Transcriptional response to lactic acid stress in the hybrid yeast *Zygosaccharomyces parabailii*

Raúl A. Ortiz-Merino[a], Nurzhan Kuanyshev[b,c], Kevin P. Byrne[a], Javier A. Varela[c], John P. Morrissey[c], Danilo Porro[b], Kenneth H. Wolfe[a], Paola Branduardi[b]#

UCD Conway Institute, School of Medicine, University College Dublin, Dublin, Ireland[a;]
Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milano, Italy[b];
School of Microbiology/Centre for Synthetic Biology and Biotechnology/Environmental Research Institute/APC Microbiome Institute, University College Cork, Cork, Ireland[c]

Running head: Hybrid yeast transcriptomics under lactic acid stress

#Address correspondence to Paola Branduardi, paola.branduardi@unimib.it
R.A.O.-M. and N.K. contributed equally to this work.

Abstract

Lactic acid has a wide range of applications starting from its undissociated form, and its production using cell factories requires stress-tolerant microbial hosts. The interspecies hybrid yeast *Zygosaccharomyces parabailii* has great potential to be exploited as a novel host for lactic acid production, due to high organic acid tolerance at low pH, and a fermentative metabolism with a fast growth rate. Here we used RNA-seq to analyze *Z. parabailii*'s transcriptional response to lactic acid added exogenously, and we explore the biological mechanisms involved in tolerance. *Z. parabailii* contains two homeologous copies of most genes. Under lactic acid stress, the two genes in each homeolog pair tend to diverge in expression to a significantly greater extent than in control conditions, indicating that stress tolerance is facilitated by interactions between the two gene sets in the hybrid. Lactic acid induces downregulation of genes related to cell wall and plasma membrane functions, possibly altering the rate of diffusion of lactic acid into cells. Genes related to iron transport and redox processes were upregulated, suggesting an important role for respiratory functions and oxidative stress defense. We found differences in the expression profiles of genes putatively regulated by Haa1 and Aft1/2, previously described as lactic acid-responsive in *Saccharomyces cerevisiae*. Furthermore, formate dehydrogenase (*FDH*) genes form a lactic acid-responsive gene family that has been specifically amplified in *Z. parabailii* as compared to other closely related species. Our study provides a useful starting point for the engineering of *Z. parabailii* as a host for lactic acid production.

Importance

Hybrid yeasts are important in biotechnology because of their tolerance to harsh industrial conditions. The molecular mechanisms of tolerance can be studied by analyzing differential gene expression in conditions of interest, and relating gene expression patterns to protein functions. However, hybrid organisms present a challenge to the standard use of mRNA sequencing (RNA-seq) to study transcriptional responses to stress, because their genomes contain two similar copies of almost every gene. Here we used stringent mapping methods and a high-quality genome sequence to study the transcriptional response to lactic acid stress in *Zygosaccharomyces parabailii* ATCC60483, a natural interspecies hybrid yeast that contains two complete subgenomes that are approximately 7% divergent in sequence. Beyond the insights we gained into lactic acid tolerance in this study, the methods we developed will be broadly applicable to other yeast hybrid strains.

Introduction

Species belonging to the *Zygosaccharomyces bailii sensu lato* clade have a remarkable resilience against stress induced by weak acids, some of which are widely used as food preservatives or are versatile chemical platforms (1, 2). Therefore, on the one hand these yeasts represent a challenging problem in the food industry because they are often found as contaminants in production pipelines for wine, high sugar products, and canned foods. On the other hand, they are promising cell factories for biotechnological applications involving organic acids that can be produced by microbial fermentation (3, 4) or released by lignocellulosic pretreatment of biomass (5).

Lactic acid is one of the useful organic acids that can be produced by yeasts as a microbial factory. This compound has a wide range of industrial applications including food preservation, additives and pharmaceuticals (6), and potential to be used for bioplastic production from a renewable source (7). Natural fermentation by lactic acid bacteria has long been the main source of industrial lactic

61 acid production (8). Despite important progress using lactic acid bacteria (9), some of these
62 organisms have complex nutritional requirements posing a negative impact on cost effectiveness and
63 on product purity (10). Moreover, operational costs are also increased by the need to convert lactate
64 to lactic acid, which is not the case with engineered yeast cells cultivated at pH well below the pK_a
65 of lactic acid (3.78) (7, 11). The production of several weak organic acids, including lactic acid, has
66 reached the industrial scale (4) but there is still room for further production improvement by
67 enhancing production host robustness and/or exploiting novel microbial hosts. Therefore,
68 understanding the mechanism of weak acid tolerance in non-*Saccharomyces* yeasts such as
69 *Zygosaccharomyces* is important for the future development of ultra-efficient production platforms
70 in which these yeasts are genetically engineered to produce lactic acid.
71
72 The mechanisms of weak acid stress tolerance and response have been studied extensively in the
73 model yeast *S. cerevisiae* (12-16). However, this knowledge is far from complete and cannot be
74 applied easily to non-*Saccharomyces* species. Previous research on tolerance to weak organic acids
75 revealed the capability of *Z. bailii sensu lato* to catabolize acetic and benzoic acids even in the
76 presence of glucose (17, 18). In addition, different *Z. bailii* strains display specific adaptation traits
77 such as the ability to modulate their cell wall and membrane composition in order to decrease the
78 influx of weak acids (19, 20).
79
80 Importantly, the *Z. bailii sensu lato* clade is characterized by substantial genetic diversity. Some
81 strains that were previously considered to be *'Z. bailii*' were reclassified in 2013 into two new
82 species called *Z. parabailii* and *Z. pseudobailii* (21). The name '*Z. bailii sensu lato*' is used to refer
83 to the species complex that includes these two new species as well as other strains that were not
84 reclassified (*Z. bailii sensu stricto*). The widely studied strains CLIB213^T and IST302 are *Z. bailii*
85 *sensu stricto* (22, 23). The strains ATCC60483 (used in this study) and ISA1307 are *Z. parabailii*,

which is a hybrid that was formed naturally by mating between *Z. bailii sensu stricto* and an unidentified *Zygosaccharomyces* species (24, 25). *Z. parabailii* genomes contain two copies of almost every gene, differing by 7% in nucleotide sequence on average (25). These genes are referred as homeologs because they are derived from different organisms; homeologs, or homoeologs, are a particular type of paralog (duplicated gene) (26).

We are exploring the possibility of using *Z. bailii sensu lato* species as alternative yeast hosts for lactic acid production. We focused on *Z. parabailii* strain ATCC60483 because our previous work demonstrated its high tolerance to lactic acid at low pH, characterized by growth without any detectable lag phase or acid consumption (20), under microaerobic conditions. These natural characteristics are promising in terms of possible exploitation for organic acid production and the potential to develop commercial strains will be enhanced when the molecular basis of its unusual tolerance to low pH, high inhibitor concentrations, and other traits of interest are clarified. As a preliminary step towards metabolic engineering, in this study we sought to investigate the molecular mechanisms of lactic acid tolerance in ATCC60483 by means of RNA-seq. In general, we found that the *Z. parabailii* transcriptome responds to lactic acid stress by inducing genes related to oxidative stress response and iron homeostasis in a different way than *S. cerevisiae* does. In addition, *Z. parabailii* modulates the transcription of genes related to the cell wall, in agreement with our previous data.

Results

Transcriptional profile of *Z. parabailii* homeolog pairs and duplicated genes in lactic acid stress

In our previous study we found that *Z. parabailii* ATCC60483, when treated with 40 g L^{-1} of lactic acid during microaerobic fermentation of glucose, neither consumes the lactic acid nor exhibits a

111 reduction in cell viability, although we could observe acid-induced phenotypic and morphological
112 changes (20). During the fermentation, both control and lactic acid treated cells consumed glucose
113 and produced ethanol as the main metabolite. No other fermentation metabolites were detected at the
114 end of fermentation. The cells treated with lactic acid showed a 25% reduction in growth rate and a
115 15% reduction in final biomass titer. In addition, the specific glucose consumption rate was 13%
116 lower than in a control condition. The yield of ethanol at the end of fermentation was similar in both
117 conditions (20). Our aim here was to study the transcriptional response of *Z. parabailii* when
118 exposed to a high concentration of lactic acid.
119
120 We compared the transcriptomes of *Z. parabailii* ATCC60483 cultures grown in the presence or
121 absence of lactic acid (40 g L^{-1}), at time points (18 h and 42 h) specifically chosen to ascertain
122 comparable growth kinetics and exclude growth phase related bias (**Fig. 1**). After normalizing and
123 filtering the raw RNA-seq read counts, we detected expression for >95% of the *Z. parabailii* genes
124 in at least one condition, including 36 genes that were transcribed only in lactic acid and 31 that
125 were transcribed only in control conditions (**Table 1, Table S1** and **Table S2**).
126
127 We used stringent mapping of RNA-seq reads to the genome (see Methods), in order to capture
128 expression differences between homeologous gene pairs even when they are highly similar in
129 sequence. About 82% of the 10,072 genes in the *Z. parabailii* nuclear genome show the pattern
130 characteristic of hybrid genomes, forming pairs of 'A' and 'B' homeologs, where the A-gene came
131 from one parent in the hybridization and the B-gene came from the other (25). Most of the
132 remaining loci in the genome are also present in two copies, but are either A:A or B:B pairs due to
133 loss of heterozygosity after hybridization (25). We calculated the ratio of expression between the A
134 and B homeologs for each of 4136 A:B gene pairs as described in Methods (**Fig. 2**). All but 21 gene
135 pairs showed evidence of expression of both homeologs.

136

137 Strikingly, the distribution of expression ratios is broader in lactic acid than in control conditions, at
138 both time points. In other words, in the stress condition one of the two genes in each homeolog pair
139 tends to become predominantly expressed. If we define unbalanced expression as an expression ratio
140 that lies outside the range 0.4-0.6 (**Fig. 2**), the proportion of homeolog pairs with unbalanced
141 expression is 13.8-18.7% in the control conditions but increases to 31.0-33.4% in lactic acid
142 conditions. The difference in variance of expression ratios is statistically significant (Fligner-Killeen
143 test; P = 3e-98 at 18 h, and P = 9e-61 at 42 h).

144

145 The distribution of expression ratios is approximately symmetrical (**Fig. 2**) indicating that in some
146 homeolog pairs the A-gene is more highly expressed than the B-gene (ratios > 0.5) whereas in others
147 the B-gene is higher (ratios < 0.5). The A-genes were derived from the parental species *Z. bailii* in
148 the hybridization, and the B-genes were derived from the other parent (an unidentified
149 *Zygosaccharomyces* species) (25). Thus, broadly speaking the cell responds to lactic acid stress by
150 inducing greater divergence of expression between the genes in a homeolog pair, without a strong
151 preference as to whether the A-gene or the B-gene is the higher-expressed one. There is a trend
152 towards higher expression of the A-genes, as illustrated by the larger numbers of loci for which the
153 expression level of the A-gene exceeds that of the B-gene, as opposed to the converse (**Table 2**, Ab
154 and aB columns). Statistical analysis indicates a weak bias towards A-genes, showing slight but
155 consistent negative skew values, but this bias is only significant in the cells treated with lactic acid
156 for 18h (**Table 2**). In summary, *Z. parabailii* has a slight tendency to express its A-genes more
157 highly than its B-genes, and this tendency is maintained under lactic acid stress, but the magnitude
158 of this tendency is small compared to the grossly increased divergence of expression levels between
159 homeologs that occurs in lactic acid stress.

160

Our method of high-stringency mapping of RNA-seq reads to a high-quality genome sequence detected transcriptional profiles of homeologous gene pairs even where the gene pairs were highly similar. Nevertheless, we needed to modify it to determine read counts for genes that occur in identical pairs (See Methods). Specifically, among the 402 genes that have no evidence of expression in the "full" count dataset, 42% are genes that were affected by loss of heterozygosity (genes in A:A pairs or B:B pairs). The modified method enabled us to measure the combined expression of 230 duplicated genes in *Z. parabailii* including the two orthologs of the *S. cerevisiae* major mitochondrial D-lactate dehydrogenase *DLD1* (*I04780_A* and *N05010_A*, which are identical) and the minor isoform *DLD2* (*B01190_N* and *G05430_N*, which are identical). Although we cannot investigate the expression of each of these identical gene pairs individually, their RPKM (Reads Per Kilobase of transcript per Million mapped reads) values are low when compared with all duplicated genes on lactic acid conditions. For example, the combined RPKM values on lactic acid at 18 h for the two *DLD1* genes was 340.2, and for the two *DLD2* genes was 710.5, while the average RPKM values for duplicated genes on this condition and time point was 4028. Furthermore, *DLD1* shows a statistically significant 2-fold expression decrease in lactic acid at both timepoints whereas *DLD2* shows no significant expression changes (**Data Set S4**). These observations are consistent with the previously reported lack of lactic acid consumption of *Z. parabailii* (20).

Upregulated genes are related to oxidation-reduction processes and ion transport in the mitochondria.

The "full" set of counts (See Methods) for the 9683 genes in the union set of expressed genes (Table 1) were then used for differential expression analysis, filtering the results for adjusted P value < 0.05 and $|\log_2\text{-fold change}| \geq 1$ (**Data Set S5**). This analysis is independent from that of the duplicated genes mentioned above, and identified a total of 227 genes upregulated in lactic acid, of which 117 are specific to 18 h and 83 to 42 h (**Table 3**). Similarly, a total of 1019 downregulated genes were

found, including 430 specific to 18 h and 431 to 42 h. We then performed a Gene Ontology (GO) term enrichment analysis to identify GO terms that were enriched at both time points in either the upregulated genes (**Fig. 3A**) or the downregulated genes (**Fig. 3B**).

When we use an *S. cerevisiae* gene name in the following functional analysis, we refer to either one or both of its orthologs in a *Z. parabailii* homeolog pair. These genes are included in our functional analysis if at least one of the members of a *Z. parabailii* homeolog pair was differentially expressed.

The enriched GO term associated with the highest number of genes upregulated by lactic acid in our dataset is GO:0055114 for "oxidation-reduction process" (**Fig. 3A**, **Data Set S6**). This term is associated with 33 genes including homologs of the *S. cerevisiae* genes *GOR1*, *AIM17*, *CCP1*, *MET13*, *SOD2*, *SOD1*, *GND1/2*, and *GRX1/2*, some of which are also related to enriched mitochondrial terms (GO:0005758 for example). We also observed enrichment for genes in the glyoxylate cycle (GO:0006097) and the glyoxysome (GO:0009514) including homologs of *ICL1* and *IDP2*. The upregulation of these genes along with *FBP1* could indicate activation of the anaplerotic reactions, probably caused by oxygen limitation. *ACH1* with CoA transferase activity, and *Z. parabailii* gene *L05300_N* (predicted to have epoxide hydrolase activity), were upregulated at both time points and are presumably involved in enzymatic detoxification process.

The siderophore transmembrane transport term (GO:0044718) was also found enriched in upregulated genes. Genes in this category are members of the MFS_1 family of transporters, potentially involved in iron retention and/or transport (genes *A10040_B*, *B02380_A*, *G04250_B*, *I00120_N*, *I05800_A*, *O00120_N*), and upregulated at 18 h. These genes are all classified as integral components of the membrane. Other genes specifically upregulated at 42 h include *FIT2*, *STL1* and

210 *K05040_N* which shows no sequence homology to *S. cerevisiae* genes but is predicted to be a
211 transmembrane transporter (see Methods).
212
213 **Downregulated genes are mainly related to components of the cell boundaries and protein**
214 **translation.**
215 The GO term enrichment analysis for downregulated genes showed 63 genes related to ribosomal
216 functions (GO:0003735), and 38 to cytoplasmic translation (GO:0002181) (**Fig. 3B**, **Data Set S6**).
217 Most of those genes were downregulated at 42 h, implying a general decrease in protein synthesis.
218 This response seems to correspond to a general mechanism observed also in other yeasts used as cell
219 factories, *e.g. S. cerevisiae* under stress conditions (27) and *Komagataella phaffii* (*Pichia pastoris*)
220 used for heterologous protein production induced by methanol (28) during stress, possibly related to
221 resilience or energy maintenance. Some of these genes are also related to the enriched terms
222 GO:0000932 and GO:0010494 for P-body and cytoplasmic stress granules involved in mRNA
223 translation and turnover during different stress conditions in *S. cerevisiae* (29). These categories are
224 downregulated at 18 h in lactic acid treated cells. One of the components of stress granules that is
225 also downregulated at 42 h codes for a homolog of *S. cerevisiae* Pab1, the major polyA binding
226 protein which has been demonstrated to promote the formation of stress granules (30). A recent
227 study conducted in *S. cerevisiae* reported that stress granules are not formed in lactic acid treated
228 cells (31) and a similar situation might be also true for *Z. parabailii*.
229
230 Among the downregulated genes, we also identified many with functions that we summarize as
231 being related to the boundaries of the cell, *i.e.* to the cell wall and the plasma membrane. The GO
232 terms in this group include the actin cortical patch (GO:0030479), cell cortex (GO:0005938),
233 extracellular region (GO:0005576), fungal-type cell wall (GO:0009277), and structural constituents
234 of the cell wall (GO:0005199) (**Fig. 3B**). Consistent with this, we noticed enrichment of the GO

terms for glucan endo-1,3-beta-D-glucosidase activity (GO:0042973) and chitin binding (GO:0008061). These observations indicate that the cell wall is modulated upon lactic acid stress, in agreement with our previous findings (20). Other genes downregulated at 42 h predicted to be integral components of the membrane are *H01670_B* with unknown function, *OPT1*, and *HBT1*. We also found downregulation of *CWP1*, a cell wall protein homolog, and *LDS2*, which is involved in the assembly of the *S. cerevisiae* spore wall.

Involvement of Haa1 and Aft1/Aft2 regulated genes in lactic acid stress response

Previous studies on lactic acid stress response mechanisms in *S. cerevisiae* indicated an important role of the transcription factors Haa1 and Aft1/Aft2 (32, 33). Therefore, we extracted all the *S. cerevisiae* genes reported to be targets of either Haa1 or Aft1/Aft2 in YEASTRACT (34), in addition to those identified as lactic acid-responsive (32). We then tested whether the *Z. parabailii* orthologs of these *S. cerevisiae* genes are differentially expressed in our dataset. In this case we ignored the $\log_2$-fold change cut-off to enable detection of small but still significant changes. The results are shown in **Data Set S7**.

We found differential expression of 42 orthologs of *S. cerevisiae* genes putatively controlled by Haa1 (**Fig. 4A**). These include the membrane-bound and major weak acid response genes *YPC1/YDC1*, *TPO2/3*, *VPS62/TDA6*, *PDR16*, and *PDR12*. This set also includes the transcription factors *MSN4/2*, *COM2*, and the transcription factor itself (*HAA1/CUP2*). Interestingly, we observed the major weak acid stress response genes, *TPO2/TPO3* and *SPS100/YGP1* to be downregulated in *Z. parabailii*, although they are upregulated during lactic acid stress in *S. cerevisiae* (32). Here we also found *PFK27* with a response changing from upregulated at 18 h to downregulated at 42 h, and *MTH1/STD1* going from downregulated at 18 h to upregulated at 42 h. These changes in glucose-responsive genes could possibly reflect the diauxic shift.

We performed a similar search strategy for genes putatively under the control of Aft1/Aft2 which in *S. cerevisiae* are related to iron utilization and homeostasis (35). Results are shown in **Fig. 4B** and **Data Set S7**. This detected changes in expression of orthologs of 27 *S. cerevisiae* genes, of which 6 are downregulated at both time points: *AFT1/AFT2* coding for the transcription factor itself; *AKR1/AKR2*, an integral component of the membrane with palmitoyltransferase activity; *LEU2*, involved in leucine biosynthesis; *MRS3/MRS4*, iron transporters; *APE1*, with metalloaminopeptidase activity; and *AHP1*, a thiol-specific peroxiredoxin. The rest of these genes are upregulated, at both timepoints, and include: *TIS11/CTH1*, involved in mRNA processing; *CCC2*, a Cu^{++}-transporting P-type ATPase; *UBC8*, which negatively regulates gluconeogenesis; *ECL1*, which increases oxygen consumption and respiratory activity; *SMF3*, a putative divalent metal ion transporter; *ARA2*, a NAD-dependent arabinose dehydrogenase; *FET3*, a Ferro-O_2-oxidoreductase; and *PEP4*, a vacuolar protease. Most of these upregulated genes are related to ion transport and redox functions, in agreement with our GO term enrichment analysis.

Multigene families significantly modulated upon lactic acid exposure

We identified an unusual regulatory pattern in a family of genes related to *S. cerevisiae FDH1*, which codes for formate dehydrogenase. This enzyme is known to be induced upon formate exposure in *S. cerevisiae* and is widely found in methylotrophic yeasts (36, 37). Recent studies showed that the Fdh1 enzyme contributes to oxidative stress resistance in bacteria (38, 39). The *Z. parabailii* genome contains six genes in this family (*I01900_B*, *O01850_A*, *P02220_N*, *H05680_N*, *N02280_N*, and *F04070_N*) although formate dehydrogenase activity has not been demonstrated for any of them. All six *FDH*-like genes were highly upregulated at 18 h of lactic acid exposure. *P02220_N* is lactic acid-specific (**Table S2**), and *N02280_N* and *F04070_N* were also significantly downregulated at 42 h (**Data Set S5**). Formate dehydrogenases perform the NAD^+-dependent

oxidation of formate to carbon dioxide. The *S. cerevisiae* strain CEN.PK 113-7D contains two *FDH* genes (*FDH1* and *FDH2*), whereas only *FDH1* is intact in the laboratory strain BY4741 because *FDH2* is truncated (37). The function of *FDH* genes in *S. cerevisiae* is not well characterized, but these enzymes have been better studied in methylotrophic yeasts such as *Komagataella phaffi* where they are involved in the last step of the methanol dissimilation pathway (36).

Interestingly, the phylogenetic distribution of *FDH* genes among sequenced yeast genomes is rather patchy (36) and indicative both of recent gene amplifications and of multiple gene losses. We searched for *FDH* homologs in the NCBI databases and constructed a phylogenetic tree (**Fig. 5**). Many yeast species lack *FDH* genes completely, containing only homologs of distantly related genes such as *GOR1* (glyoxylate reductase). Nevertheless, the phylogenetic relationship among the *FDH* genes of the few species that retain this gene agrees well with the expected relationship among these species (**Fig. 5**). This observation suggests that the patchy distribution is due to numerous losses of an ancestral *FDH* gene (for example, in the genera *Torulaspora*, *Lachancea* and *Kluyveromyces*), and not the result of horizontal gene transfer. There is essentially no conservation of synteny among the existing *FDH* genes, which shows that multiple species-specific gene duplications and gene relocations have occurred. Of the six *Z. parabailii FDH*-like genes, four are closely related and form a phylogenetic cluster with *Saccharomyces* species (**Fig. 5**). The other two form a cluster with the only *FDH*-like gene we identified in the genome of CLIB213^{T}, a *Z. bailii sensu stricto* strain. The sister species *Z. rouxii* has four *FDH*-like genes that cluster together in the tree. Thus, amplifications of *FDH*-like genes by gene duplication have occurred separately in *Z. parabailii* and *Z. rouxii*, and in the former species they are highly induced by lactic acid. This difference in *FDH* gene copy number between *Z. parabailii* and *Z. bailii* may be a contributory factor to the difference in their tolerance to lactic acid as our previous study showed that *Z.*

Applied and Environmental Microbiology

309 *parabailii* ATCC60483 is more resilient to lactic acid than the *Z. bailii sensu stricto* strains
310 ATCC8766 and ATCC58445^T (synonymous with CLIB213^T) (20).
311
312 We searched systematically for other *Z. parabailii* genes assigned into multigene families, which
313 have significant expression changes in lactic acid. This was done by searching for sets of three or
314 more *Z. parabailii* genes that share the same *Z. bailii* ortholog. We examined a total of 123 *Z.*
315 *parabailii* genes in multigene families of this type, of which 22 are differentially expressed in at
316 least one timepoint (classified as category 9 in **Data Set S5**). These 22 genes belong to 12 different
317 multigene families significantly modulated in lactic acid. For example, *F06230_N*, *N00190_A* and
318 *O04100_A* are homologs of the *FFZ2* transporters, which are specific to *Zygosaccharomyces* species
319 and able to transport fructose and glucose when overexpressed in *S. cerevisiae* (40, 41). In this
320 family, *F06230_N* is upregulated in lactic acid at both time points whereas *N00190_A* is upregulated
321 only at 18 h, and *O04100_A* did not show significant expression changes. Another interesting family
322 is *A10020_N*, *G00240_N* and *P00180_N* which are all lactic acid-specific (**Table S2**), significantly
323 upregulated in lactic acid (when ignoring the $\log_2$-fold change cut-off) and are homologous to the
324 iron siderophore transporter *FIT2* putatively under the control of Aft1/2 (**Fig. 4B, Data Set S7**).
325 This family also includes *K00140_A* and *C00210_N* for which we did not observe any evidence of
326 expression. Furthermore, given that *K00140_A* is identical to the only *FIT2* homolog annotated in *Z.*
327 *bailii* strain CLIB213^T (BN860_19394g1_1), and it is not differentially regulated, the *Z. parabailii*-
328 specific genes in this family may have functional relevance.
329
330 **Discussion**
331 We aim to engineer a yeast strain able to produce lactic acid which, at high concentrations, is toxic
332 to the cells. Our results indicate that, in general terms, *Z. parabailii* counteracts the toxicity of lactic
333 acid by modulating its oxidation-reduction processes and the composition of its cell boundaries.

Although some of these responses overlap with *S. cerevisiae*'s, *Z. parabailii* additionally shows an interplay between its two homeologous gene sets, and utilizes expanded multigene families. Validation of these observations awaits the development of better molecular tools for manipulation of *Z. parabailii*, but our work nevertheless represents a significant step towards engineering this non-conventional yeast to produce lactic acid.

The toxicity of lactic acid, and weak acids in general, involves dissipation of the pH gradient at the plasma membrane (42). Other secondary effects result from the intracellular accumulation of the weak acid counteranions. For example, acetate has been shown to trigger programmed cell death and an increase in the formation of reactive oxygen species (ROS) (12), while sorbate affects the membrane structure (43). Microorganisms have developed different mechanisms to tolerate these toxic effects. For example, *S. cerevisiae* responds to weak acids by using H^+-ATPases to control intracellular pH (44). In *Pichia anomala*, a higher tolerance is achieved by coupling H^+-ATPases with increased mitochondrial ATP production (45). *Candida krusei* also has higher tolerance to lactic and acetic acid than *S. cerevisiae*, postulated to involve a quicker H^+-ATPase response (46). *Z. parabailii* shares this ATP-dependent tolerance response but has some unique features, connected to the hybrid nature of its genome, as discussed below.

We observed upregulation of genes related to detoxification of ROS which could be linked to the upregulation of the respiratory chain and the glyoxylate cycle. Lactic acid stress has been reported to imbalance the prooxidant/antioxidant ratio (47), and trigger the accumulation of ROS via the Fenton reaction (48). Accordingly, overexpression of cytosolic catalase or introduction of the pathway for biosynthesis of L-ascorbic acid (a well-known antioxidant) into *S. cerevisiae* improved resistance to oxidative and lactic acid stress (49, 50). The alleviating effect of antioxidants indicates the importance of controlling the concentration of H_2O_2, which can catalyze the conversion of

359 glyoxylate into formate and CO_2 (51). The upregulation of the glyoxylate cycle, combined with
360 respiratory chain, would then result in the production of H_2O_2 and formate. The expansion and
361 upregulation of the *FDH* multigene family in *Z. parabailii* would then serve to convert the otherwise
362 toxic formate into NADH and CO_2. This mechanism was described in the bacterium *Pseudomonas*
363 *fluorescens* as an anti-oxidative defence mechanism (39, 52) and we speculate that the multiple Fdh
364 enzymes in *Z. parabailii* might serve a similar role.
365
366 There are significant differences between the response to lactic acid that we observed in *Z.*
367 *parabailii* and the responses previously reported in *S. cerevisiae* (32, 33). While many of these
368 differences may reflect differences in the physiology of the two yeasts, there were also differences
369 in the experimental setup used. We used microaerobic conditions, whereas previous studies used
370 anaerobic chemostat conditions (32), and batch flask fermentation (33). Nevertheless, we also
371 identified some similarities between the lactic acid responses in *S. cerevisiae* and *Z. parabailii*,
372 involving iron homeostasis genes such as siderophore transporters and iron transporters. In *S.*
373 *cerevisiae* a high concentration of lactate ions in the growth medium chelates free iron, reducing its
374 availability for cellular functions (32), and triggering a strong regulation of iron homeostasis (32,
375 33). We observed a similar response to lactic acid stress in *Z. parabailii*.
376
377 We found that *Z. parabailii* appears to modulate its cell wall in response to lactic acid stress. The
378 cell wall is generally considered to be a barrier for large molecules (53, 54). Nevertheless, studies on
379 *S. cerevisiae* have reported changes in expression of genes coding for cell wall components (55), or
380 related to cell wall integrity (56), in response to acetic acid or a low pH environment (57). In *Z.*
381 *parabailii*, the downregulation of cell wall related genes we observed in this study can be linked to
382 the decrease of cell wall mannoprotein and β 1→3 glucan levels that we previously found by FTIR

383 analysis (20). Together with the peculiar plasma membrane composition (19), these changes in the
384 cell wall may contribute to the superior lactic acid tolerance of *Z. parabailii*.
385
386 The expression of Haa1-regulated genes during stress in *Z. parabailii* is rather different from *S.*
387 *cerevisiae*. Haa1 has been reported to be a transcriptional activator of genes in response to both
388 acetic acid and lactic acid in *S. cerevisiae* (32, 58, 59), and in response to acetic acid in *Z. bailii*
389 (lactic acid was not investigated) (23, 60). It is intriguing to observe a different expression pattern
390 for those genes in *Z. parabailii* during lactic acid stress, but further studies will be necessary to fully
391 characterize the divergence of the roles of the Haa1 orthologs in the two species.
392
393 We found that lactic acid stress induces robust and statistically significant divergent expression
394 responses between the homeologous gene pairs in *Z. parabailii.* These differences need to be further
395 explored when considering differentially expressed genes as engineering targets, but the overall
396 stress response we saw among them is striking. Homeologous gene pairs are present in all hybrid
397 (allopolyploid) organisms (26). Most previous transcriptomic analyses involving homeologous pairs
398 have been carried out in plant species (61-63), although there are examples with fungi (64) and
399 yeasts (65, 66). We are not aware of any previous studies that found a similar genome-wide increase
400 in homeolog expression divergence under stress conditions. Our study differs from the previous
401 work on yeast hybrids because we examined gene expression in a natural hybrid isolate, whereas
402 preceding studies analyzed synthetic hybrids (65, 66). Furthermore, we compared expression
403 between homeolog pairs under two different growth conditions, whereas previous comparisons were
404 done against the parental genes (65, 66), even when using more than one condition (65).
405
406 Our study is a pioneering approach to examining the transcriptome of a hybrid yeast. It was made
407 possible by the availability of a high quality reference genome sequence (25) which is often not

available for other hybrid organisms. It also required highly-stringent and tailored methods to measure the expression of highly similar genes and even identical copies. We showed that homeologous gene pairs have different expression patterns when subjected to acid stress, which could reflect or override transcriptional control mechanisms inherited from the parents of this hybrid. This hybrid nature is one of a few differences we observed in comparison with the lactic acid responses reported for *S. cerevisiae* and *Z. bailii*. Our observations need further experimental validation given that changes in transcript levels are not always reflected in protein activities *in vivo*. Nevertheless, our observations that the duplicated homologs of *DLD1* and *DLD2* are expressed only at a low level in lactic acid, and *DLD1* is even repressed relative to control conditions, are consistent with the absence of lactic acid consumption by *Z. parabailii* in these conditions (20), which is a key feature needed for a lactic acid producing host. Our study provides methods and data to facilitate the understanding of molecular responses during acid stress in this or other hybrid yeasts, which is important both for fundamental and applied science.

Materials and Methods

Cell growth, RNA extraction and sequencing

Z. parabailii strain ATCC60483 was used for bioreactor fermentation. Cell aliquots, stored at -80°C in YPD glycerol stock, were grown to mid exponential phase before being inoculated to the bioreactor at final absorbance of OD_{660} 0.1. We used 2x Verduyn growth medium (67) at pH 3 containing 40 g L^{-1} glucose with 40 g L^{-1} lactic acid or no lactic acid (control condition). The fermentations were performed in 2 L volume bioreactors (BIOSTAT B, Sartorius AG, Germany) with operative volume of 1.5 L. The temperature was maintained at 30°C, pH at 3 by the addition of 4 M NaOH and the stirrer speed was set to 400 rpm. The inlet gas flow was adjusted by two mass flow controllers (Bronkhornst®High Tech- EL-FLOW®Select). The mass flow was set to obtain a

mixture of N_2 and air with final concentration of inlet oxygen of 5%. The mixture was sparged at 0.75 vvm. Antifoam (Antifoam 204, Sigma Aldrich) was used for foaming control.

The samples for RNA sequencing were taken in triplicate at 18 h and 42 h from the bioreactor fermentation, corresponding to log phase and post diauxic shift, respectively (20). Total RNA was extracted using Zymo Research Fungal/Bacterial RNA MiniPrep™ kit (Irvine, USA) and the quality of RNA samples were evaluated with an Agilent Bioanalyzer. The RNA samples were sequenced using the Illumina HiSeq2000 platform with 100 nt-long paired-end reads at Parco Tecnologico Padano (Lodi, Italy).

RNA-seq analysis

We used our *Z. parabailii* ATCC60483 genome annotation as a reference (25). This annotation consists of 10,072 nuclear and 13 mitochondrial protein-coding genes obtained using an improved version of the Yeast Genome Annotation Pipeline (68), and includes additional metadata as an aid for functional interpretation. Briefly, because of its hybrid nature, the *Z. parabailii* genome contains two homeologous copies of most genes. We use suffixes _A and _B in gene names to indicate the two copies, where _A indicates gene copies that are virtually identical to their *Z. bailii sensu stricto* orthologs, and _B indicates copies that are more divergent (5-25% synonymous sequence divergence). A few genes have the suffix _N because they could not be assigned to either of these two groups. There are 4139 homeologous A:B gene pairs in the annotated genome sequence, but for 3 pairs we did not detect transcription of either of the genes in any condition, so these 3 pairs were not analyzed further.

Some extra information (**Data Set S1**) was added to the original annotation, including functional domains and protein family memberships, which were obtained by aligning all the *Z. parabailii* ATCC60483 amino acid sequences against the PFAM database (69) using HMMER v. 3.0 (70). A

457 genome-wide annotation of transmembrane proteins was also made by comparing the *Z. parabailii*
458 proteome against the TransportDB 2.0 (71) database using BLAST v. 2.2.22 (72). The sequences
459 were then filtered based on identity (>35 %) and coverage (>80%) and submitted to the TMHMM
460 server v. 2.0 (73) to determine a minimum of 2 potential transmembrane domains per sequence.
461 Blast2GO (74) was then used to generate a custom Gene Ontology (GO) annotation for *Z. parabailii*
462 (**Data Set S2**).
463
464 The raw RNA-seq reads were mapped against the *Z. parabailii* ATCC60483 nuclear and
465 mitochondrial genomes (25) using bowtie v1.1.2 (75) with the parameters -v 0 -k 10 --best -M 1.
466 The parameter -v 0 gives high stringency by allowing no mismatches in the alignments
467 discriminating between highly similar regions in the genome, and discarding reads with sequencing
468 artefacts. The parameters -k 10 --best -M 1 report only the best possible alignment out of up to 10
469 alternatives and, in case there are two equivalent best hits, only one is reported at random. This
470 procedure reports multi-mapping reads with the tag "XM:i:2" and a mapping quality (MAPQ) equal
471 to 0.
472
473 The mapped reads were subsequently counted using htseq-count v0.6.0 (76) with two different
474 settings. In the first case, htseq-count was applied to the full set of *Z. parabailii* genes with default
475 parameters, to generate what we refer to as "full" counts. This setting discards the alignments for
476 multi-mapping reads, because their quality is artificially set to the lowest possible value. Then, to
477 obtain data from pairs of identical genes, a different htseq-count run was performed with the
478 parameter -a 0 allowing for MAPQ >= 0. To avoid spurious low-quality alignments, this second run
479 used only the alignments with the "XM:i:2" tag, and was applied only to a set of 232 duplicated
480 genes that have 100% sequence identity (measured using blastn; 72), over their full length, to one or
481 more other *Z. parabailii* genes. The counts from this second htseq-run for duplicated genes with

multi-mapping reads are referred as "duplicated" counts. The duplicated counts represent a composite signal from two or more identical genes, and potential different quality values, which is not the case for the "full" counts. Therefore, the two sets of counts were analysed separately. All the counts reported are "full" counts unless stated otherwise.

The RNA-seq read counts were split in 4 groups according to condition and time point, each group containing 3 libraries. One of the libraries for the control condition at 18 h contained few reads (5.9 million, compared to the average of 30.5 million from the other libraries) and was excluded from further analyses. We therefore used the TMM method (77) implemented in edgeR v. 3.18.1 (78) to normalize the read counts and provide better comparability across the different sized samples. Counts per million (CPM) were calculated from the normalized counts using edgeR. Genes with less than 1 CPM in at least 3 samples from the same condition were considered to have no evidence of expression. We also calculated Reads Per Kilobase of transcript per Million mapped reads (RPKM) using edgeR. This was done for the normalized and filtered sets of "full" counts for 4136 homeolog pairs, and for the "duplicated" counts for the 232 duplicated genes.

An expression ratio index was calculated for the 4136 A: B homeolog pairs for which at least one gene in the pair showed evidence of expression. There are 4139 homeologous A:B gene pairs in the annotated genome sequence, but for 3 pairs we did not detect transcription of either of the genes in any condition, so these 3 pairs were not analyzed further. The expression ratio is calculated as: *Expression ratio* = $avg\ RPKM_A / (avg\ RPKM_A + avg\ RPKM_B)$ where the subscripts A and B indicate the parental origin of each gene. This index ranges from 0 to 1, with 0.5 meaning equal expression of the A and B homeologs. RPKM values for each homeolog in a pair, averaged among replicates, are given in **Data Set S3**. We calculated descriptive statistics from the expression ratio for the different groups using the R package psych v. 1.7.5 (79). Exact binomial tests and Fligner-Killeen

tests were performed using the R functions binom.test and fligner.test. The R function p.adjust was used for Bonferroni correction of P values for multiple testing.

The normalized and filtered datasets were Voom-transformed (80) to consider the differences in count sizes (or sequencing depth) and the overall dataset variability. This was followed by differential expression analysis (DEA) with adjusted P value < 0.05 and $|\log_2\text{-fold change}| \geq 1$ for statistical significance (**Data Set S4** for the "full" set; **Data Set S5** for the "duplicated" set). Both the Voom transformation and the differential expression analysis were done using Limma v. 3.32.2 (81). The *Z. parabailii* GO annotation was utilized for GO term enrichment analysis with the R package goseq v. 1.28.0 (82), applied to the 3 sets of upregulated genes (18 h, 42 h, and both time points), as well as to the corresponding 3 sets of downregulated genes (**Data Set S6**). The output of goseq was visualized using UpsetR v. 1.3.3 (83).

Phylogenetic analysis of formate dehydrogenase sequences

Protein sequences of homologs of the *Z. parabailii* Fdh-like proteins were identified by blastp (72) searches against the non-redundant protein sequence database of the National Center for Biotechnology Information (NCBI) with default parameters. The search was restricted to yeast species (taxid:4932). Representative sequences were selected from the species indicated in Figure 5. For yeast species that lacked an apparent Fdh, we retained the next-most similar protein instead, which in all cases had higher similarity to *S. cerevisiae* Gor1, another protein with an NAD(P)-binding domain. *Escherichia coli* Fdh1, which is more closely related to yeast Fdh1 than to yeast Gor1, was included for reference. Alignments and phylogenetic trees were generated in SeaView v4 (84) with the included versions of Clustal Omega (85) and PhyML (86) using 100 bootstrap replicates. The output was visualized using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Nucleotide sequence accession numbers.

The RNA-seq data reported here have been deposited in NCBI's Gene Expression Omnibus (87) with accession number GSE104654. The submitted data include the RNAseq fastq files, counts for the "full" and "duplicated" sets (both raw and normalized), and RPKMs for the duplicated genes.

Acknowledgments

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° 606795. P.B. and D.P. acknowledge support by FAR (Fondo di Ateneo per la Ricerca) of the University of Milano-Bicocca. R. A. O.-M. was partially supported by CONACyT, Mexico (fellowship number 440667).

The authors declare that there is no conflict of interest.

References

1. **Martorell P, Stratford M, Steels H, Fernandez-Espinar MT, Querol A.** 2007. Physiological characterization of spoilage strains of Zygosaccharomyces bailii and Zygosaccharomyces rouxii isolated from high sugar environments. Int J Food Microbiol **114:**234-242.
2. **Stratford M, Steels H, Nebe-von-Caron G, Novodvorska M, Hayer K, Archer DB.** 2013. Extreme resistance to weak-acid preservatives in the spoilage yeast Zygosaccharomyces bailii. Int J Food Microbiol **166:**126-134.

3. **Kuanyshev N, Adamo GM, Porro D, Branduardi P.** 2017. The spoilage yeast Zygosaccharomyces bailii: Foe or friend? Yeast **34:**359-370.

4. **Becker J, Lange A, Fabarius J, Wittmann C.** 2015. Top value platform chemicals: bio-based production of organic acids. Curr Opin Biotechnol **36:**168-175.

5. **Limayem A, Ricke SC.** 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. Progress in Energy and Combustion Science **38:**449-467.

6. **Castillo Martinez FA, Balciunas EM, Salgado JM, Domínguez González JM, Converti A, Oliveira RPdS.** 2013. Lactic acid properties, applications and production: A review. Trends in Food Science & Technology **30:**70-83.

7. **Sauer M, Porro D, Mattanovich D, Branduardi P.** 2010. 16 years research on lactic acid production with yeast - ready for the market? Biotechnol Genet Eng Rev **27:**229-256.

8. **Datta R, Henry M.** 2006. Lactic acid: recent advances in products, processes and technologies — a review. Journal of Chemical Technology & Biotechnology **81:**1119-1129.

9. **Ma K, Maeda T, You H, Shirai Y.** 2014. Open fermentative production of L-lactic acid with high optical purity by thermophilic Bacillus coagulans using excess sludge as nutrient. Bioresour Technol **151:**28-35.

10. **Fitzpatrick JJ, Murphy C, Mota FM, Pauli T.** 2003. Impurity and cost considerations for nutrient supplementation of whey permeate fermentations to produce lactic acid for biodegradable plastics. International Dairy Journal **13:**575-580.

11. **Chen Y, Nielsen J.** 2016. Biobased organic acids production by metabolically engineered microorganisms. Curr Opin Biotechnol **37:**165-172.

12. **Giannattasio S, Guaragnella N, Zdralevic M, Marra E.** 2013. Molecular mechanisms of Saccharomyces cerevisiae stress adaptation and programmed cell death in response to acetic acid. Front Microbiol **4:**33.

13. **Mira NP, Teixeira MC, Sa-Correia I.** 2010. Adaptive response and tolerance to weak acids in Saccharomyces cerevisiae: a genome-wide view. OMICS **14:**525-540.

14. **Piper P, Calderon CO, Hatzixanthis K, Mollapour M.** 2001. Weak acid adaptation: the stress response that confers yeasts with resistance to organic acid food preservatives. Microbiology **147:**2635-2642.

15. **Berterame NM, Porro D, Ami D, Branduardi P.** 2016. Protein aggregation and membrane lipid modifications under lactic acid stress in wild type and OPI1 deleted Saccharomyces cerevisiae strains. Microb Cell Fact **15:**39.

16. **Martani F, Marano F, Bertacchi S, Porro D, Branduardi P.** 2015. The Saccharomyces cerevisiae poly(A) binding protein Pab1 as a target for eliciting stress tolerant phenotypes. Sci Rep **5:**18318.

17. **Rodrigues F, Sousa MJ, Ludovico P, Santos H, Corte-Real M, Leao C.** 2012. The fate of acetic acid during glucose co-metabolism by the spoilage yeast Zygosaccharomyces bailii. PLoS One **7:**e52402.

18. **Mollapour M, Piper PW.** 2001. The ZbYME2 gene from the food spoilage yeast Zygosaccharomyces bailii confers not only YME2 functions in Saccharomyces cerevisiae, but also the capacity for catabolism of sorbate and benzoate, two major weak organic acid preservatives. Mol Microbiol **42:**919-930.

19. **Lindberg L, Santos AX, Riezman H, Olsson L, Bettiga M.** 2013. Lipidomic profiling of Saccharomyces cerevisiae and Zygosaccharomyces bailii reveals critical changes in lipid composition in response to acetic acid stress. PLoS One **8:**e73936.

20. **Kuanyshev N, Ami D, Signori L, Porro D, Morrissey JP, Branduardi P.** 2016. Assessing physio-macromolecular effects of lactic acid on Zygosaccharomyces bailii cells during microaerobic fermentation. FEMS Yeast Res **16**.

21. **Suh SO, Gujjari P, Beres C, Beck B, Zhou J.** 2013. Proposal of Zygosaccharomyces parabailii sp. nov. and Zygosaccharomyces pseudobailii sp. nov., novel species closely related to Zygosaccharomyces bailii. Int J Syst Evol Microbiol **63:**1922-1929.

22. **Galeote V, Bigey F, Devillers H, Neuveglise C, Dequin S.** 2013. Genome Sequence of the Food Spoilage Yeast Zygosaccharomyces bailii CLIB 213T. Genome Announc **1**.

23. **Palma M, Munsterkotter M, Peca J, Guldener U, Sa-Correia I.** 2017. Genome sequence of the highly weak-acid-tolerant Zygosaccharomyces bailii IST302, amenable to genetic manipulations and physiological studies. FEMS Yeast Res **17**.

24. **Mira NP, Munsterkotter M, Dias-Valada F, Santos J, Palma M, Roque FC, Guerreiro JF, Rodrigues F, Sousa MJ, Leao C, Guldener U, Sa-Correia I.** 2014. The genome sequence of the highly acetic acid-tolerant Zygosaccharomyces bailii-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. DNA Res **21:**299-313.

25. **Ortiz-Merino RA, Kuanyshev N, Braun-Galleani S, Byrne KP, Porro D, Branduardi P, Wolfe KH.** 2017. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. PLoS Biol **15:**e2002128.

26. **Glover NM, Redestig H, Dessimoz C.** 2016. Homoeologs: What Are They and How Do We Infer Them? Trends Plant Sci **21:**609-621.

27. **Simpson CE, Ashe MP.** 2012. Adaptation to stress in yeast: to translate or not? Biochem Soc Trans **40:**794-799.

28. **Sauer M, Branduardi P, Gasser B, Valli M, Maurer M, Porro D, Mattanovich D.** 2004. Differential gene expression in recombinant Pichia pastoris analysed by heterologous DNA microarray hybridisation. Microb Cell Fact **3:**17.

29. **Decker CJ, Parker R.** 2012. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harb Perspect Biol **4:**a012286.

30. **Swisher KD, Parker R.** 2010. Localization to, and effects of Pbp1, Pbp4, Lsm12, Dhh1, and Pab1 on stress granules in Saccharomyces cerevisiae. PLoS One **5:**e10006.

31. **Iwaki A, Izawa S.** 2012. Acidic stress induces the formation of P-bodies, but not stress granules, with mild attenuation of bulk translation in Saccharomyces cerevisiae. Biochem J **446:**225-233.

32. **Abbott DA, Suir E, van Maris AJ, Pronk JT.** 2008. Physiological and transcriptional responses to high concentrations of lactic acid in anaerobic chemostat cultures of Saccharomyces cerevisiae. Appl Environ Microbiol **74:**5759-5768.

33. **Kawahata M, Masaki K, Fujii T, Iefuji H.** 2006. Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in Saccharomyces cerevisiae cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. FEMS Yeast Res **6:**924-936.

34. **Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sa-Correia I.** 2006. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. Nucleic Acids Res **34:**D446-451.

35. **Rutherford JC, Jaron S, Winge DR.** 2003. Aft1p and Aft2p mediate iron-responsive gene expression in yeast through related promoter elements. J Biol Chem **278:**27636-27643.

36. **Tishkov VI, Popov VO.** 2004. Catalytic mechanism and application of formate dehydrogenase. Biochemistry (Mosc) **69:**1252-1267.

37. **Overkamp KM, Kotter P, van der Hoek R, Schoondermark-Stolk S, Luttik MA, van Dijken JP, Pronk JT.** 2002. Functional analysis of structural genes for NAD(+)-dependent formate dehydrogenase in Saccharomyces cerevisiae. Yeast **19:**509-520.

38. **Iwadate Y, Funabasama N, Kato JI.** 2017. Involvement of formate dehydrogenases in stationary phase oxidative stress tolerance in Escherichia coli. FEMS Microbiol Lett **364**.

39. **Thomas SC, Alhasawi A, Auger C, Omri A, Appanna VD.** 2016. The role of formate in combatting oxidative stress. Antonie Van Leeuwenhoek **109:**263-271.

40. **Leandro MJ, Sychrova H, Prista C, Loureiro-Dias MC.** 2011. The osmotolerant fructophilic yeast Zygosaccharomyces rouxii employs two plasma-membrane fructose uptake systems belonging to a new family of yeast sugar transporters. Microbiology **157:**601-608.

41. **Cabral S, Prista C, Loureiro-Dias MC, Leandro MJ.** 2015. Occurrence of FFZ genes in yeasts and correlation with fructophilic behaviour. Microbiology **161:**2008-2018.

42. **Bracey D, Holyoak CD, Coote PJ.** 1998. Comparison of the inhibitory effect of sorbic acid and amphotericin B on Saccharomyces cerevisiae: is growth inhibition dependent on reduced intracellular pH? J Appl Microbiol **85:**1056-1066.

43. **Ullah A, Chandrasekaran G, Brul S, Smits GJ.** 2013. Yeast adaptation to weak acids prevents futile energy expenditure. Front Microbiol **4:**142.

44. **de Kok S, Yilmaz D, Daran JM, Pronk JT, van Maris AJ.** 2012. In vivo analysis of Saccharomyces cerevisiae plasma membrane ATPase Pma1p isoforms with increased in vitro H+/ATP stoichiometry. Antonie Van Leeuwenhoek **102:**401-406.

45. **Fletcher E, Feizi A, Kim S, Siewers V, Nielsen J.** 2015. RNA-seq analysis of Pichia anomala reveals important mechanisms required for survival at low pH. Microb Cell Fact **14:**143.

46. **Halm M, Hornbaek T, Arneborg N, Sefa-Dedeh S, Jespersen L.** 2004. Lactic acid tolerance determined by measurement of intracellular pH of single cells of Candida krusei and Saccharomyces cerevisiae isolated from fermented maize dough. Int J Food Microbiol **94:**97-103.

47. **Piper PW.** 1999. Yeast superoxide dismutase mutants reveal a pro-oxidant action of weak organic acid food preservatives. Free Radic Biol Med **27:**1219-1227.

679 48. **Ali MA, Yasui F, Matsugo S, Konishi T.** 2000. The lactate-dependent enhancement of
680 hydroxyl radical generation by the Fenton reaction. Free Radic Res **32:**429-438.
681 49. **Abbott DA, Suir E, Duong GH, de Hulster E, Pronk JT, van Maris AJ.** 2009. Catalase
682 overexpression reduces lactic acid-induced oxidative stress in Saccharomyces cerevisiae.
683 Appl Environ Microbiol **75:**2320-2325.
684 50. **Branduardi P, Fossati T, Sauer M, Pagani R, Mattanovich D, Porro D.** 2007.
685 Biosynthesis of vitamin C by yeast leads to increased stress resistance. PLoS One **2:**e1092.
686 51. **Yokota A, Kawabata A, Kitaoka S.** 1983. Mechanism of Glyoxylate Decarboxylation in
687 the Glycolate Pathway in Euglena gracilis Z : Participation of Mn-Dependent NADPH
688 Oxidase in Chloroplasts. Plant Physiol **71:**772-776.
689 52. **Alhasawi A, Castonguay Z, Appanna ND, Auger C, Appanna VD.** 2015. Glycine
690 metabolism and anti-oxidative defence mechanisms in Pseudomonas fluorescens. Microbiol
691 Res **171:**26-31.
692 53. **Aguilar-Uscanga B, Francois JM.** 2003. A study of the yeast cell wall composition and
693 structure in response to growth conditions and mode of cultivation. Lett Appl Microbiol
694 **37:**268-274.
695 54. **Lesage G, Bussey H.** 2006. Cell wall assembly in Saccharomyces cerevisiae. Microbiol Mol
696 Biol Rev **70:**317-343.
697 55. **Simoes T, Mira NP, Fernandes AR, Sa-Correia I.** 2006. The SPI1 gene, encoding a
698 glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the
699 development of yeast resistance to lipophilic weak-acid food preservatives. Appl Environ
700 Microbiol **72:**7168-7175.
701 56. **Rego A, Duarte AM, Azevedo F, Sousa MJ, Corte-Real M, Chaves SR.** 2014. Cell wall
702 dynamics modulate acetic acid-induced apoptotic cell death of Saccharomyces cerevisiae.
703 Microb Cell **1:**303-314.

57. **Kapteyn JC, ter Riet B, Vink E, Blad S, De Nobel H, Van Den Ende H, Klis FM.** 2001. Low external pH induces HOG1-dependent changes in the organization of the Saccharomyces cerevisiae cell wall. Mol Microbiol **39:**469-479.

58. **Mira NP, Becker JD, Sa-Correia I.** 2010. Genomic expression program involving the Haa1p-regulon in Saccharomyces cerevisiae response to acetic acid. OMICS **14:**587-601.

59. **Keller G, Ray E, Brown PO, Winge DR.** 2001. Haa1, a protein homologous to the copper-regulated transcription factor Ace1, is a novel transcriptional activator. J Biol Chem **276:**38697-38702.

60. **Palma M, Dias PJ, Roque FC, Luzia L, Guerreiro JF, Sa-Correia I.** 2017. The Zygosaccharomyces bailii transcription factor Haa1 is required for acetic acid and copper stress responses suggesting subfunctionalization of the ancestral bifunctional protein Haa1/Cup2. BMC Genomics **18:**75.

61. **Rapp RA, Udall JA, Wendel JF.** 2009. Genomic expression dominance in allopolyploids. BMC Biol **7:**18.

62. **Yoo MJ, Szadkowski E, Wendel JF.** 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity (Edinb) **110:**171-180.

63. **Combes MC, Hueber Y, Dereeper A, Rialle S, Herrera JC, Lashermes P.** 2015. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. Genome Biol Evol **7:**1110-1121.

64. **Cox MP, Dong T, Shen G, Dalvi Y, Scott DB, Ganley AR.** 2014. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. PLoS Genet **10:**e1004180.

65. **Tirosh I, Reikhav S, Levy AA, Barkai N.** 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. Science **324:**659-662.

66. **Wang Z, Sun X, Zhao Y, Guo X, Jiang H, Li H, Gu Z.** 2015. Evolution of gene regulation during transcription and translation. Genome Biol Evol **7:**1155-1167.

67. **Verduyn C, Postma E, Scheffers WA, Van Dijken JP.** 1992. 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.

68. **Proux-Wera E, Armisen D, Byrne KP, Wolfe KH.** 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. BMC Bioinformatics **13:**237.

69. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res **42:**D222-230.

70. **Finn RD, Clements J, Eddy SR.** 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res **39:**W29-37.

71. **Elbourne LD, Tetu SG, Hassan KA, Paulsen IT.** 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res **45:**D320-D324.

72. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

73. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-580.

74. **Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A.** 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res **36:**3420-3435.

75. **Langmead B.** 2010. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics **Chapter 11:**Unit 11 17.

76. **Anders S, Pyl PT, Huber W.** 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31:**166-169.

77. **Robinson MD, Oshlack A.** 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol **11:**R25.

78. **McCarthy DJ, Chen Y, Smyth GK.** 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res **40:**4288-4297.

79. **Revelle W.** 2017. psych: Procedures for Psychological, Psychometric, and Personality Research.

80. **Law CW, Chen Y, Shi W, Smyth GK.** 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol **15:**R29.

81. **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK.** 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res **43:**e47.

82. **Young MD, Wakefield MJ, Smyth GK, Oshlack A.** 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol **11:**R14.

83. **Conway JR, Lex A, Gehlenborg N.** 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics **33:**2938-2940.

84. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221-224.

85. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of

776 high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol
777 **7:**539.
778 86. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New
779 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
780 performance of PhyML 3.0. Syst Biol **59:**307-321.
781 87. **Edgar R, Domrachev M, Lash AE.** 2002. Gene Expression Omnibus: NCBI gene
782 expression and hybridization array data repository. Nucleic Acids Res **30:**207-210.

783

784

785

Category	Control	Lactic Acid	Intersect	Union
Expressed	9647	9652	9616	9683
Condition-specific	31	36	0	67
No evidence of expression	438	433	402	469

786

787 **Table 1. General overview of the *Z. parabailii* transcriptional profile.** The numbers of genes in
788 each category of expression are shown for each condition. After filtering and normalizing the
789 RNAseq counts, genes were categorized based on their expression profiles in both conditions,
790 pooling data from the two time points. Genes showing condition-specific expression only in control
791 conditions, or only in lactic acid, are listed in Tables S1 and S2. "No evidence of expression"
792 includes genes that were discarded by the filtering procedure and genes with no read counts.

Group	Mean	Median	Standard deviation	Skew	Ab	aB	P value	Unbalanced (%)
C18	0.500	0.500	0.079	-0.045	2077	2059	1	13.8
LA18	0.503	0.505	0.113	-0.009	2151	1985	0.041	31.0
C42	0.501	0.501	0.089	-0.146	2088	2048	1	18.7
LA42	0.500	0.503	0.115	-0.085	2119	2017	0.465	33.4

Table 2. Expression ratio between homeologous gene pairs. Mean, median, standard deviation and skew refer to the A/(A+B) expression ratios for 4136 homeologous gene pairs, as described in Methods. P values refer to the comparison between the number of homeolog pairs where the A-gene shows higher expression ("Ab"), and the opposite case ("aB"), in each of the four conditions. P values were obtained from two-sided exact binomial tests of the null hypothesis that the numbers of Ab and aB loci are equal, and were corrected for multiple testing using the Bonferroni method. C18: control at 18 h; LA18: lactic acid at 18 h; C42: control at 42 h; LA42: lactic acid at 42 h.

803

Category	18 h specific	42 h specific	At both time points	At either time point
Upregulated	117	83	27	227
Downregulated	430	431	158	1019

804

805 **Table 3. *Z. parabailii* differential expression analysis.** The upregulated and downregulated rows
806 show the numbers of genes with an adjusted P value < 0.05 and a $\log_2$-fold change ≥ 1, or $\log_2$-fold
807 change ≤ 1 respectively, between lactic acid and control conditions. The sets of genes were further
808 classified into those with altered expression only at 18 h, only at 42 h, at both time points, or at
809 either time point.

Figure 1. *Z. parabailii* fermentation profile. Batch bioreactor fermentation was performed in Verduyn medium at pH 3 with addition of 40 g L^{-1} lactic acid (red lines) or without lactic acid (black lines). The samples for RNA sequencing were taken at 18 h and 42 h (indicated by arrows), corresponding to exponential phase and post diauxic shift. Solid lines represent glucose consumption rate while dash lines corresponding optical density values at 660 nm.

Figure 2. Expression ratios in 4136 homeologous gene pairs. Expression ratio is defined as A/(A+B) where A and B are the RPKM values (reads per kilobase of mRNA per million transcripts) of the A- and B- homeologous genes, respectively, averaged among replicates. Histograms show the distribution of expression ratio values in (A) control conditions at 18 h; (B) lactic acid at 18 h; (C) control conditions at 42 h; (D) lactic acid at 42 h.

Figure 3 Enriched GO terms among differentially expressed genes. Bar plots show the numbers of differentially expressed genes associated with a GO term (dots) or with a group of GO terms (dots connected by vertical lines). Upregulated genes are shown in panel **A** and downregulated genes in panel **B**. For example, among the 33 upregulated genes with the term GO:0055114 for oxidation reduction process in panel A, 19 show only this term, 2 also show the term GO:0001320 for age−dependent response to reactive oxygen species, and so forth. The GO terms are ordered by ontology type (BP biological process, CC cellular component and MF molecular function) and by decreasing adjusted P value, always < 0.05 (values are in **Dataset S6**).

Figure 4 Log_2-fold changes for *Z. parabailii* genes putatively controlled by the Haa1 or the Aft1/2 transcription factors. Genes under Haa1 control are shown in panel A and genes controlled by Aft1/2 are shown in panel B. Asterisks (*) are used to mark *S. cerevisiae* genes reported as lactic

835 acid-responsive by Abbot *et al*., whose *Z. parabailii* homologs display an opposite response profile
836 (*i.e.* upregulated in *S. cerevisiae* and downregulated in *Z. parabailii*). Positive $\log_2$-fold change
837 values in lactic acid vs. control are coloured in red as a sign of upregulation whereas negative values
838 are blue. All the changes shown have an adjusted P value < 0.05 (values are in **Dataset S7**).
839
840 **Figure 5. Phylogenetic tree of formate dehydrogenase amino acid sequences in yeast species.**
841 The six *Z. parabailii* Fdh-like genes are shown (names ending in _A, _B, and _N). Prefixes ZYRO
842 and ZYBA indicate genes from *Z. rouxii* and *Z. bailii*, respectively. The tree was rooted using the
843 paralogous yeast gene Gor1 (glyoxylate reductase), and *Escherichia coli* Fdh is included for
844 reference. The tree was constructed using PhyML. Bootstrap values from 100 replicates are shown.
845