
105. De Melo, H. F., Bonini, B. M., Thevelein, J., Simões, D. A. & Morais, M. A. Physiological and molecular analysis of the stress response of *Saccharomyces cerevisiae* imposed by strong inorganic acid with implication to industrial fermentations. *J. Appl. Microbiol.* **109**, 116–127 (2010).
106. Morrisette, V. A. & Rolfes, R. J. The intersection between stress responses and inositol pyrophosphates in *Saccharomyces cerevisiae*. *Curr. Genet.* **66**, 901–910 (2020).
107. Dong, Y., Hu, J., Fan, L. & Chen, Q. RNA-Seq-based transcriptomic and metabolomic analysis reveal stress responses and programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Sci. Rep.* **7**, 1–16 (2017).
108. García, R. *et al.* The global transcriptional response to transient cell wall damage in *Saccharomyces cerevisiae* and its regulation by the cell integrity signaling pathway. *J. Biol. Chem.* **279**, 15183–15195 (2004).
109. Lodi, T. & Guaid, B. Complex transcriptional regulation of the *Saccharomyces cerevisiae* CYB2 gene encoding cytochrome b2: CYP1(HAP1) activator binds to the CYB2 upstream activation site UAS1-B2. *Mol. Cell. Biol.* **11**, 3762–3772 (1991).

110. Zhang, T., Bu, P., Zeng, J. & Vancura, A. Increased heme synthesis in yeast induces a metabolic switch from fermentation to respiration even under conditions of glucose repression. *J. Biol. Chem.* **292**, 16942–16954 (2017).
111. Zitomer, R. S. & Lowry, C. V. Regulation of gene expression by oxygen in *Saccharomyces cerevisiae*. *Microbiol. Rev.* **56**, 1–11 (1992).
112. Bermingham-McDonogh, O., Gralla, E. B. & Valentine, J. S. The copper, zinc-superoxide dismutase gene of *Saccharomyces cerevisiae*: Cloning, sequencing, and biological activity. *Proc. Natl. Acad. Sci. USA.* **85**, 4789–4793 (1988).
113. Reddi, A. R. & Culotta, V. C. SOD1 integrates signals from oxygen and glucose to repress respiration. *Cell* **152**, 224–235 (2013).
114. Montllor-Albalade, C. *et al.* Extra-mitochondrial Cu/Zn superoxide dismutase (Sod1) is dispensable for protection against oxidative stress but mediates peroxide signaling in *Saccharomyces cerevisiae*. *Redox Biol.* **21**, 101064 (2019).
115. Petrova, V. Y., Drescher, D., Kujumdzieva, A. V. & Schmitt, M. J. Dual targeting of yeast catalase A to peroxisomes and mitochondria. *Biochem. J.* **380**, 393–400 (2004).
116. Kurita, O. Overexpression of peroxisomal malate dehydrogenase MDH3 gene enhances cell death on H<sub>2</sub>O<sub>2</sub> stress in the ald5 mutant of *Saccharomyces cerevisiae*. *Curr. Microbiol.* **47**, 0192–0197 (2003).
117. Giannattasio, S., Guaragnella, N., Corte-Real, M., Passarella, S. & Marra, E. Acid stress adaptation protects *Saccharomyces cerevisiae* from acetic acid-induced programmed cell death. *Gene* **354**, 93–98 (2005).
118. Davidson, J. F., Whyte, B., Bissinger, P. H. & Schiestl, R. H. Oxidative stress is involved in heat-induced cell death in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA.* **93**, 5116–5121 (1996).
119. Carman, G. M. & Han, G. S. Regulation of phospholipid synthesis in the yeast *Saccharomyces cerevisiae*. *Annu. Rev. Biochem.* **80**, 859 (2011).
120. Henry, S. A., Kohlwein, S. D. & Carman, G. M. Metabolism and regulation of glycerolipids in the yeast *Saccharomyces cerevisiae*. *Genetics* **190**, 317–349 (2012).
121. Roncoroni, M. *et al.* The yeast IRC7 gene encodes a  $\beta$ -lyase responsible for production of the varietal thiol 4-mercapto-4-methylpentan-2-one in wine. *Genetics* **28**, 926–935 (2011).
122. Santiago, M. & Gardner, R. C. The IRC7 gene encodes cysteine decaldehyde activity and confers on yeast the ability to grow on cysteine as a nitrogen source. *Yeast* **32**, 519–532 (2015).
123. Cordente, A. G. *et al.* Inactivating mutations in Irc7p are common in wine yeasts, attenuating carbonylsulfur  $\beta$ -lyase activity and volatile sulfur compound production. *Appl. Environ. Microbiol.* **85**, e02684 (2019).
124. Huang, C. W., Walker, M. E., Fedrizzi, B., Gardner, R. C. & Jiranek, V. Hydrogen sulfide and its roles in i in a winemaking context. *FEMS Yeast Res.* **17**, 058 (2017).
125. Mendoza-Cózatl, D., Loza-Tavera, H., Hernández-Navarro, A. & Moreno-Sánchez, R. Sulfur assimilation and glutathione metabolism under cadmium stress in yeast, protists and plants. *FEMS Microbiol. Rev.* **29**, 653–671 (2005).
126. Zhang, M. M. *et al.* Enhanced acetic acid stress tolerance and ethanol production in *Saccharomyces cerevisiae* by modulating expression of the de novo purine biosynthesis genes. *Biotechnol. Biofuels* **12**, 1–13 (2019).
127. de Lucena, R. M. *et al.* Transcriptomic response of *Saccharomyces cerevisiae* for its adaptation to sulphuric acid-induced stress. *J. Gen. Mol. Microbiol.* **108**, 1147–1160 (2015).
128. Auesukaree, C. *et al.* Genome-wide identification of genes involved in tolerance to various environmental stresses in *Saccharomyces cerevisiae*. *J. Appl. Genet.* **50**, 301–310 (2009).
129. Luttik, M. A. H. *et al.* The *Saccharomyces cerevisiae* ICL2 gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme a metabolism. *J. Bacteriol.* **182**, 7007–7013 (2000).
130. Verduyn, C., Postma, E., Scheffers, W. A. & Van Dijken, J. P. Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8**, 501–517 (1992).
131. Procópio, D. P. Metabolic engineering of *Saccharomyces cerevisiae* for second generation ethanol from Xylo-Oligosaccharides. (Universidade de São Paulo, 2021). <https://doi.org/10.11606/T.3.2021.tde-16032022-134640>.
132. Regenber, B. *et al.* Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol.* **7**, R107 (2006).
133. Olsson, L. & Nielsen, J. On-line and in situ monitoring of biomass in submerged cultivations. *Trends Biotechnol.* **15**, 517–522 (1997).
134. Della-Bianca, B. E., de Hulster, E., Pronk, J. T., van Maris, A. J. A. & Gombert, A. K. Physiology of the fuel ethanol strain *Saccharomyces cerevisiae* PE-2 at low pH indicates a context-dependent performance relevant for industrial applications. *FEMS Yeast Res.* **14**, 1196–1205 (2014).
135. Medina, V. G., Almering, M. J. H., Van Maris, A. J. A. & Pronk, J. T. Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor. *Appl. Environ. Microbiol.* **76**, 190–195 (2010).
136. Kammerer, D., Claus, A., Carle, R. & Schieber, A. Polyphenol screening of pomace from red and white grape varieties (*Vitis vinifera* L.) by HPLC-DAD-MS/MS. *J. Agric. Food Chem.* **52**, 4360–4367 (2004).
137. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
138. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
139. Encode. ENCODE Guidelines and Best Practices for RNA-Seq. Revised December 2016. 1–5 (2016). <https://doi.org/10.1101/044578>.
140. Howe, K. L. *et al.* Ensembl genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz890> (2020).
141. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
142. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, 605–612 (2021).
143. Luo, W., Pant, G., Bhavnasi, Y. K., Blanchard, S. G. & Brouwer, C. Pathview Web: User friendly pathway visualization and data integration. *Nucleic Acids Res.* **45**, W501–W508 (2017).
144. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-10-161> (2009).
145. Jia, C., Li, Y., Carson, M. B., Wang, X. & Yu, J. Node attribute-enhanced community detection in complex networks. *Sci. Rep.* <https://doi.org/10.1038/s41598-017-02751-8> (2017).
146. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/30.7.1575> (2002).
147. Van Dongen, S. & Abreu-Goodger, C. Using MCL to extract clusters from networks. *Methods Mol. Biol.* [https://doi.org/10.1007/978-1-61779-361-5\\_15](https://doi.org/10.1007/978-1-61779-361-5_15) (2012).
148. Wilczyński, B. & Dojer, N. BNFinder: Exact and efficient method for learning Bayesian networks. *Bioinformatics* **25**, 286–287 (2009).
149. Sturges, H. A. The choice of a class interval. *J. Am. Stat. Assoc.* **21**, 65–66 (1926).
150. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* <https://doi.org/10.1186/s13059-016-0974-4> (2016).



151. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. in *Proceedings of the 7th Python in Science Conference* (eds. Varoquaux, G., Vaught, T. & Millman, J.) 11–15 (2008).

### Acknowledgements

All analyses were performed using LaBIS Cloud. The authors would like to thank Espaço da Escrita (Pró-Reitoria de Pesquisa, UNICAMP) for the language services provided.

### Author contributions

F.E.C., M.M.B., T.O.B., D.P.P. and T.T.F. conceived and designed the research. D.P.P. and F.E.C. have equally contributed to this manuscript. D.P.P. conducted yeast chemostat cultivations and sample preparation. F.E.C. and N.F.M. performed computational biology analysis. F.E.C. wrote the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

### Funding

This project was funded by São Paulo Research Foundation (FAPESP) Grants #2018/17172-2, #2015/50612-8 and #2011/00417-3. F.E.C. was funded by CNPq Grant #141406/2018-6. D.P.P. was funded by FAPESP Grant #2018/01759-4. N.F.M. was funded by FAPESP Grant #2019/13946-6.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-26843-2>.

**Correspondence** and requests for materials should be addressed to T.O.B. or M.M.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022