Type

Research Letter

Title

Design of Experiments (DOE) methodology to build a multifactorial statistical model describing the metabolic interactions of alcohol dehydrogenase isozymes in the ethanol biosynthetic pathway of the yeast *Saccharomyces cerevisiae*

Authors

Steven R. Brown[1]; Marta Staff[1]; Rob Lee[2]; John Love[1]; David A. Parker[2]; Stephen J. Aves[1]; Thomas P. Howard[3,*]

Author Addresses

[1]Biosciences, Geoffrey Pope Building, College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4QD, UK

[2]Biodomain, Shell Technology Center Houston, 3333 Highway 6 South, Houston, Texas, 77082-3101, USA

[3]School of Natural and Environmental Sciences, Devonshire Building, Faculty of Science, Agriculture and Engineering, Newcastle University, Newcastle-upon-Tyne, NE1 7RU, UK

***Corresponding Author**

Thomas P. Howard

School of Natural and Environmental Sciences, Devonshire Building, Faculty of Science, Agriculture and Engineering, Newcastle University, Newcastle-upon-Tyne, NE1 7RU, UK

Email: thomas.howard@ncl.ac.uk

Telephone: +44 (0)191 208 48 54

Author Contact Details

S.R.B. s.brown@synthace.com

Synthace Ltd., The London Bioscience Innovation Centre, 2 Royal College Street, London, NW1 0NH, United Kingdom

M.S. Marta.e.staff@gmail.com

Biosciences, Geoffrey Pope Building, College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4QD, UK

D.A.P. David.A.Parker@shell.com

Biodomain, Shell Technology Center Houston, 3333 Highway 6 South, Houston, Texas, 77082-3101, USA

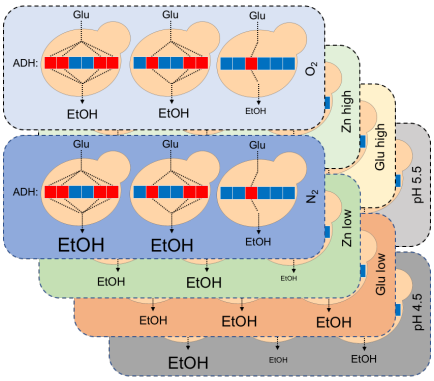
R.L. R.Lee@exeter.ac.uk
Biodomain, Shell Technology Center Houston, 3333 Highway 6 South, Houston, Texas, 77082-3101, USA

J.L. J.Love@exeter.ac.uk
Biosciences, The BioEconomy Centre, College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4QD, UK

S.J.A. S.J.Aves@exeter.ac.uk
Biosciences, Geoffrey Pope Building, College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4QD, UK

ABSTRACT

Multifactorial approaches can quickly and efficiently model complex, interacting natural or engineered biological systems in a way that traditional one-factor-at-a-time experimentation can fail to do. We applied a Design of Experiments (DOE) approach to model ethanol biosynthesis in yeast, which is well-understood and genetically tractable, yet complex. Six alcohol dehydrogenase (ADH) isozymes catalyse ethanol synthesis, differing in their transcriptional and post-translational regulation, subcellular localisation, and enzyme kinetics. We generated a combinatorial library of all ADH gene deletions, and measured the impact of gene deletion(s) and environmental context on ethanol production of a subset of this library. The data were used to build a statistical model that described known behaviours of ADH isozymes and identified novel interactions. Importantly, the model described features of ADH metabolic behaviour without explicit *a priori* knowledge. The method is therefore highly suited to understanding and optimising metabolic pathways in less well understood systems.



Keywords
Saccharomyces cerevisiae; ethanol biosynthesis; Design of Experiments (DOE); metabolic engineering; alcohol dehydrogenase

Wide accessibility to DNA synthesis, combined with increasing affordability of laboratory automation, are powerful drivers of biological research and engineering. The ability to design gene constructs *in silico*, to the specifications required by the researcher rather than evolution, and to have these assembled swiftly, enables scientists and engineers to test design concepts largely inaccessible through traditional cloning methods. There is, therefore, an unprecedented opportunity to examine and engineer biological function. Despite these advances, much research effort rests on a traditional approach to performing experimentation, i.e. the sequential alteration of a single factor whilst all other factors remain constant. Such an approach, frequently termed one-factor-at-a-time (OFAT), is not the most efficient or effective means to understand or optimise multifactorial systems. Biological systems are inherently multifactorial and demonstrate strong, non-linear interactions between both genetic and environmental factors. If OFAT is the only method employed, these properties confound understanding and make optimisation difficult[1]. The requirement to use multivariate statistical methodologies to evaluate complex and often noisy biological processes is therefore crucial.

Design of Experiments (DOE) is an empirical statistical methodology applicable to both the design and analysis of experiments. This multifactorial method explores the design space with less bias, and with fewer experimental runs and resource requirements than traditional OFAT. The DOE approach can identify strong, non-linear interactions and provide well-structured datasets against which statistical and mechanistic models can be validated. Where it has been applied, DOE has proven extremely powerful for exploring complex, multifactor situations. Examples include optimising cell-free protein expression[2], engineering codon-bias[3-4], enzyme improvements and repurposing[5-6] and optimisation of biochemical systems within an artificial cell[7]. Recently, DOE has been applied to metabolic engineering of violacein pathway in *Escherichia coli*[8]. The approach therefore has clear advantages to metabolic optimisation efforts.

We wished to assess the DOE approach in investigating a well-studied pathway in which genetic and environmental factors are known to interact, and to determine if this method is capable of capturing known behaviour and elucidating new information. We chose to investigate ethanol production in *S. cerevisiae*. This process is well-defined, with a significant body of both biochemical and genetic research. Furthermore, it is industrially important not just for the production of ethanol, but also because ethanol represents a major carbon sink that affects the synthesis of alternative products in yeast[9-11]. Ethanol production in *S. cerevisiae* is catalysed by multiple alcohol dehydrogenase (ADH) isozymes[12]. Adh1 is the dominant isoform and strains with a *Δadh1* genotype have a significantly reduced rate of ethanol production. Complete removal of ethanol production in yeast, permitting redirection of carbon flux to other products, is however not trivial. A quadruple

ADH deletion strain (*Δadh1 Δadh2 Δadh3 Δadh4*), for example, still produces significant quantities of ethanol. To generate a stable disruption of ethanol biosynthesis, removal of six ADH isozymes (*Δadh1 Δadh2 Δadh3 Δadh4 Δadh5 Δsfa1*) is required under standard growth conditions[13]. The principal functions of these six ADH isozymes are summarised in **Figure S1** and their phylogenetic relationships are shown in **Figure S2**. Removal of ethanol production is accompanied by a reduction in growth rate and a concomitant accumulation of glycerol[13]. This latter phenomenon is attributed to fermentation in which reducing equivalents, normally recycled during ethanol production, are instead transferred to dihydroxyacetone phosphate to form glycerol-3-phosphate, which is converted to glycerol. Other complications in understanding and redirecting carbon flux from ethanol biosynthesis arise from the fact that ADH isozymes functionally overlap making simple, single-gene-to-function relationships difficult to capture. For example, experiments using ethanol as a carbon source for growth have demonstrated that Adh1 is capable of assimilating ethanol, a function attributed to Adh2[12], whilst gene function may also be strain dependent. mRNA levels of Adh4 in the brewing yeast strain NCYC1245 indicate it to be the main cytosolic ADH, whereas in other laboratory strains there is a complete lack of Adh4 gene expression[12, 14-15].

Generating a library of *S. cerevisiae* strains that includes all permutations of the ADH genotypes would provide a useful tool for understanding how these enzymes functionally substitute for each other *in vivo*. Here we describe the generation of a 64-strain combinatorial ADH gene knockout library in an industrially-relevant prototrophic strain of *S. cerevisiae*. We use this library to appraise the DOE methodology for evaluating genotype-by-genotype and genotype-by-environment interactions and we use the data generated from these evaluations to build a statistical model of ethanol biosynthesis. This model is congruent with our understanding of ethanol biosynthesis in yeast, confirms hypothesised functions of ADH isozymes and indicates new roles in fermentation. We propose that such a methodology is therefore a viable method for optimising novel metabolic pathways and is advantageous for the exploitation of non-model organism metabolism for which there is limited prior knowledge.

We used the prototrophic, haploid *S. cerevisiae* strain CEN.PK113-7D for construction of a full combinatorial knockout library of ADH genes *ADH1*, *ADH2*, *ADH3*, *ADH4*, *ADH5* and *SFA1* (**Figure S3**). This yeast strain offers a good compromise between genetic accessibility and physiological properties based on its growth characteristics, transformation efficiency and industrial applicability[16]. To rapidly generate multiple ADH deletion strains, we adapted the *amdSYM* method for sequential gene deletion[17]. The acetamidase *amdSYM* marker gene from *Aspergillus nidulans* allows selection of transformants on nitrogen-deficient media, through the conversion of acetamide into acetate and ammonia[17]. Counter-selection for marker loss is achieved by culturing cells in the presence of fluoroacetamide, which is converted by acetamidase to the toxic product fluoroacetate. The *amdSYM* method allows sequential rounds of open reading frame (ORF) deletion without retaining short sequence "scars" in the genome, such as the *loxP* sites of the Cre-loxP system which

can lead to genome instability and altered physiological response in *S. cerevisiae*[18]. We adapted the *amdSYM* method in order to precisely delete ORFs without any loss of adjacent genomic sequence (see **Figure S4** for details). For rapid rounds of ADH gene deletion and marker recycling we established a cyclic workflow incorporating colony PCR-based checkpoints for the integration and excision steps (**Figure 1**). Using this workflow, we successfully generated a combinatorial ADH gene knockout library comprising all 64 ADH gene deletion permutations in *S. cerevisiae* strain CEN.PK113-7D. All genotypes were verified by diagnostic PCR (**Figure S5**).

In addition to the six genetic factors (*ADH1-5* and *SFA1*) we identified six environmental factors as having the potential to influence the production of ethanol in *S. cerevisiae*, and used a combination of experimental scoping exercises and literature to assign upper and lower boundaries for these (**Table 1**). The upper level for the dilution rate was set at the critical dilution rate threshold required to elicit the Crabtree effect in aerobic chemostat culture, with a D_{crit} of 0.29 (refs [19-20]). The carbon substrate levels were chosen to provide glucose concentrations shown to be either limiting or previously used to evaluate the Crabtree effect in *S. cerevisiae* in chemostat culture[19]. The temperatures represented the laboratory standard growth temperature for yeast culture and a higher value balancing ethanol yield and culture viability[21]. Zinc concentration has been shown to regulate the expression of the *ADH4* gene, therefore the effects of ADH isozymes on ethanol production were assessed in zinc-limiting and zinc-replete media[22]. The initial pH of the media included the level used for standard growth under laboratory conditions as well as a more acidic condition, corresponding to the pH of industrial consolidated bioprocessing of lignocellulose to bioethanol[23]. Evaluating all permutations of the levels of factors in **Table 1** would require 4096 experiments. By applying DOE, we were able to reduce this to a set of 88 experiments that statistically delineate the main effects and the effect of all two-way-interactions on the measured response, ethanol production (**Table S1**).

Shake flask experiments do not provide the level of control required for experimentation of this nature, which ideally should be conducted under controlled chemostat steady-state fermentation conditions. We established a parallel experimentation system – the ministat[24] – that could provide sufficient throughput for our experimental designs (**Figure S6**). In total, 32 parallel ministat bioreactors were constructed with adaptations including individual media reservoirs and effluent burettes to measure the media flow rate for each. Dilution rate was controlled using a single peristaltic pump with 32 channels, which could be individually fine-tuned. Two heat blocks controlled temperature. Experimentation was performed at an individual bioreactor gas flow rate of 200 ml min^{-1} with either air or nitrogen. Inoculation of the ministat bioreactors used 3.0 ml of overnight culture added to 17 ml of culture medium, grown in batch culture for 24 h prior to operation as a ministat. Experiments were performed in a medium-throughput manner and permitted steady state measurements to be made providing a direct comparative evaluation of the ADH gene deletion library. Data from the 88 experiment runs are shown in **Table S1**.

Statistical modelling of the fermentation data was performed. A Partial Least Squares (or Projection to Latent Structures, PLS) analysis including KFold cross validation and the SIMPLS algorithm were applied to a subset of the data shown in **Table S1**. The PLS model highlighted 60 genotype and environmental factors, including their interactions, as being important predictors (VIP > 1.0; ref [25]) for modelling ethanol production. A PLS model fitted using three latent variables (also known as factors) to describe the relationship between X and Y matrices had a root mean PRESS of 0.6772 and explained 13.9 % of the cumulative variation in the X score and 95.0 % of the cumulative variation in the Y score. The PLS modelling platform is a flexible, effective method for modelling experimentation where the variables are expected to be correlated and noisy. The run order of the designed experiment was constrained to four whole plots (experiment batches) and eight subplots due to experimental limitations, including a single pump to control dilution rate, and the presence of only two heat blocks for temperature control of the ministats. Potential effects of the reduction in run randomisation were assessed by modelling the effect of the whole and subplots factors. The run order had no significant impact on the model projection of ethanol production. Calculated values for ethanol yield per gram of glucose were in broad agreement with values previously reported for ADH deletion strains cultured under similar conditions[13] (**Table S1**).

In total, 60 different factor and factor interactions (VIP score > 1.0) were determined to be important for the model projection of ethanol production in the *S. cerevisiae* ADH gene deletion library under different environmental conditions (**Figure 2)**. This result highlights the complexity of regulation and the adaptability of the natural system and shows the necessity to use structured experimental design. Model diagnostics demonstrated that the model was internally consistent (**Figure 3**). The actual-by-predicted diagnostic plot confirms that the data are predicted well by the model, and the residual by predicted, residual by row, and residual normal quantile plots demonstrate a balanced standard distribution with no distinct patterns, with the following caveat: aerobic cultures, operated at a dilution rate of 0.06 (which corresponds to a culture growth rate below the D_{crit} threshold value for mixed respiro-fermentative growth) perform cellular respiration without exhibiting the Crabtree effect. This subset of experiments results in no ethanol production and the abrupt edge effect seen in **Figure 3b**.

We next evaluated whether the model reflects known responses of ethanol biosynthesis in yeast to environmental cues, and whether it describes the relative importance of the various ADH isoforms. Consistent with prior findings, the model describes key factors influencing ethanol production that include the presence of *ADH1*, aeration and the interaction between many ADH isoforms and aeration (**Figure 2, Table S2**). Given that Adh1 has been shown to be primarily responsible for catalysing the reduction of acetaldehyde to ethanol[26-27] it may be surprising that the interaction of *ADH3* and nitrogen was identified as a similarly important predictor for the model (**Figure 2** (1); VIP = 2.99), in comparison with the interaction between *ADH1* and nitrogen **(Figure 2** (2); VIP = 2.71). It is known, however, that Adh3 functions as part of the ethanol acetaldehyde shuttle

in the mitochondria, and under anaerobic conditions it is required to re-oxidise mitochondrial NADH[28]. This is because pyridine nucleotides NAD^+ and NADH cannot cross the mitochondrial inner membrane[29] but ethanol and acetaldehyde can. Providing there is a cytosolic alcohol dehydrogenase[30], re-oxidation of mitochondrial NADH during anaerobic growth can occur due to the action of Adh3, which addresses the redox imbalance within the system, increasing metabolic flux toward ethanol biosynthesis. This phenomenon can be visualised using an associated prediction profiler of ethanol production (**Figure 4a**). This shows that the effect of Adh3 on ethanol production decreases when aeration is applied to the system.

The model of ethanol production generated is therefore congruent with well-documented dynamics of the ADH isozymes with regard to metabolism in *S. cerevisiae*. This gives confidence in the ability of the PLS model to describe the role genetic and environmental factors play in determining ethanol production. We can use the model to probe other scenarios. Adh2 for example, has been shown to catalyse the reaction producing ethanol from acetaldehyde[12]. Its primary role has therefore been identified in ethanol re-assimilation back into primary metabolism. It has been proposed that Adh2 can also operate in the direction of ethanol biosynthesis in the absence of a functioning Adh1. Our model agrees, but provides an additional caveat: biosynthesis of ethanol via Adh2 does not occur under anaerobic conditions when the glucose concentration is high (**Figure 4b**). The presence of the *ADH2* gene only increases ethanol production when either oxygen is present or when glucose concentration is limiting. In identifying these details our statistical model supports hypotheses generated by gene expression studies of *ADH2*. *ADH2* gene expression has been shown to be oxygen inducible and catabolite repressed (transcription is markedly but not completely repressed by growth on glucose[15]), an effect further augmented because the Adr1 transcription factor, which positively regulates *ADH2* expression, is inactivated during growth on both glucose and ethanol[15]. In identifying these factors our model serves two purposes. Firstly, it provides a further line of evidence in support for the complex response of *ADH2* identified through gene expression studies. Secondly, it provides a clear indication that this data-driven, statistical approach is capable of capturing subtleties in metabolic regulation that were not explicitly programmed *a priori*.

Adh5 was discovered from sequence homology to Adh1 (77 % amino acid identity) during sequencing of the *S. cerevisiae* genome[31]. The distinct function and transcriptional regulation of *ADH5* remain unclear. Chemostat glucose-pulse experiments concluded that *ADH5* was expressed at a constant level during experimentation, and in a study deleting *ADH1/2/3/4*, the strain still produced ethanol and this was proposed to be due to the function of Adh5[15, 26]. The prediction profiler (**Figure 4c**) for the modelled response of ethanol production indicates that the *ADH5* gene does indeed function in the production of ethanol in *S. cerevisiae*. Importantly however, our model indicates that this function occurs predominantly at pH 4.5 with the activity reduced by 76 % at pH 5.5. This may indicate why its function has yet to be fully elucidated; laboratory evaluation of yeast strains often occurs at the standardised pH of 5.5.

The expression and importance of *ADH4* in *S. cerevisiae* also remain to be fully elucidated, with reports of a lack of expression in some laboratory strains of *S. cerevisiae*[12, 15] and expression in others[32]. Adh4 shares very little sequence similarity with the other ADH isozymes; it is more closely related to the iron-activated ADH from *Zymomonas mobilis* and other bacteria but requires zinc for activation[32-33]. The expression of *ADH4* is upregulated under low zinc conditions[12] and it has been postulated that its expression is advantageous during zinc starvation either due to the protein binding only one atom of zinc per subunit (it is a dimer as opposed to the other tetrameric ADH isozymes), or that it has a more efficient catalytic activity than Adh1 under conditions of zinc limitation[34]. The interaction of the ADH isozymes with zinc in the media and the corresponding impact on ethanol production can be seen in **Figure 5**. Here it is shown that the contribution of all isozymes to ethanol biosynthesis, bar Adh4, is lower under zinc limiting conditions. By contrast, the contribution of Adh4 to ethanol biosynthesis is greater under zinc limiting conditions and drops as zinc concentration rises. This provides the first empirical evidence that the Adh4 isozyme does indeed provide ADH activities for the cell under zinc limiting conditions.

The presence or absence of *SFA1* alone is not identified as a significant predictor of ethanol concentration (**Figure 2**), though its importance to ethanol production was identified in a full ADH knockout strain[13]. Here, *SFA1* is only seen as an important factor in ethanol production in combination with other genetic or environmental factors. The greatest effects on the model projection are seen in the absence of *SFA1*. The most significant of these is the interaction between *SFA1* and the aeration conditions. Specifically, in aerobic conditions the loss of *SFA1* has a strong negative effect within the model, whilst in anaerobic conditions the loss of *SFA1* has a strong positive effect on ethanol production in the model (**Figure 2**). The model however does not clarify the dynamics of its function and it is difficult to assign a model coefficient to this factor. For example, under anaerobic conditions both the presence and absence of *SFA1* are deemed to have positive impacts on ethanol production (**Figure 2** (13, 32); VIP = 1.78 and VIP = 1.34 respectively). This may be due to further regulation by additional factors (genotypic or environmental) that were not included in the analysis, or by higher order interactions.

In summary, a multivariate approach using a structured experimental design, including both genetic and environmental factors, requires an initial investment in planning but has a high return on understanding. Using this approach to examine the impact of ADH genes on ethanol production in yeast, sensitivities to cellular or environmental context are taken into account by testing each factor in different contexts. This is valuable for understanding and engineering metabolic performance. Here, 88 experiments (compared to a full factorial 4096 experimental combinations) provided data for a PLS model of carbon flux to ethanol that captures current understanding and behaviour of all of the ADH isozymes within different genetic and environmental contexts. The resulting model detailed 60 different genotype-by-genotype and genotype-by-environment interactions important to model projection. Additionally, the model has furthered our understanding of ethanol production under

different environmental conditions in *S. cerevisiae*. It supports the hypothesised role for Adh4 in zinc-limited environments and for the regulation of *ADH2* gene expression under different environmental conditions, and identifies physiologically-relevant pH sensitivity of Adh5. The results highlight both the sophistication contained within a single reaction catalysed by multiple ADH isozymes and the power of the multivariate methodology for exploring complex systems. Given the recent interest in this area for metabolic engineers, we expect these approaches to become more commonplace in the coming years to understand and engineer metabolic networks. Finally, model-guided biological engineering, whether genome scale metabolic models or constraint-based techniques such as Flux Balance Analysis, are popular tools to computationally predict phenotypes under environmental and genetic perturbations in steady state[35]. Validation of these engineering strategies is likely to benefit from the growth of statistically structured experimental datasets as exemplified within this study.

MATERIALS AND METHODS

Strains and plasmids. *Saccharomyces cerevisiae* strain CEN.PK113-7D (MATa *MAL2-8^c SUC2*) was from EUROSCARF (Frankfurt, Germany). ADH gene deletion cassettes for the *amdSYM* gene deletion method were synthesised in *Escherichia coli* vector pJ154 series by DNA 2.0 Inc. (California, US) flanked by *Bmr*I sites such that plasmid linearization left no remnants of the restriction site or the expression vector backbone (**Figure S7; Table S3**). Plasmids were maintained in *E. coli* NEB 5-alpha (New England Biolabs, Ipswich, US). Details of media and growth conditions are given in the supporting information (**Methods S1**).

Generation of the combinatorial deletion library. ADH gene deletions followed the workflow in **Figure 1** and used the *amdSYM* gene deletion method (**Figure S4**). Transformations of linearised deletion cassettes into *S. cerevisiae* CEN.PK113-7D were performed with the yeast transformation kit (Sigma-Aldrich, Dorset, UK), adapted as described in the supporting information (**Methods S1**). The *ADH1* gene deletion cassette required an extended 100 bp 5′ homologous sequence (**Table S3**) to achieve a transformation efficiency >20 transformants μg^{-1} DNA for all deletion cassettes (**Table S4**). For yeast colony PCR screening, a single colony was picked using a sterile pipette tip and suspended in 50 µl Tris-EDTA buffer pH 8.0. The mixture was heated to 99.9 °C for 5.0 min, cooled to 4.0 °C and placed on ice for at least 5.0 min. The mixture was centrifuged at 13 000 $\times$g and 5 µl of the supernatant was used as template DNA for diagnostic PCR for gene deletions using primers detailed in **Table S5**. Primers flanked each of the ADH genes permitting verification of the intact wild type gene, the integrated *amdSYM* deletion cassette, or scarless deletion. Integration efficiencies were ≥83 % for each deletion cassette (**Table S4**). In the final library 21 strains contained a single small remnant of marker cassette due to a historic *loxP* site and multiple cloning site left in the original plasmid used for transformation (**Figure S8**).

Miniature chemostat assembly. A bioprocess system was built that included 32 parallel ministat bioreactors (**Figure S6**)[24]. The equipment was adapted to include multiple feed vessels: 32 Drechsel bottles contained 250 ml of sterile media and permitted media-associated factors to be changed for each ministat bioreactor using Synthetic Defined media as described in **Materials S1**. The dilution rate for each reactor was measured in-line using 32 x 50 ml burettes connected to the effluent line of the ministat bioreactors and was adjusted as required. Ministats were operated at the set dilution rate (D) of 1.2 ml h^{-1} = (μ) 0.06 h^{-1} to prevent wash out of cultures, which occurs when the dilution rate exceeds the maximum specific growth of a culture. When cultured anaerobically, *S. cerevisiae* is unable to synthesise sterols and unsaturated fatty acids, therefore media were supplemented with Tween [420 mg l^{-1}] and ergosterol [10 mg l^{-1}] for anaerobic cultures. Two heat blocks controlled temperature. The gas addition manifold was adapted enabling robust gas flow of either air or nitrogen to each ministat at a flow rate of 200 ml min^{-1} ensuring adequate mixing of the culture broth. Inoculation of the ministat bioreactors used 3.0 ml of overnight culture (10 ml SD broth cultures in 50 ml tube, 30 °C with shaking at 200 rpm) added to 17 ml of culture medium, grown in batch culture for 24 h prior to operation as a ministat. The ministat bioreactors were operated as chemostats for 83.3 hours, which permitted 5 reactor volume changes to occur prior to harvesting and subsequent analysis. Ethanol was determined by high performance liquid chromatography with a refractive index detector.

Design of Experiments and statistical analysis. Design of Experiments comprising optimal design of experimentation, data modelling and visualisation was performed using JMP Pro v.12 (SAS Institute Inc. USA). The number and choice of runs was determined using the Custom Design platform in JMP Pro v.12 to balance experimental resources with the collection of appropriate data to permit estimations of the main effects (single effects of all factors) and all power terms for two-factor interactions. A statistical interaction is indicated in the manuscript by * meaning that two factors combine to influence the impact on ethanol production. This may be a synergistic or antagonistic interaction. Model validation was performed using KFold (10) cross validation and the SIMPLS (Statistically Inspired Modification of the PLS method) algorithm applied to the dataset. The coefficient estimates for X and Y for different models were comparatively evaluated using the root mean PRESS (Predicted REsidual Sum of Squares) statistic. Root mean PRESS provides an estimate of the squared prediction error between an observed validation set value and the model predicted value; a lower root mean PRESS value indicates a lower variance of the model prediction for a given response compared to the validation dataset. The Prediction Profiler is an interactive representation of the underpinning PLS model. When a factor is altered the profiler recalculates the predicted response allowing exploration of the underlying model predictions.

ASSOCIATED CONTENT

Supporting Information

Supporting Information is available free of charge on the ACS Publications website at DOI:

Supporting Information: Figures S1-S8; Tables S1-S6; Supporting Material and Methods (.pdf)

AUTHOR INFORMATION

***Corresponding Author**

Tel.: +44 (0)191 208 4854. Email: thomas.howard@ncl.ac.uk

Present Address

S.R.B. Synthace Ltd., The London Bioscience Innovation Centre, 2 Royal College Street, London, NW1 0NH, United Kingdom

ORCID ID

Steven R. Brown: 0000-0001-9715-6092

Stephen J. Aves: 0000-000205600-900X

Thomas P. Howard: 0000-0002-5546-4043

Author Contributions

S.R.B., D.A.P. R.L., J.L. S.J.A. and T.P.H. designed the study; S.R.B. and M.S. performed the experiments; S.R.B., D.A.P., S.J.A. and T.P.H. analysed the data; S.R.B., S.J.A. and T.P.H. wrote the manuscript. All authors commented on and revised the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

We wish to thank Dr Alex Johns for helpful discussions. S.R.B. would also like to thank Shell Biodomain for funding for this PhD research project.

REFERENCES

1. Jansen, R. C., (2003) Studying complex biological systems using multifactorial perturbation. *Nature Reviews Genetics 4*, 145-151.

2. Caschera, F., Bedau, M. A., Buchanan, A., Cawse, J., de Lucrezia, D., Gazzola, G., Hanczyc, M. M., and Packard, N. H., (2011) Coping with complexity: machine learning optimization of cell-free protein synthesis. *Biotechnol. Bioeng. 108*, 2218-2228.

3. Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., and Welch, M., (2012) Engineering genes for predictable protein expression. *Protein Expression and Purification 83*, 37-46.

4. Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C., (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One 4*, e7002.

5. Govindarajan, S., Mannervik, B., Silverman, J. A., Wright, K., Regitsky, D., Hegazy, U., Purcell, T. J., Welch, M., Minshull, J., and Gustafsson, C., (2015) Mapping of amino acid substitutions conferring herbicide resistance in wheat glutathione transferase. *ACS Synth Biol 4*, 221-7.
6. Liao, J., Warmuth, M. K., Govindarajan, S., Ness, J. E., Wang, R. P., Gustafsson, C., and Minshull, J., (2007) Engineering proteinase K using machine learning and synthetic genes. *BMC biotechnology 7*, 16.
7. Caschera, F., Rasmussen, S., and Hanczyc, M., (2011) Machine learning optimization of evolvable artificial cells. *Procedia Computer Science 7*, 187-189.
8. Xu, P., Rizzoni, E. A., Sul, S. Y., and Stephanopoulos, G., (2017) Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synth Biol 6*, 148-158.
9. Abbott, D. A., Zelle, R. M., Pronk, J. T., and van Maris, A. J., (2009) Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. *FEMS Yeast Res 9*, 1123-36.
10. Runguphan, W., and Keasling, J. D., (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab Eng 21*, 103-13.
11. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., et al., (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature 496*, 1-9.
12. de Smidt, O., du Preez, J. C., and Albertyn, J., (2012) Molecular and physiological aspects of alcohol dehydrogenases in the ethanol metabolism of *Saccharomyces cerevisiae*. *FEMS Yeast Res 12*, 33-47.
13. Ida, Y., Furusawa, C., Hirasawa, T., and Shimizu, H., (2012) Stable disruption of ethanol production by deletion of the genes encoding alcohol dehydrogenase isozymes in *Saccharomyces cerevisiae*. *J Biosci Bioeng 113*, 192-5.
14. Mizuno, A., Tabei, H., and Iwahuti, M., (2006) Characterization of low-acetic-acid-producing yeast isolated from 2-deoxyglucose-resistant mutants and its application to high-gravity brewing. *J Biosci Bioeng 101*, 31-7.
15. van den Berg, M. A., De Jong-Gubbels, P., and Steensma, H. Y., (1998) Transient mRNA responses in chemostat cultures as a method of defining putative regulatory elements: Application to genes involved in *Saccharomyces cerevisiae* acetyl-coenzyme A metabolism. *Yeast 14*, 1089-1104.
16. van Dijken, J. P., Bauer, J., Brambilla, L., Duboc, P., Francois, J. M., Gancedo, C., Giuseppin, M. L., Heijnen, J. J., Hoare, M., Lange, H. C., et al., (2000) An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol 26*, 706-714.

17. Solis-Escalante, D., Kuijpers, N. G. A., Bongaerts, N., Bolat, I., Bosman, L., Pronk, J. T., Daran, J.-M., and Daran-Lapujade, P., (2013) *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Research 13*, 126-139.

18. Solis-Escalante, D., van den Broek, M., Kuijpers, N. G., Pronk, J. T., Boles, E., Daran, J. M., and Daran-Lapujade, P., (2015) The genome sequence of the popular hexose-transport-deficient *Saccharomyces cerevisiae* strain EBY.VW4000 reveals LoxP/Cre-induced translocations and gene loss. *FEMS Yeast Res 15*.

19. van Hoek, P., van Dijken, J. P., and Pronk, J. T., (2000) Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of *Saccharomyces cerevisiae*. *Enzyme Microb Technol 26*, 724-736.

20. Vemuri, G. N., Eiteman, M. A., McEwen, J. E., Olsson, L., and Nielsen, J., (2007) Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A 104*, 2402-7.

21. Torija, M. J., Rozes, N., Poblet, M., Guillamon, J. M., and Mas, A., (2003) Effects of fermentation temperature on the strain population of *Saccharomyces cerevisiae*. *Int J Food Microbiol 80*, 47-53.

22. North, M., Steffen, J., Loguinov, A. V., Zimmerman, G. R., Vulpe, C. D., and Eide, D. J., (2012) Genome-wide functional profiling identifies genes and processes important for zinc-limited growth of *Saccharomyces cerevisiae*. *PLoS Genet 8*, e1002699.

23. Hasunuma, T., and Kondo, A., (2012) Development of yeast cell factories for consolidated bioprocessing of lignocellulose to bioethanol through cell surface engineering. *Biotechnology Advances 30*, 1207-1218.

24. Miller, A. W., Befort, C., Kerr, E. O., and Dunham, M. J., (2013) Design and use of multiplexed chemostat arrays. *J Vis Exp*, e50262.

25. Wold, S., Sjöström, M., and Eriksson, L., (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems 58*, 109-130.

26. Drewke, C., Thielen, J., and Ciriacy, M., (1990) Ethanol formation in adh0 mutants reveals the existence of a novel acetaldehyde-reducing activity in *Saccharomyces cerevisiae*. *J Bacteriol 172*, 3909-17.

27. Thomson, J. M., Gaucher, E. A., Burgan, M. F., De Kee, D. W., Li, T., Aris, J. P., and Benner, S. A., (2005) Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet 37*, 630-5.

28. Bakker, B. M., Bro, C., Kotter, P., Luttik, M. A. H., van Dijken, J. P., and Pronk, J. T., (2000) The mitochondrial alcohol dehydrogenase adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae*. *J. Bacteriol. 182*, 4730-4737.

29. Heux, S., Cachon, R., and Dequin, S., (2006) Cofactor engineering in *Saccharomyces cerevisiae*: Expression of a H_2O-forming NADH oxidase and impact on redox metabolism. *Metab Eng 8*, 303-14.

30. de Smidt, O., du Preez, J. C., and Albertyn, J., (2008) The alcohol dehydrogenases of *Saccharomyces cerevisiae*: a comprehensive review. *FEMS Yeast Res 8*, 967-78.

31. Feldmann, H., Aigle, M., Aljinovic, G., Andre, B., Baclet, M. C., Barthe, C., Baur, A., Becam, A. M., Biteau, N., Boles, E., et al., (1994) Complete DNA sequence of yeast chromosome II. *EMBO J 13*, 5795-809.

32. Drewke, C., and Ciriacy, M., (1988) Overexpression, purification and properties of alcohol dehydrogenase IV from *Saccharomyces cerevisiae*. *Biochim Biophys Acta 950*, 54-60.

33. Yuan, D. S., (2000) Zinc-regulated genes in *Saccharomyces cerevisiae* revealed by transposon tagging. *Genetics 156*, 45-58.

34. Bird, A. J., Gordon, M., Eide, D. J., and Winge, D. R., (2006) Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *EMBO J 25*, 5726-34.

35. VanderSluis, B., Hess, D. C., Pesyna, C., Krumholz, E. W., Syed, T., Szappanos, B., Nislow, C., Papp, B., Troyanskaya, O. G., Myers, C. L., et al., (2014) Broad metabolic sensitivity profiling of a prototrophic yeast deletion collection. *Genome Biol 15*, R64.

TABLES AND FIGURES

Table 1. Experimental factors and the associated levels of each chosen for evaluation for their impact on ethanol metabolism.

Factor	Role	Factor Level	
ADH1	Categorical	OFF (Δ*adh1*)	ON (wild type)
ADH2	Categorical	OFF (Δ*adh2*)	ON (wild type)
ADH3	Categorical	OFF (Δ*adh3*)	ON (wild type)
ADH4	Categorical	OFF (Δ*adh4*)	ON (wild type)
ADH5	Categorical	OFF (Δ*adh5*)	ON (wild type)
SFA1	Categorical	OFF (Δ*sfa1*)	ON (wild type)
Aeration	Categorical	Nitrogen	Air
Dilution rate (D)	Continuous	0.06	0.29
Glucose (g l^{-1})	Continuous	7.5	25
Temperature (°C)	Continuous	30	33
Zinc (μM)	Continuous	1.0	38
Media pH	Continuous	4.5	5.5

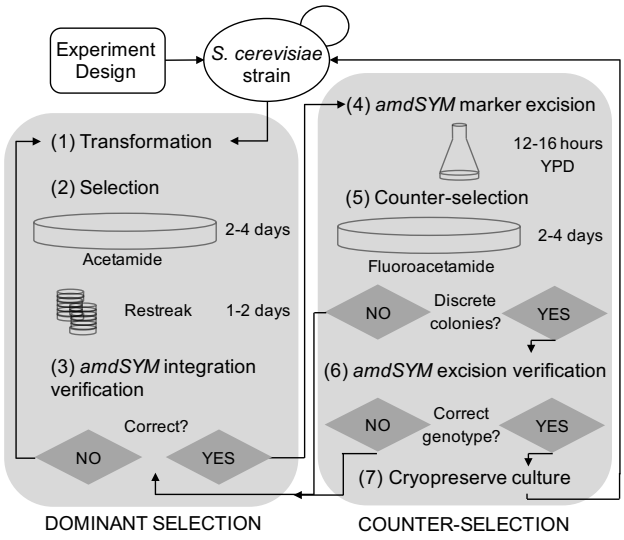
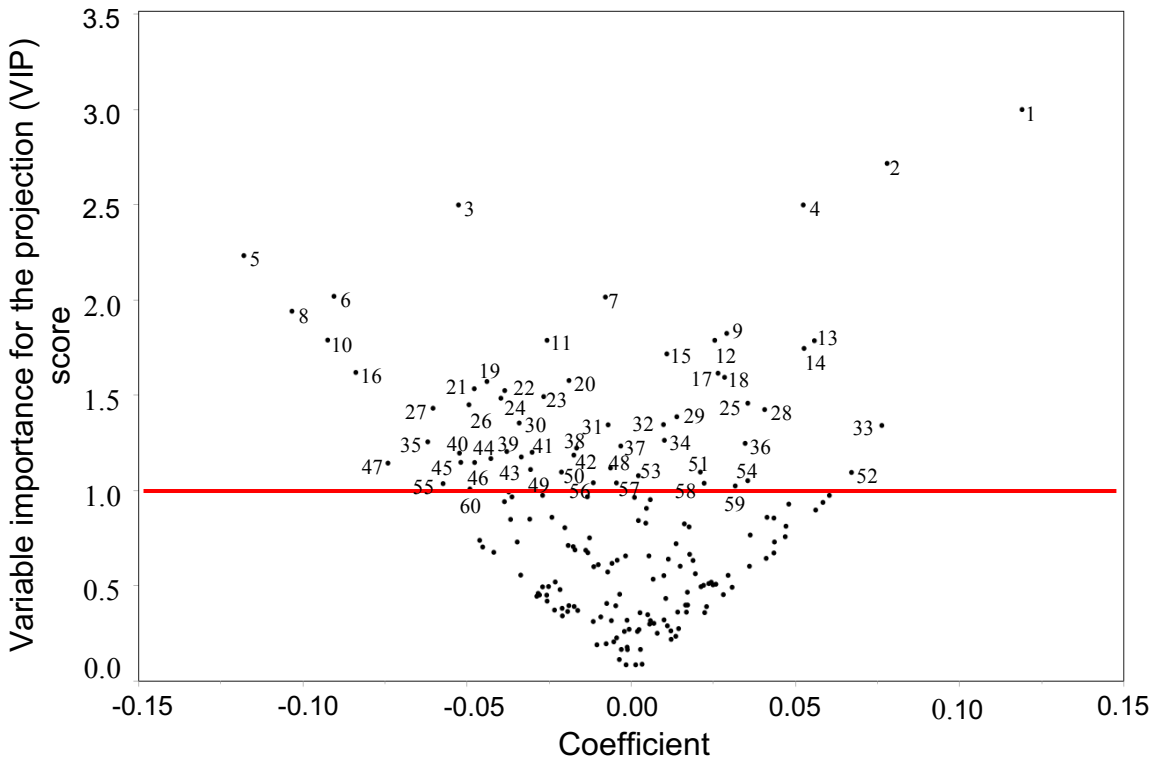


Figure 1. Workflow for generating *S. cerevisiae* combinatorial ADH isozyme gene deletions. Workflow was based on the use of the *amdSYM* counter-selectable marker. The left-hand panel of the workflow shows the dominant selection process: (1) Transformation using ADH gene-specific linearised deletion cassettes; (2) Selection using acetamide as the sole nitrogen source; (3) Verification of correct *amdSYM* integration by diagnostic colony PCR on three transformants. The right-hand panel of the workflow shows subsequent *amdSYM* removal by counter-selection: (4) *amdSYM* removal using homologous recombination by culturing in a non-selective medium to promote marker excision; (5) Counter-selection using fluoroacetamide; (6) Verification of *amdSYM* cassette removal by replica-plating and diagnostic colony PCR on discrete colonies. Finally (7) correctly identified genotypes were cryopreserved. Checkpoints after stages (3), (5) and (6) were employed for quality control.



1	*ADH3* * Aeration [N_2]	16	*ADH3* * Aeration [Air]	31	*SFA1* * Aeration [Air]	46	*ADH5* *Glucose
2	*ADH1* * Aeration [N_2]	17	Δ*adh3* * Aeration [Air]	32	*SFA1* * Aeration [N_2]	47	Zinc * Glucose
3	Aeration [Air]	18	*ADH5* * Aeration [N_2]	33	Δ*adh1* * Glucose	48	Δ*adh5* * Glucose
4	Aeration [N_2]	19	Δ*adh1* * Δ*adh3*	34	*ADH1* * Δ*adh5*	49	Dilution rate
5	*ADH1* * Glucose	20	*ADH5* * Aeration [Air]	35	Δ*adh5* * Dilution rate	50	Δ*adh3*
6	*ADH3* * Glucose	21	Δ*adh1* * Δ*adh4*	36	*ADH4* * Aeration [N_2]	51	*ADH3*
7	Δ*adh1* * Aeration [Air]	22	Glucose	37	Δ*adh1* * Δ*adh2*	52	Δ*adh4* * Zinc
8	Aeration [N_2] * Glucose	23	Δ*adh4* * Aeration [Air]	38	*ADH2* * Aeration [Air]	53	Δ*adh1* * Δ*sfa1*
9	Δ*adh4* * Aeration [N_2]	24	Δ*adh2* * Aeration [Air]	39	Δ*adh5* * Aeration [Air]	54	Aeration [Air] * Glucose
10	*ADH1* * Dilution rate	25	Δ*adh5* * Aeration [N_2]	40	Δ*adh4* * Dilution rate	55	Δ*adh3* * Aeration [N_2]
11	Δ*adh1*	26	Δ*sfa1* * Aeration [Air]	41	*ADH4* * Aeration [Air]	56	Δ*sfa1* * Glucose
12	*ADH1*	27	Δ*adh2* * Glucose	42	Δ*adh1* * Δ*adh5*	57	*ADH1* * Δ*sfa1*
13	Δ*sfa1* * Aeration [N_2]	28	*ADH1* * Δ*adh4*	43	Δ*adh1* * *SFA1*	58	*ADH1* * *ADH2*
14	Δ*adh2* * Aeration [N_2]	29	*ADH2* * Aeration [N_2]	44	*SFA1* * Glucose	59	*ADH1* * *SFA1*
15	*ADH1* * *ADH3*	30	Δ*adh4* * Glucose	45	Δ*adh5* * Media pH	60	*ADH1* * Aeration [Air]

Figure 2. Model analysis for ethanol production in the ADH gene deletion library using the ministat bioprocess equipment. The variable importance for the projection (VIP) scores for experimental predictors are plotted against the centred and scaled data coefficients. Standardised coefficients indicate if a predictor is having a positive or negative impact, as well as the magnitude of that impact on the measured response. Predictors above the VIP value of 1.0 (red line) are important to the explanatory model for ethanol production.
* indicates an interaction of the factors.

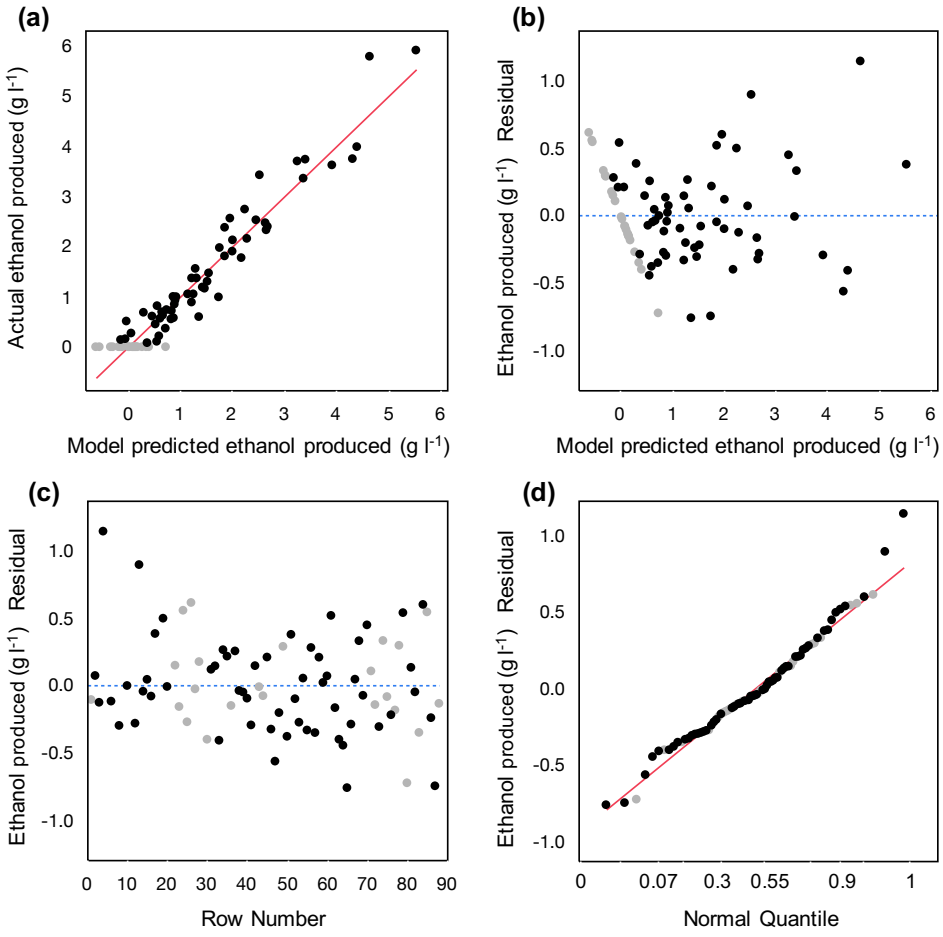


Figure 3. Diagnostic plots assessing the fit of the PLS model. The model evaluates the effect of ADH gene deletions in combination with different environmental factors on ethanol production in *S. cerevisiae*. (a) The observed (actual) data for ethanol production are plotted against the values predicted by the model (n = 81, R^2 = 0.932). The plots of Residual by Predicted (b), Residual by Row (c), and Residual Normal Quantile (d) show the distribution of the data to be homoscedastic. The grey data points in each of the plots represent 21 experiments where zero ethanol was produced, a subset of 15 experiments were cultures that performed cellular respiration without exhibiting the Crabtree effect due to the dilution rate set below the $D_{critical}$ threshold for mixed respire-fermentative growth. A comparison of the root mean PRESS statistic (minimised) was made during model selection.

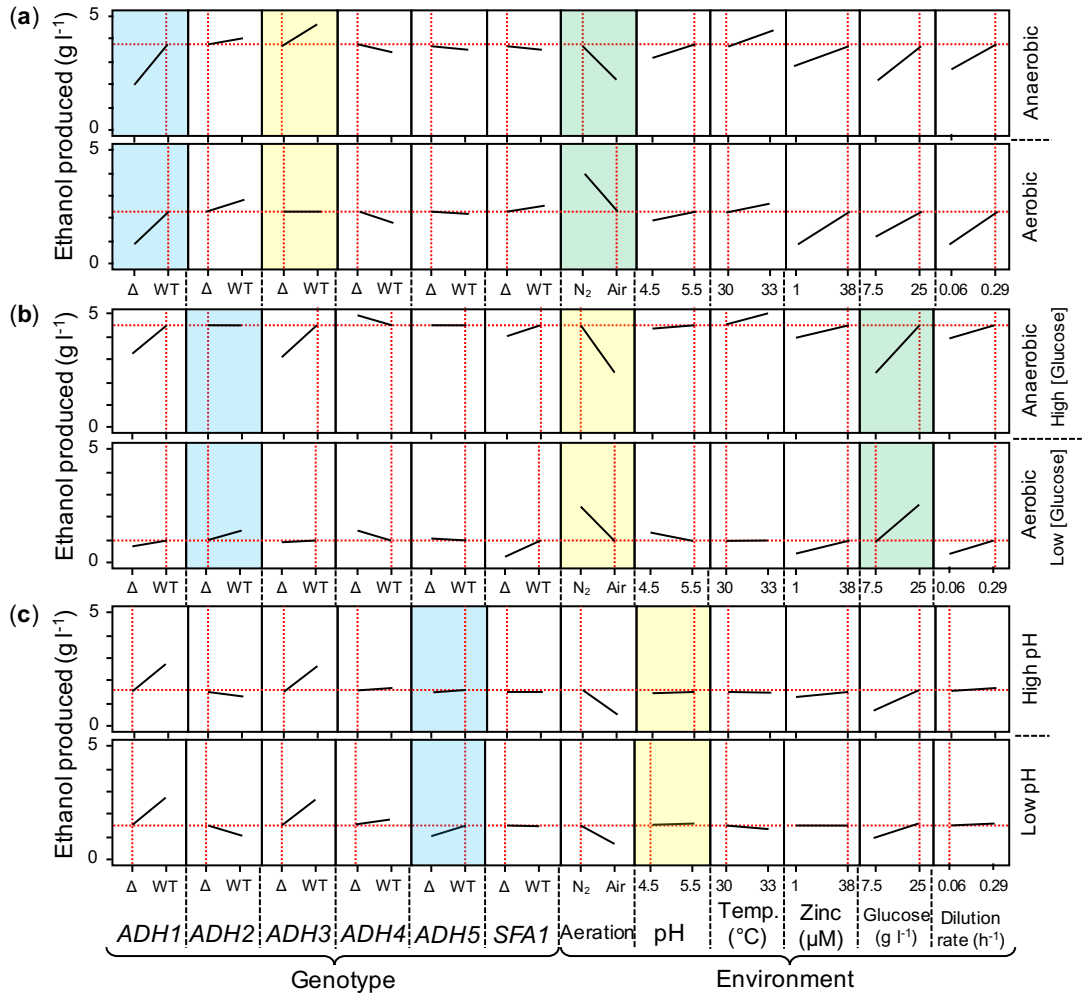


Figure 4. Prediction profiler for the model projection of ethanol production in the *Saccharomyces cerevisiae* ADH gene deletion library. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates ethanol production dependent upon the validated model prediction formula. (**a**) The importance of *ADH1* and *ADH3* on ethanol production. In the upper panel cultures are anaerobic (i.e. aeration is set to nitrogen), in the lower panel cultures are grown under aerobic conditions (i.e. aeration is set to air). These setting are highlighted in green. Every other factor is constant. The blue boxes indicate that under both aeration conditions *ADH1* has a large positive impact on ethanol production. The yellow boxes indicate that *ADH3* has a positive impact on ethanol production only under anaerobic conditions. (**b**) The importance of *ADH2* on model prediction. In this instance the upper panel represents cultures grown under anaerobic conditions in the presence of high glucose (yellow and green boxes respectively); the lower panel represents cultures under aerobic conditions and low glucose. Every other factor is constant. The blue boxes indicate that *ADH2* does not contribute to our model of ethanol production under anaerobic aeration and in high glucose medium. (**c**) The interaction between *ADH5* and pH. Here the upper panel represents cultures grown at pH 5.5 (yellow box), typical of laboratory culture conditions. In these circumstances the contribution of *ADH5* to ethanol production is minimal (blue box). The lower panel represents cultures grown at pH 4.5. Under these conditions *ADH5* has a positive impact on ethanol production.

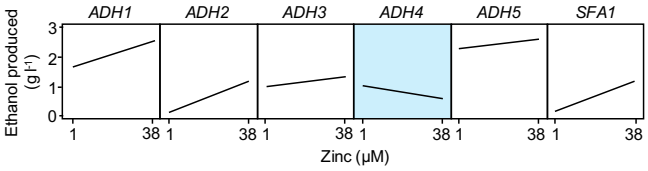
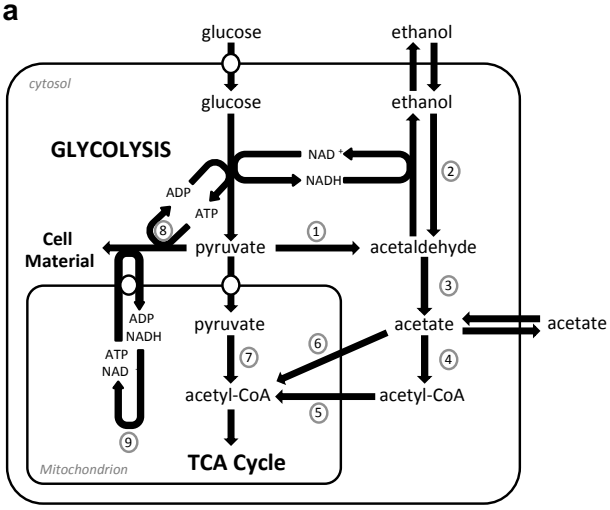


Figure 5. Interaction between each of the ADH isozymes in zinc limiting and zinc replete media. The interaction of *ADH4* is highlighted in blue. In contrast to the other isozymes, Adh4 produces more ethanol under zinc limited than zinc replete conditions.

SUPPORTING FIGURES

a



b

Enzymes directly involved in step ② of glucose metabolism in *Saccharomyces cerevisiae*

Gene	ORF	Functions
ADH1	YOL086c	Main cytosolic ADH, forms ethanol in glycolysis
ADH2	YMR303c	Cytosolic, glucose repressed, used for growth on ethanol
ADH3	YMR083w	Mitochondrial, role as redox shuttle
ADH4	YGL256w	Induced in response to zinc deficiency
ADH5	YBR145w	Paralog of *ADH1*, arose from whole genome duplication
SFA1	YDL168w	Part of a bifunctional formaldehyde dehydrogenase

Text

Fig S1. (**a**) Glucose metabolism in *Saccharomyces cerevisiae* adapted from [1] ① Pyruvate decarboxylase; ② alcohol dehydrogenase(s); ③ acetaldehyde dehydrogenase(s); ④ acetyl-CoA synthetase (cytoplasmic); ⑤ transport of acetyl-CoA into mitochondria via the carnitine shuttle; ⑥ transport of acetate into mitochondria and formation of acetyl-CoA via mitochondrial acetyl-CoA synthetase; ⑦ pyruvate dehydrogenase complex; ⑧ pyruvate carboxylase; ⑨ formation of ATP via oxidative phosphorylation. (**b**) The alcohol dehydrogenase genes of *Saccharomyces cerevisiae* adapted from [2].

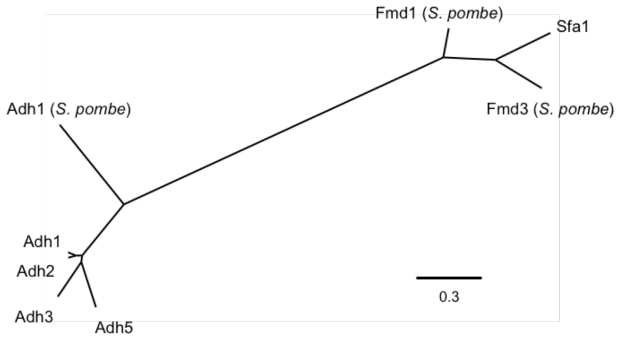


Fig S2. Phylogenetic tree of the amino acid sequences of ADH isozymes of *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* specifically involved in ethanol metabolism. Adh4 proteins have little sequence similarity to the other ADH isozymes and are not included. Branch lengths are proportional to the number of substitutions per site, as indicated by the marker bar. Phylogenetic analysis of ADH isozyme protein sequence performed according to method as 3.

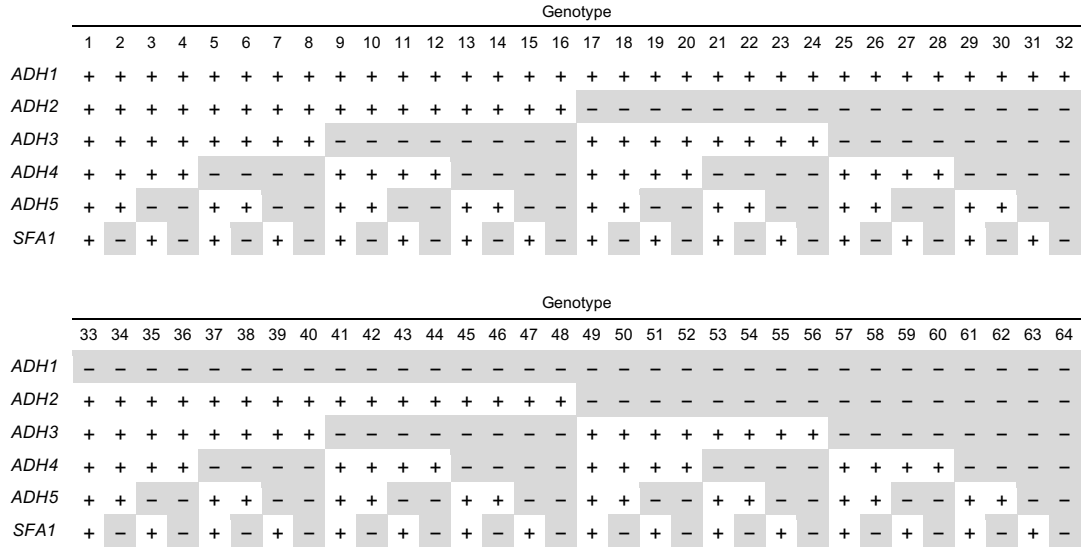


Fig S3. An orthogonal array highlighting the 64 genotypes required for the evaluation of all permutations of the ADH isozymes of the yeast *Saccharomyces cerevisiae*.

(+) represent intact ADH isozyme genes; (-) represent gene deletions.

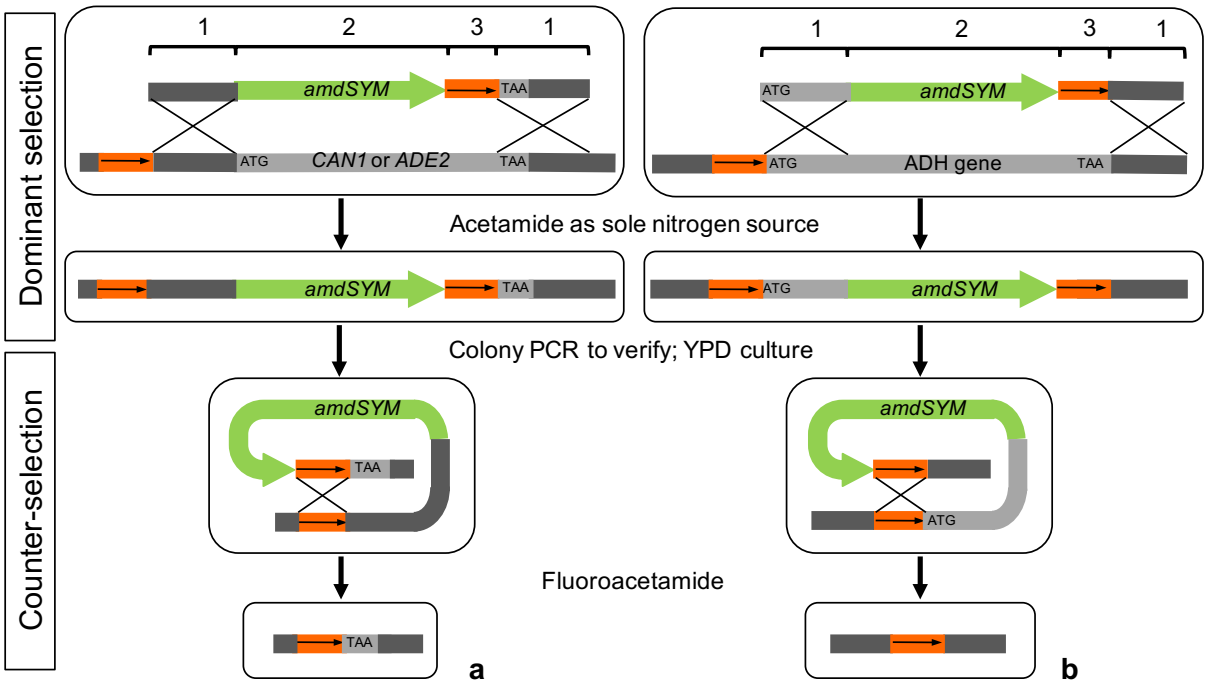


Fig S4. The *amdSYM* method of gene deletion in *Saccharomyces cerevisiae*. The deletion cassette includes **1**) 50-55 base pairs (bp) gene-specific sequences for homologous recombination; **2**) *amdSYM* marker module; and **3**) 40 bp direct repeat sequence for scarless marker recovery. (**a**) As described by Solis-Escalante, et al. [4]; note that this leads to deletion of upstream flanking sequences of the target gene (dark grey) during the counter-selection step. (**b**) Adaptation of homologous recombination sequences, which after marker excision permit exact open reading frame (ORF) deletion with no loss of upstream flanking sequences.

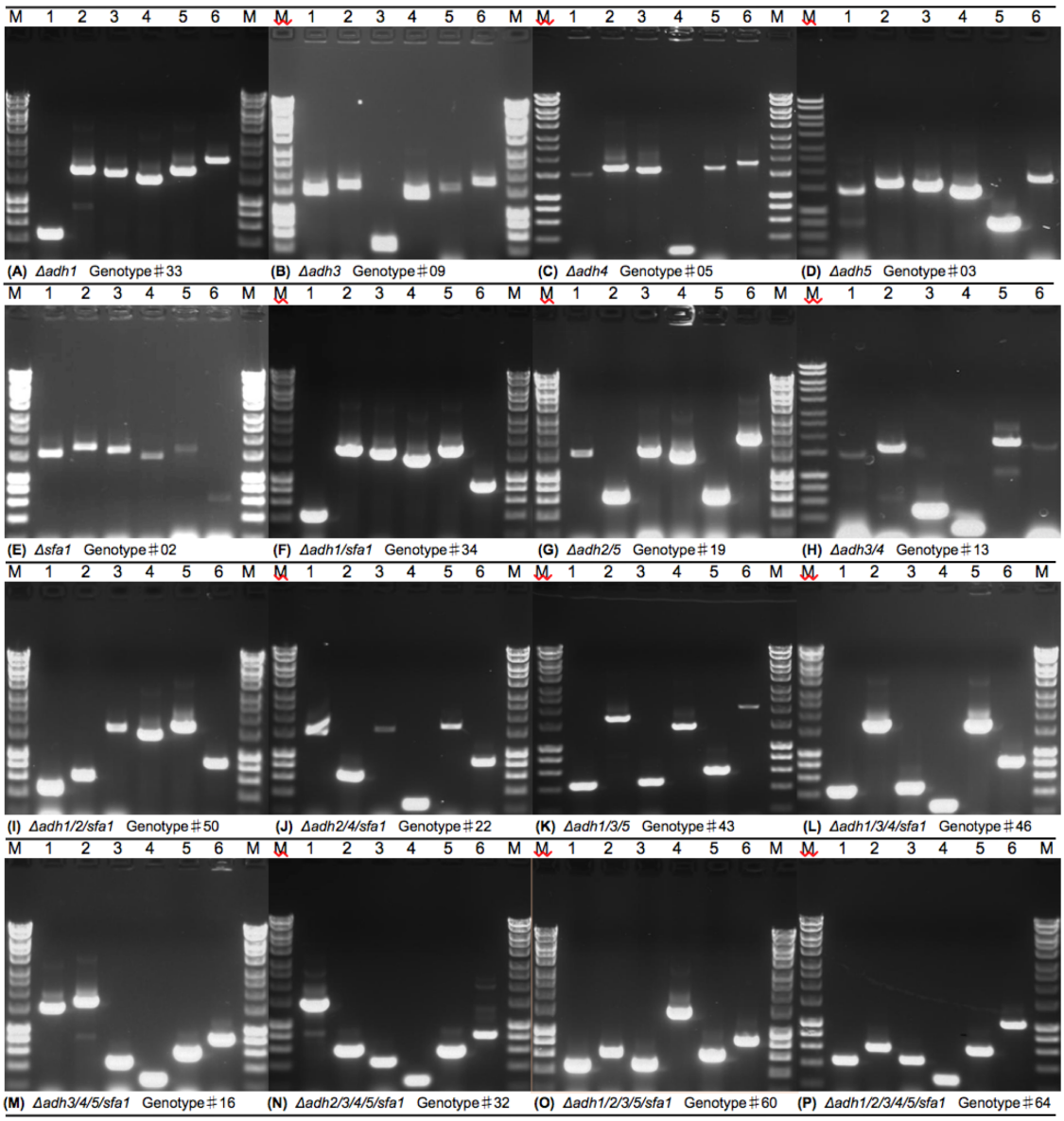


Fig S5. Diagnostic colony PCR verification examples of ADH knockout genotypes. All PCRs were performed at the standardised annealing temperature of 63.3 °C. Diagnostic fragment sizes are: Lane 1, *ADH1* (1,509 bp), ***Δadh1*** (462 bp); Lane 2, *ADH2* (1,675 bp), ***Δadh2*** (628 bp); Lane 3, *ADH3* (1,609 bp), ***Δadh3*** (482 bp); Lane 4, *ADH4* (1,426 bp), ***Δadh4*** (277 bp); Lane 5, *ADH5* (1,678 bp), ***Δadh5*** (672 bp); Lane 6, *SFA1* (1,997 bp), ***Δsfa1*** (836 bp). The genotype number corresponds to the orthogonal array list of required genotypes, Figure S3.

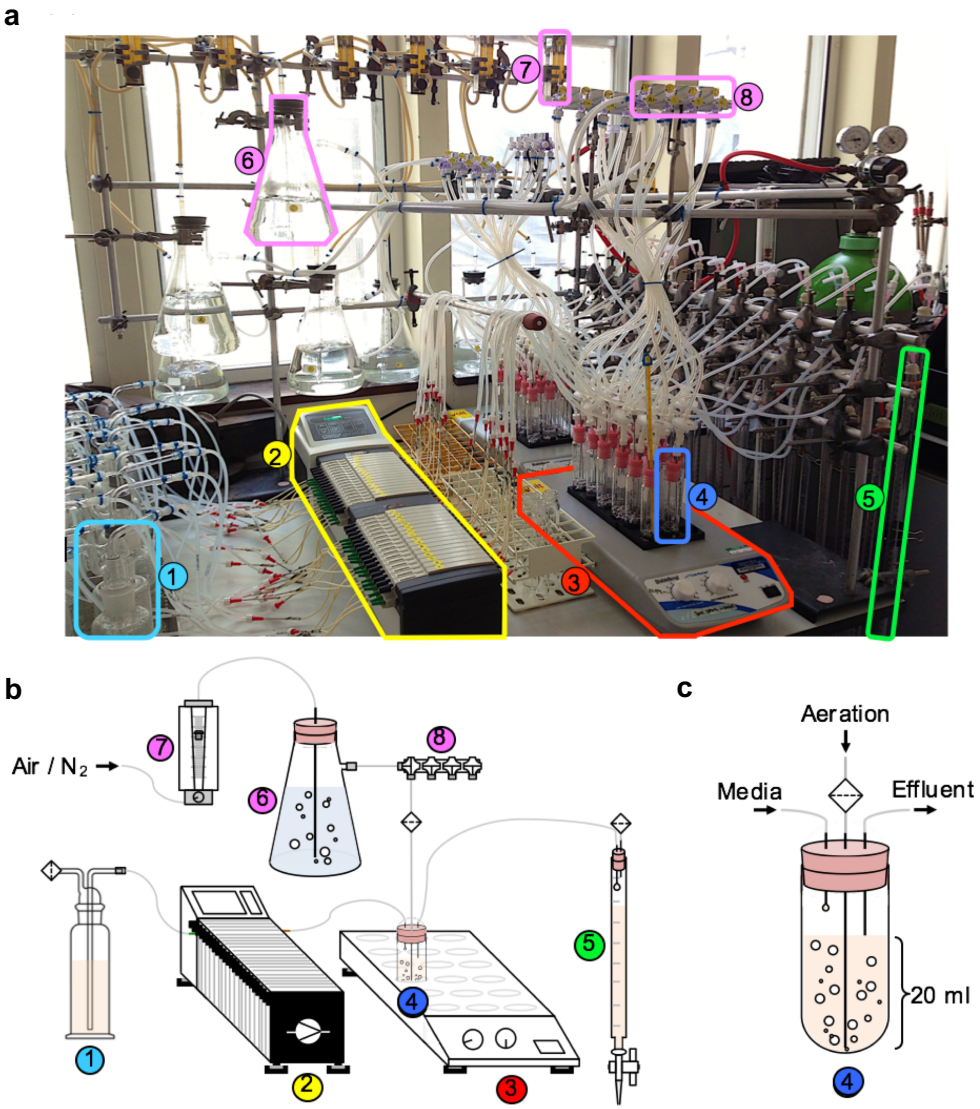


Fig S6. The ministat equipment including: (**a**) in situ photograph; (**b**) a detailed schematic of the units of operation; and (**c**) a 20 ml ministat bioreactor.

1. Drechsel bottle for sterile media (×32);
2. a 32 channel peristaltic pump;
3. heatblock (×2);
4. 20 ml working volume ministat bioreactors (×32);
5. burette (×32);
6. gas washing bottle (×8);
7. gas rotameter (×8);
8. gas four-port manifold (×8).

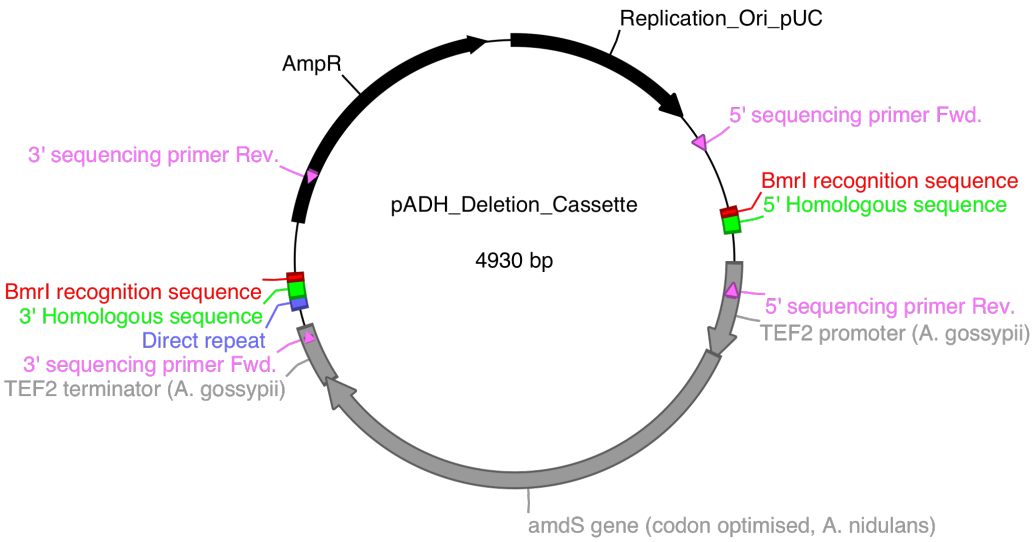


Fig S7. Plasmid map of the pADH_Deletion_Cassette plasmid used for the deletion of the alcohol dehydrogenase isozymes in *Saccharomyces cerevisiae* CEN.PK113-7D. Isozyme-specific homologous recombination and direct repeat marker excision sequences are detailed in SI Table 1. The bacterial amplification backbone was the pJ154 series of expression vector (DNA2.0 Inc., California, US), allowing the use of the type II restriction enzyme, *BmrI*, for linearization of the deletion cassette. The *amdS* gene is the prototrophic selection marker, in *Saccharomyces cerevisiae*.

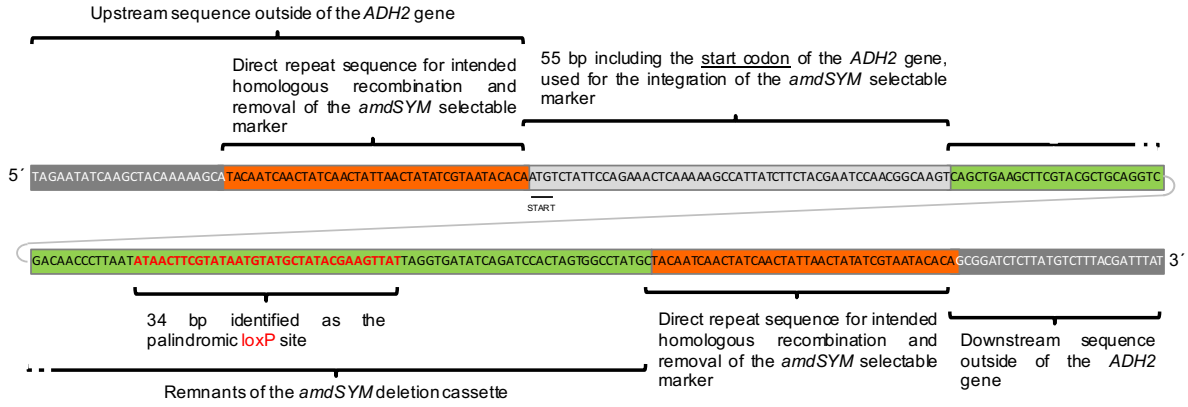


Fig S8. Annotated sequence of an incorrect excision of the Δ*adh2*::*amdSYM* selectable marker. The features described were present in all incorrect Δ*adh2*::*amdSYM* excisions, and for gene deletions of the other five ADH genes. Highlight colours match those in SI Figure 4.

SUPPORTING TABLES

Table S1. Raw data evaluating the impact on ethanol metabolism of the ADH isozymes and environmental factors using the ministat bioprocess equipment.

Whole Plots	Subplots	ADH genotype (Wild Type or deleted)						Aeration	Media pH	Temp. (ºC)	Zinc (µM)	Glucose (g/l)	Dilution rate	Ethanol produced (g/l)	Glucose consumed (g/l)	Yp/s (g/g)	Cell count at culture harvest ($\times 10^{8}$)	Total free fatty acids (mg/l)
		ADH1	ADH2	ADH3	ADH4	ADH5	SFA1											
1	1	Δ	Δ	Δ	Δ	Δ	Δ	Air	5.5	33	0.1	7.5	0.29	0.00	4.42	0.00	0.62	15.91
1	1	WT	Δ	Δ	WT	WT	Δ	Air	4.5	33	0.1	7.5	0.29	1.00	7.38	0.14	1.19	24.05
1	1	WT	WT	Δ	WT	Δ	Δ	Nitrogen	5.5	33	38.0	7.5	0.29	2.16	7.50	0.29	1.12	27.59
1	1	WT	Δ	WT	WT	Δ	Δ	Nitrogen	5.5	33	0.1	25.0	0.29	5.79	20.30	0.29	2.19	49.24
1	1	Δ	Δ	Δ	Δ	WT	Δ	Nitrogen	4.5	33	38.0	7.5	0.29	0.00	0.02	0.00	0.00	1.74
1	1	Δ	WT	WT	WT	WT	Δ	Air	5.5	33	38.0	25.0	0.29	0.72	8.18	0.09	0.32	11.44
1	1	Δ	Δ	WT	Δ	WT	WT	Nitrogen	5.5	33	0.1	25.0	0.29	0.00	0.00	0.00	0.01	3.85
1	1	Δ	Δ	Δ	Δ	WT	WT	Air	4.5	33	0.1	7.5	0.29	0.58	2.75	0.21	0.36	6.68
1	1	Δ	WT	Δ	WT	WT	WT	Nitrogen	4.5	33	0.1	25.0	0.29	0.00	0.00	0.00	0.01	2.78
1	1	Δ	WT	WT	Δ	Δ	Δ	Air	4.5	33	38.0	7.5	0.29	0.74	6.79	0.11	0.71	14.95
1	1	Δ	WT	Δ	Δ	WT	WT	Nitrogen	5.5	33	38.0	7.5	0.29	0.00	0.00	0.00	0.01	2.96
1	2	WT	Δ	WT	WT	WT	WT	Nitrogen	4.5	30	0.1	7.5	0.29	2.40	7.50	0.32	1.60	35.97
1	2	WT	WT	Δ	Δ	Δ	Δ	Air	5.5	30	38.0	25.0	0.29	3.43	12.71	0.27	0.69	22.19
1	2	WT	Δ	Δ	WT	WT	Δ	Air	5.5	30	0.1	25.0	0.29	0.85	1.39	0.61	0.34	6.21
1	2	Δ	WT	Δ	WT	Δ	WT	Air	4.5	30	38.0	7.5	0.29	0.70	7.02	0.10	0.56	18.06
1	2	WT	WT	Δ	WT	WT	WT	Nitrogen	5.5	30	38.0	7.5	0.29	1.47	7.50	0.20	0.93	21.19
1	2	WT	Δ	Δ	Δ	Δ	WT	Air	5.5	30	0.1	7.5	0.29	0.69	3.81	0.18	0.64	9.10
1	2	Δ	WT	WT	Δ	WT	Δ	Nitrogen	5.5	30	0.1	7.5	0.29	0.00	0.06	0.00	0.00	2.92
1	2	WT	WT	Δ	Δ	WT	WT	Nitrogen	5.5	30	0.1	25.0	0.29	2.74	25.00	0.11	0.66	17.51
1	2	WT	Δ	WT	Δ	Δ	WT	Nitrogen	4.5	30	0.1	25.0	0.29	3.36	14.34	0.23	1.33	33.76
1	2	Δ	WT	WT	Δ	WT	WT	Nitrogen	4.5	30	38.0	25.0	0.29	0.00	0.00	0.00	0.00	2.86
1	2	Δ	WT	Δ	Δ	Δ	Δ	Air	4.5	30	0.1	25.0	0.29	0.00	3.15	0.00	0.26	10.64
2	3	Δ	WT	Δ	Δ	Δ	WT	Nitrogen	4.5	33	38.0	25.0	0.06	0.00	0.00	0.00	0.01	1.67
2	3	Δ	Δ	WT	Δ	WT	Δ	Air	5.5	33	38.0	7.5	0.06	0.00	7.50	0.00	1.22	15.48
2	3	Δ	WT	WT	Δ	WT	WT	Air	4.5	33	0.1	25.0	0.06	0.00	7.95	0.00	0.92	15.22
2	3	WT	Δ	WT	Δ	WT	Δ	Air	4.5	33	0.1	7.5	0.06	0.00	7.18	0.00	0.57	14.66
2	3	WT	WT	WT	WT	Δ	WT	Air	4.5	33	38.0	7.5	0.06	0.00	7.50	0.00	1.05	12.20
2	3	WT	Δ	Δ	WT	WT	WT	Air	5.5	33	38.0	7.5	0.06	0.00	7.50	0.00	2.19	17.37
2	3	Δ	Δ	Δ	WT	Δ	Δ	Nitrogen	4.5	33	0.1	25.0	0.06	0.00	0.00	0.00	0.00	1.02
2	3	WT	WT	Δ	Δ	WT	Δ	Air	5.5	33	0.1	25.0	0.06	0.00	13.94	0.00	0.94	16.71
2	3	Δ	WT	WT	Δ	Δ	Δ	Nitrogen	5.5	33	0.1	25.0	0.06	2.13	25.00	0.09	0.85	10.93
2	3	WT	WT	Δ	Δ	Δ	Δ	Nitrogen	4.5	33	0.1	7.5	0.06	1.37	7.50	0.18	1.48	15.65
2	3	WT	Δ	WT	Δ	WT	WT	Nitrogen	4.5	33	38.0	25.0	0.06	3.99	25.00	0.16	2.40	20.55
2	4	Δ	WT	Δ	WT	Δ	Δ	Nitrogen	4.5	30	0.1	7.5	0.06	1.56	7.50	0.21	0.50	6.23
2	4	Δ	WT	WT	WT	WT	WT	Nitrogen	4.5	30	38.0	7.5	0.06	1.98	7.50	0.26	0.57	9.86
2	4	WT	WT	Δ	Δ	Δ	WT	Air	4.5	30	0.1	25.0	0.06	0.00	7.67	0.00	0.81	13.02
2	4	WT	WT	WT	WT	Δ	Δ	Air	4.5	30	0.1	25.0	0.06	0.82	11.21	0.07	0.70	11.93
2	4	Δ	WT	WT	WT	Δ	Δ	Air	5.5	30	38.0	7.5	0.06	0.63	7.50	0.08	1.77	14.52
2	4	Δ	WT	Δ	WT	WT	WT	Air	5.5	30	38.0	25.0	0.06	0.57	25.00	0.02	0.90	15.25
2	4	WT	Δ	Δ	WT	Δ	WT	Nitrogen	4.5	30	0.1	7.5	0.06	1.05	7.50	0.14	1.01	18.13
2	4	WT	WT	WT	Δ	WT	Δ	Nitrogen	5.5	30	38.0	25.0	0.06	3.62	25.00	0.14	3.19	28.02
2	4	Δ	Δ	WT	Δ	Δ	WT	Air	5.5	30	38.0	25.0	0.06	0.61	25.00	0.02	0.89	15.74
2	4	Δ	WT	Δ	Δ	WT	Δ	Air	5.5	30	0.1	7.5	0.06	0.00	6.77	0.00	1.43	12.89
2	4	WT	Δ	WT	WT	WT	Δ	Air	4.5	30	38.0	7.5	0.06	0.00	7.50	0.00	0.18	23.56

Whole Plots	Subplots	ADH genotype (Wild Type or deleted)						Aeration	Media pH	Temp. (ºC)	Zinc (µM)	Glucose (g/l)	Dilution rate	Ethanol produced (g/l)	Glucose consumed (g/l)	Yp/s (g/g)	Cell count at culture harvest ($\times 10^{8}$)	Total free fatty acids (mg/l)
		ADH1	ADH2	ADH3	ADH4	ADH5	SFA1											
3	5	Δ	WT	WT	WT	WT	WT	Air	5.5	33	0.1	7.5	0.29	0.27	3.95	0.07	0.53	8.64
3	5	WT	Δ	WT	Δ	WT	Δ	Nitrogen	5.5	33	38.0	7.5	0.29	2.33	7.50	0.31	1.20	25.06
3	5	WT	WT	WT	Δ	WT	Δ	Nitrogen	4.5	33	0.1	25.0	0.29	3.75	18.09	0.21	1.03	31.69
3	5	WT	Δ	WT	WT	Δ	WT	Air	5.5	33	38.0	7.5	0.29	1.05	7.50	0.14	1.74	28.07
3	5	Δ	Δ	Δ	WT	Δ	WT	Nitrogen	5.5	33	38.0	25.0	0.29	0.00	4.10	0.00	0.00	2.49
3	5	Δ	Δ	WT	Δ	WT	WT	Air	4.5	33	38.0	7.5	0.29	0.22	7.50	0.03	0.68	22.41
3	5	WT	WT	WT	Δ	Δ	WT	Nitrogen	5.5	33	38.0	25.0	0.29	5.91	24.51	0.00	0.90	34.73
3	5	WT	Δ	Δ	Δ	Δ	Δ	Air	4.5	33	38.0	25.0	0.29	1.90	16.11	0.12	0.47	21.54
3	5	Δ	Δ	WT	WT	Δ	WT	Air	4.5	33	0.1	25.0	0.29	0.55	7.30	0.00	0.53	11.43
3	5	Δ	WT	WT	WT	Δ	WT	Nitrogen	4.5	33	0.1	7.5	0.29	1.37	7.50	0.18	1.00	18.35
3	5	WT	WT	Δ	WT	Δ	WT	Air	5.5	33	0.1	25.0	0.29	0.89	12.75	0.00	0.91	14.96
3	6	Δ	Δ	Δ	Δ	Δ	Δ	Nitrogen	5.5	30	38.0	25.0	0.29	0.14	3.25	0.04	0.01	1.39
3	6	Δ	WT	Δ	WT	WT	Δ	Nitrogen	4.5	30	38.0	25.0	0.29	0.37	11.71	0.03	0.14	10.71
3	6	Δ	Δ	Δ	Δ	WT	Δ	Air	5.5	30	38.0	7.5	0.29	0.16	4.68	0.03	0.44	14.72
3	6	Δ	Δ	WT	WT	Δ	Δ	Air	4.5	30	0.1	7.5	0.29	0.93	6.71	0.14	1.34	25.73
3	6	WT	WT	WT	WT	WT	WT	Air	4.5	30	38.0	25.0	0.29	2.53	19.05	0.13	0.52	20.18
3	6	WT	Δ	Δ	Δ	WT	Δ	Nitrogen	4.5	30	0.1	7.5	0.29	2.38	7.50	0.32	1.47	33.28
3	6	WT	Δ	Δ	Δ	WT	WT	Air	4.5	30	38.0	25.0	0.29	2.47	17.87	0.14	0.71	23.28
3	6	WT	WT	Δ	Δ	Δ	WT	Nitrogen	4.5	30	38.0	7.5	0.29	1.78	7.50	0.24	0.85	21.46
3	6	Δ	WT	WT	WT	Δ	WT	Nitrogen	5.5	30	0.1	25.0	0.29	0.11	11.71	0.01	0.00	5.90
3	6	WT	Δ	WT	Δ	Δ	Δ	Air	5.5	30	0.1	25.0	0.29	0.60	8.99	0.00	0.81	19.08
3	6	Δ	Δ	Δ	Δ	WT	WT	Nitrogen	4.5	30	0.1	7.5	0.29	0.08	1.52	0.05	0.02	3.05
4	7	Δ	WT	WT	Δ	Δ	Δ	Air	4.5	30	38.0	25.0	0.06	0.70	24.65	0.00	0.18	29.59
4	7	WT	Δ	WT	WT	Δ	WT	Nitrogen	5.5	30	38.0	25.0	0.06	3.74	24.75	0.15	1.21	28.32
4	7	WT	Δ	WT	Δ	WT	WT	Air	5.5	30	38.0	7.5	0.06	0.45	7.26	0.06	0.86	23.98
4	7	Δ	Δ	WT	WT	WT	Δ	Nitrogen	4.5	30	0.1	25.0	0.06	3.70	24.76	0.15	1.68	30.05
4	7	Δ	Δ	WT	Δ	Δ	WT	Nitrogen	5.5	30	0.1	7.5	0.06	0.00	0.41	0.00	0.01	4.62
4	7	Δ	Δ	Δ	Δ	Δ	Δ	Air	4.5	30	38.0	7.5	0.06	0.00	7.26	0.00	1.14	28.84
4	7	Δ	Δ	Δ	WT	WT	WT	Nitrogen	4.5	30	38.0	25.0	0.06	1.17	9.04	0.00	0.18	10.08
4	7	Δ	WT	WT	Δ	Δ	WT	Air	4.5	30	0.1	7.5	0.06	0.00	7.25	0.00	1.47	29.53
4	7	Δ	Δ	Δ	WT	Δ	Δ	Air	5.5	30	0.1	25.0	0.06	0.00	24.76	0.00	1.92	44.40
4	7	WT	WT	WT	WT	WT	Δ	Nitrogen	5.5	30	0.1	7.5	0.06	1.30	7.26	0.18	1.11	22.83
4	7	WT	WT	Δ	WT	WT	WT	Air	4.5	30	0.1	7.5	0.06	0.00	7.16	0.00	1.18	17.63
4	8	Δ	Δ	Δ	Δ	WT	WT	Air	5.5	33	0.1	25.0	0.06	0.00	24.44	0.00	0.46	32.48
4	8	Δ	WT	Δ	WT	Δ	WT	Nitrogen	5.5	33	0.1	7.5	0.06	0.51	1.89	0.27	0.07	4.12
4	8	Δ	Δ	Δ	WT	WT	Δ	Nitrogen	5.5	33	0.1	7.5	0.06	0.00	0.00	0.00	0.01	3.47
4	8	WT	Δ	Δ	Δ	Δ	WT	Nitrogen	5.5	33	38.0	7.5	0.06	1.00	7.24	0.14	0.32	13.41
4	8	WT	WT	WT	Δ	WT	WT	Nitrogen	5.5	33	0.1	7.5	0.06	1.81	7.24	0.25	0.25	14.84
4	8	Δ	Δ	Δ	WT	WT	Δ	Air	4.5	33	38.0	25.0	0.06	0.00	24.64	0.00	0.26	21.01
4	8	WT	WT	Δ	WT	WT	Δ	Nitrogen	4.5	33	38.0	25.0	0.06	2.56	24.70	0.10	0.79	27.51
4	8	WT	WT	WT	Δ	WT	Δ	Air	4.5	33	38.0	7.5	0.06	0.00	7.25	0.00	0.33	18.07
4	8	WT	Δ	WT	WT	WT	WT	Air	5.5	33	0.1	25.0	0.06	1.19	22.72	0.05	0.61	30.57
4	8	WT	Δ	WT	WT	Δ	Δ	Nitrogen	4.5	33	38.0	7.5	0.06	0.99	7.25	0.14	0.78	18.78
4	8	Δ	Δ	Δ	WT	Δ	WT	Air	4.5	33	0.1	7.5	0.06	0.00	7.16	0.00	0.59	17.51

Table S2. The predictors with the greatest VIP (variable importance for the projection) scores for the modelled responses of ethanol production in *S. cerevisiae*, derived from the ministat bioprocess equipment evaluation in which all environmental factors were kept constant with the exception of culture aeration (air or nitrogen). The model coefficients for each predictor are not shown.

Predictors with a **negative** impact on ethanol production		Predictors with a **positive** impact on ethanol production	
Factor	**VIP Score**	**Factor**	**VIP Score**
Δ*adh1* * Δ*adh2* ◇	1.98	Aeration [Nitrogen]	1.83
Aeration [Air]	1.83	*ADH1*	1.80
Δ*adh1*	1.80	*ADH1* * Aeration [Nitrogen] ◇	1.53

◇ * indicates an interaction between the factors shown

Table S3. The pADH_Deletion_Cassette plasmid homologous recombination sequence variants used for the deletion of the alcohol dehydrogenase isozymes in *Saccharomyces cerevisiae* CEN.PK113-7D and the specific direct repeat sequences for subsequent marker excision. The two 5´ sequences for the ADH1 deletion cassette were (1) the original sequence and (2) the extended 100 bp homologous sequence used to increase integration efficiency.

Gene	Name	Sequence (5´ → 3´)
ADH1	5´sequence(1)	ATGTCTATCCCAGAAACTCAAAAAGGTGTTATCTTCTACGAATCCCACGGTAAGT
	5´sequence(2)	ATGTCTATCCCAGAAACTCAAAAAGGTGTTATCTTCTACGAATCCCACGGTAAGTT GGAATACAAAGATATTCCAGTTCCAAAGCCAAAGGCCAACGAAT
	3´ sequence	TTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTA
	Direct repeat	TCAAGCTATACCAAGCATACAATCAACTATCTCATATACA
ADH2	5´sequence	ATGTCTATTCCAGAAACTCAAAAAGCCATTATCTTCTACGAATCCAACGGCAAGT
	3´ sequence	TCTCTTATGTCTTTACGATTTATAGTTTTCATTATCAAGTATGCCTATATTAGTA
	Direct repeat	TACAATCAACTATCAACTATTAACTATATCGTAATACACA
ADH3	5´sequence	ATGTTGAGAACGTCAACATTGTTCACCAGGCGTGTCCAACCAAGCCTATTTTCTA
	3´ sequence	TGTTACGCACCCAAACTTTTTATGAAAGTCTTTGTTTATAATGATGAGGTTTATA
	Direct repeat	GTTAAAACTAGGAATAGTATAGTCATAAGTTAACACCATC
ADH4	5´sequence	ATGTCTTCCGTTACTGGGTTTTACATTCCACCAATCTCTTTCTTTGGTGAAGGTG
	3´ sequence	TCGAACGAACTCATAAACGTCAATTATGCGTGTGCCTTATTTATTTAGTTGTGCG
	Direct repeat	CAAGTTTACATTTGCAACAACTAATAGTCAAATAAGAAAA
ADH5	5´sequence	ATGCCTTCGCAAGTCATTCCTGAAAAACAAAAGGCTATTGTCTTTTATGAGACAG
	3´ sequence	TGTAACGAATTTGATGAATATATTTTTACTTTTTATATAAGCTATTTTGTAGATA
	Direct repeat	AGAAAATTATTTAACTACATATCTACAAAATCAAAGCATC
SFA1	5´sequence	ATGTCCGCCGCTACTGTTGGTAAACCTATTAAGTGCATTGCTGCTGTTGCGTATG
	3´ sequence	CTTAATTAAACTAAGTAAGCATGACTCAAATTTTCTGGAATACTTTGAAAATCAA
	Direct repeat	AATCTCCAAGTAAAGAAGGAATATAAGTAATATAAGTACA

Table S4. Transformation efficiencies of each of the ADH gene deletion cassettes.including expected PCR fragment sizes and observed integration efficiencies (12 discrete colonies) for each of the ADH genes.

Deletion cassette integration	Transformants μg^{-1} DNA	Expected PCR fragment size (bp)	Integration efficiency (%)
Δadh1::*amdSYM*	2 / 37 ◆	2963	83
Δadh2::*amdSYM*	29	3130	100
Δadh3::*amdSYM*	30	2986	83
Δadh4::*amdSYM*	21	2781	92
Δadh5::*amdSYM*	48	3124	83
Δsfa1::*amdSYM*	72	3338	83

◆ change in transformation efficiency after extending the homologous recombination sequences of the *ADH1* deletion cassette

Table S5. Primers for colony PCR assessment of gene deletions and marker removal within the genome of *Saccharomyces cerevisiae* CEN.PK113-7D.

Gene	Name	Sequence (5´ → 3´)	Diagnostic colony PCR fragment size (bp)		
			Wild type gene	*amdSYM* integration	Scarless excision
ADH1	ADH1verFwd	CAGCACCAACAGATGTCGTTGTTCC	1509	2963	462
	ADH1verRev	CGACCTCATGCTATACCTGAGAAAGCAACC			
ADH2	ADH2verFwd	CGGGAAACCATCCACTTCACGAGAC	1675	3130	628
	ADH2verRev	GAGACGATTCAGAGGAGCAGGACAAAC			
ADH3	ADH3verFwd	CGTTTCTGCGTCCGTACACTGTCC	1609	2986	482
	ADH3verRev	GTTTGGGCGGCTCGATGCTTG			
ADH4	ADH4verFwd	TTGCTGCCTCAAATATCTCACAC	1426	2781	277
	ADH4verRev	GTGCATTATACTGTACGCACAAC			
ADH5	ADH5verFwd	CTGCTATCTGCTTGTAGAAGGGTACGCTAACAGAG	1678	3124	672
	ADH5verRev	CTATTTCAGTTTGTCTTACGCACGCAGTTG			
SFA1	SFA1verFwd	AGTGCCTCAGTCGAATGG	1997	3338	836
	SFA1verRev	GGCGATGCAAGTGAAACC			

SUPPORTING MATERIALS AND METHODS

Media and growth conditions

Single *E. coli* colonies were inoculated into 6.0 ml Lysogeny broth (LB) [5] with antibiotic selection, as required, and incubated at 37 °C with 220 rpm shaking. Colonies of *S. cerevisiae* were inoculated into 10 ml yeast extract, peptone and dextrose broth (YEPD) [6] with antibiotic selection, as required, and incubated at 30 °C with 200 rpm shaking. Ampicillin was added to 100 μg ml^{-1}, or G418 (Geneticin) to 200 μg ml^{-1} as indicated.

Shake flask cultures were set up at 10-20 % (v/v) of the total flask volume. Microbial cultures were grown in a 1.0 cm orbital shaking diameter incubator. For microbial storage, 1.0 ml of stationary phase culture was added to 250 μl of sterile glycerol, thoroughly mixed, snap frozen in liquid nitrogen and stored at -80 °C.

Synthetic Medium (SM) was prepared according to Solis-Escalante, et al. [4] and was supplemented with acetamide (SM Ac) or fluoroacetamide (SM FAc) as indicated.

The basal media composition for Synthetic Defined (SD) media was Yeast Nitrogen Base (YNB) without zinc prepared according to the manufacturer's instructions (Formedium, Hunstanton, UK). The SD media additionally contained 20 g l^{-1} glucose and 400 μg l^{-1} $ZnSO_4$. For ministat experimentation SD media was adapted as described in **Table 1**.

Generation of the combinatorial deletion library

Yeast transformation. Transformations of DNA plasmids and linearised deletion cassettes into yeast were performed using lithium acetate and single stranded carrier DNA (Schiestl and Gietz, 1989). The protocol was standardised with the use of the yeast transformation kit (Sigma-Aldrich, Dorset, UK), adapted as follows: Single yeast colonies were inoculated into 10 ml YEPD broth and incubated overnight at 30 °C with 200 rpm shaking. 100 ml YEPD broth was inoculated with overnight culture at OD600nm of 0.2 and incubated at 30 °C with 200 rpm shaking until an OD600nm of 1.0 was reached (typically after 3-4 hours). Cells were washed in an equal volume of sterile Tris-EDTA buffer solution, pelleted and re-suspended in 1.0 ml transformation buffer (Sigma-Aldrich, Dorset, UK). For each transformation, 100 μl yeast cell suspension was aliquoted and to this 10 μl denatured salmon sperm DNA (100 μg) and 1.0 μg of plasmid/linearised DNA (or an equal volume of sterile Tris-EDTA buffer solution for the negative control) was added. Ethanol was added to 10 % (v/v) and the transformation mix was incubated at room temperature for 5 min. 600 μl PLATE buffer (Sigma-Aldrich, Dorset, UK) was added followed by incubation for 1.0 hour at 30 °C with 200 rpm shaking. The cells were then heat shocked at 42 °C for 20 min, pelleted, re-suspended in 500 μl YEPD broth, incubated for 1.0 hour at 30 °C with 200 rpm shaking, washed in 500 μl of sterile Tris-EDTA buffer, pelleted and re-suspended in 100 μl of sterile Tris-EDTA buffer solution. 90 μl was spread on a selection media plate and 10 μl on a

non-selection control plate. Transformation plates were incubated for 2-4 days at 30 °C until discrete colonies were visible.

Biomass determination

The BD FACS Aria II flow cytometer (Becton Dickenson, San Jose, US) was used to determine cell count and viability. The blue 488 nm excitation laser was used for analysis of samples (diluted with 1:10 with sterile PBS) and data collection continued until 1,500 events of the CountBright™ absolute counting beads (ThermoFisher Scientific, Waltham, US) had been recorded. Size (forward scatter) and internal complexity (side scatter) detectors were used to identify single cells of S. cerevisiae. Propidium iodide was used to stain yeast suspensions according to Deere, et al. [7], in order to determine yeast culture viability.

Chromatographic methods

High performance liquid chromatography. Ethanol was determined by high performance liquid chromatography with a refractive index detector (HPLC RI) 200 µl of supernatant from cell samples was combined with 10 µl of cellobiose [85.5 g l^{-1}], used as an internal standard for quantification. Analyses were performed using an Agilent Technologies (Santa Clara, US) 1200 series liquid chromatography system equipped with a 1260 Infinity refractive index detector. 60 µl of each sample and standard were injected on a Rezex™ RHMMonosaccharide H+ (8 %) column, 8.0 µm, 300 × 7.8 mm (Phenomenex, Torrance, US), at 50 °C. The refractive index detector was at 40 °C; chromatographic separation was obtained at a flow rate of 0.6 ml min^{-1} using a 0.005 % (v/v) H_2SO_4 isocratic mobile phase. A 5-point calibration curve and retention time standard including cellobiose, glucose, glycerol and ethanol were performed for each analysis. Data were analysed with OpenLAB CDS ChemStation edition for LC & LC/MS systems (Agilent Technologies, Santa Clara, US).

Gas chromatography. Sample preparation included 1.0 ml of yeast cell suspension was disrupted using lysing matrix C tubes and the FastPrep®-24 instrument (MP Biomedicals, California, US), at a speed setting of 6.0 m s^{-1} for 1.0 min and for 8 passes. 100 µl of homogenised sample was combined with 1.5 µl of heptadecanoic acid [1.70 g l^{-1}], used as an internal standard for quantification of free fatty acids. The free fatty acid analytical method performed on the prepared samples was according to Runguphan and Keasling [8] with the following adaptations: (1) 100 µl of ethyl acetate (EA)/iodomethane (Mel) was added to the samples and standards; (2) autonomous sample/standard preparation and injection was performed using a Gerstel (Mülheim an der Ruhr, Germany) multipurpose sampler attached to the GC with flame ionisation detector (FID). Analyses were performed using an Agilent Technologies (Santa Clara, US) 7890B GC system equipped with a flame ionisation detector (maintained at 300 °C, hydrogen flow of 30 ml min^{-1}, air flow of 400 ml min^{-1} and a data acquisition rate of 50 Hz). 2.0 µl of each sample and standard were injected into a non-deactivated, baffled glass liner with a 10:1 split ratio (12 ml min^{-1} split flow) and the inlet temperature was maintained at 250 °C with a 3.0 ml min^{-1} septum

purge flow. A Zebron semi-volatiles (Phenomenex, Torrance, US) column, 30 m × 250 µm × 0.25 µm, coupled to a 10 m guard column was maintained at a constant gas flow rate of 1.2 ml min^{-1}. The temperature gradient of the GC oven was initially held for 2.45 minutes at 40 °C. It was ramped at a rate of 24.52 °C min^{-1} until 310 °C was achieved and held for 4.08 minutes. A 6-point calibration curve and retention time standard including myristic acid (C14:0), pentadecanoic acid (C15:0), palmitic acid (C16:0), heptadecanoic acid (C17:0), stearic acid (C18:0) and oleic acid (C18:1) were performed for each analysis. Data were analysed with OpenLAB CDS ChemStation edition for GC systems (Agilent Technologies, Santa Clara, US).

(HPLC RI) 200 µl of supernatant from cell samples was combined with 10 µl of cellobiose [85.5 g l^{-1}], used as an internal standard for quantification. Analyses were performed using an Agilent Technologies (Santa Clara, US) 1200 series liquid chromatography system equipped with a 1260 Infinity refractive index detector. 60 µl of each sample and standard were injected on a Rezex™ RHMMonosaccharide H+ (8%) column, 8.0 µm, 300 × 7.8 mm (Phenomenex, Torrance, US), at 50 °C. The refractive index detector was at 40 °C; chromatographic separation was obtained at a flow rate of 0.6 ml min^{-1} using a 0.005 % (v/v) H_2SO_4 isocratic mobile phase. A 5-point calibration curve and retention time standard including cellobiose, glucose, glycerol and ethanol were performed for each analysis. Data were analysed with OpenLAB CDS ChemStation edition for LC & LC/MS systems (Agilent Technologies, Santa Clara, US).

SUPPORTING INFORMATION REFERENCES

1. van Dijken, J. P., Weusthuis, R. A., and Pronk, J. T., (1993) Kinetics of growth and sugar consumption in yeasts. *Antonie Van Leeuwenhoek 63*, 343-52.
2. Dickinson, J. R. and Schweizer, M., *The Metabolism and Molecular Physiology of* Saccharmoyces cerevisiae. Second edition ed.; CRC Press: UK, 2004.
3. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., et al., (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res. 36*, W465-9.
4. Solis-Escalante, D., Kuijpers, N. G., Bongaerts, N., Bolat, I., Bosman, L., Pronk, J. T., Daran, J. M., and Daran-Lapujade, P., (2013) *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res 13*, 126-39.
5. Bertani, G., (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol 62*, 293-300.
6. Green, M. R., Sambrook, J., and Sambrook, J., *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2012.
7. Deere, D., Shen, J., Vesey, G., Bell, P., Bissinger, P., and Veal, D., (1998) Flow cytometry and cell sorting for yeast viability assessment and cell selection. *Yeast (Chichester, England) 14*, 147-160.
8. Runguphan, W., and Keasling, J. D., (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab Eng 21*, 103-13.