PNAS
www.pnas.org

Supplementary Information for
Adaptations in metabolism and protein translation give rise to the Crabtree effect in yeast

Carl Malina[a,b], Rosemary Yu[a,c], Johan Björkeroth[a], Eduard J Kerkhoven[a,c], Jens Nielsen[a,b,c,d,e*]

Affiliations
[a]Department of Biology and Biological Engineering, Chalmers University of Technology, SE-412 96, Gothenburg, Sweden

[b]Wallenberg Center for Protein Research, Chalmers University of Technology, SE-412 96, Gothenburg, Sweden

[c]Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, SE-412 96, Gothenburg, Sweden

[d]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kgs Lyngby, Denmark

[e]BioInnovation Institute, Ole Måløes Vej 3, DK-2200, Copenhagen N, Denmark

*Correspondance to Jens Nielsen
Email: nielsenj@chalmers.se

This PDF file includes:

Supplementary Methods
Figures S1 to S13
Table S1
Legends for Datasets S1 to S4
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S4

Supplementary Methods

Sample preparation for proteomic analysis

Samples were suspended in 300 µl of the lysis buffer containing 2% sodium dodecyl sulfate (SDS) and 50 mM triethylammonium bicarbonate (TEAB), and homogenized using lysis matrix C (1.0 mm silica spheres) on a FastPrep®-24 instrument (MP Biomedicals, OH, USA) for 4 repeated 40 second cycles at 6.5 m/s. Lysed samples were centrifuged at 16,200xg for 10 min and the supernatants were transferred to clean tubes. The pellets were washed with 200 µl of the lysis buffer, centrifuged at 16,200xg for 10 min, and the supernatants were combined with the corresponding lysates from the previous step. Protein concentrations in the lysates were determined using Pierce™ BCA Protein Assay Kit (Thermo Fischer Scientific) and the Benchmark™ Plus microplate reader (Bio-Rad Laboratories, Hercules, CA, USA) with bovine serum albumin (BSA) solutions as standards. For each of the yeast species, aliquots containing equal protein amount from each replicate were mixed to yield the pooled reference samples.

For the tandem mass tag (TMT) relative quantification experiment, aliquots containing 50 µg of total protein from each sample were incubated at 37°C for 60 min in the lysis buffer with DL-dithiothreitol (DTT) at 100 mM final concentration. The reduced samples were processed via the modified filter-aided sample preparation (FASP) protocol (1). In short, the reduced samples were diluted to 1:4 by 8M urea solution, transferred onto Nanosep 30k Omega filters (Pall Corporation, Port Washington, NY, USA) and washed repeatedly with 8 M urea and once with digestion buffer containing 0.5% sodium deoxycholate (SDC) and 50 mM TEAB. Free cysteine residues were modified using 10 mM methyl methanethiosulfonate (MMTS) solution in digestion buffer for 20 min at room temperature and the filters were washed twice with 100 µl of digestion buffer. Pierce trypsin protease (MS Grade, Thermo Fisher Scientific) in digestion buffer was added at a ratio of 1:100 relative to total protein mass and the samples were incubated at 37°C overnight, then an additional 1:100 portion of trypsin was added and incubated for 2 more hours. Peptides were collected by centrifugation and labeled using Tandem Mass Tag (TMT 10plex) reagents (Thermo Fischer Scientific) according to the manufacturer's instructions. For each of the yeast specie, the labeled samples were combined into one pooled sample, concentrated using vacuum centrifugation, and SDC was removed by acidification with 10% TFA and subsequent centrifugation. The labeled pooled samples were treated with HiPPR detergent removal kit (Thermo Fisher Scientific) according to the manufacturer's protocol.

Pooled TMT-labeled samples were fractionated into 40 primary fractions via basic-pH reversed-phase chromatography (bRP-LC) on an XBridge BEH C18 column (3.5 µm, 3.0x150 mm, Waters Corporation) using Dionex Ultimate 3000 UPLC system (Thermo Fischer Scientific), solvent A was 10 mM ammonium formate buffer at pH 10.00 and solvent B was 90% acetonitrile, 10% solvent A. Forty primary fractions were collected linear gradient from 3% to 40% solvent B over 18 min followed by an increase to 100% B over 5 min and hold at 100% B for 5 min; the primary fractions were concatenated into 20 final fractions (1+21, 2+22 etc), dried and reconstituted in 15µl of 3% acetonitrile, 0.2% formic acid for LC-MS analysis.

For the label-free quantification, aliquots containing 25 µg of total protein from each pooled reference sample were mixed with 5.3 µg of the UPS2 Proteomics Dynamic Range Standard (Sigma-Aldrich, Saint-Louis, MO), and the samples were digested with trypsin as described above. Digested peptide samples were treated with HiPPR detergent removal kit (Thermo Fisher

Scientific) according to the manufacturer's protocol. Each sample was subsequently separated into 8 fractions on Pierce High pH Reversed-Phase spin column kit (Thermo Fischer Scientific) using stepwise elution with 0.1% aqueous trimethylamine solution containing 7.5, 10.0, 12.25, 14.5, 16.75, 19.0, 21.25 and 50.0% of acetonitrile. The fractions were dried and reconstituted in 15µl of 3% acetonitrile, 0.2% formic acid for LC-MS analysis.

LC-MS/MS Analysis

The TMT-labeled samples (**workflow 1**)) were analyzed on an Orbitrap Fusion mass spectrometer and the label-free samples (**workflow 2**) were analyzed on an Orbitrap Fusion or an Orbitrap Fusion Lumos mass spectrometer interfaced with Easy-nLC 1200 liquid chromatography systems (all from Thermo Fisher Scientific).

Peptides were trapped on an Acclaim Pepmap 100 C18 trap column (100 µm x 2 cm, particle size 5 µm, Thermo Fischer Scientific) and separated on an analytical column (75 µm x 35 cm, packed in-house with Reprosil-Pur C18, particle size 3 µm, Dr. Maisch, Ammerbuch, Germany) at 300 nL/min flow rate using 0.2% formic acid in water as solvent A and 80% acetonitrile and 0.2% formic acid as solvent B.

In workflow 1, peptides were eluted using the linear gradient from 5% to 35% B over 75 min followed by an increase to 100% B for 5 min, and 100% B for 10 min. Full MS scans were performed at 120,000 resolution in the m/z range 375-1375. The most abundant doubly or multiply charged precursors from the MS1 scans were isolated using the quadrupole with 0.7 m/z isolation window with a "top speed" duty cycle of 3 s and dynamic exclusion within 10 ppm for 45 seconds, fragmented by collision induced dissociation (CID) at 35% collision energy with the maximum injection time of 50 ms, and the MS2 spectra were detected in the ion trap. Top 10 MS2 fragment ions within the m/z range 400-1200 were simultaneously isolated, fragmented by higher-energy collision dissociation (HCD) at 65% collision energy, and the MS3 spectra were recorded in the Orbitrap at 30,000 resolution, m/z range 100-500 and maximum injection time 75 ms.

In workflow 2, each label-free fraction was injected in triplicate. Peptides were eluted using the linear gradient from 5% to 35% B over 105 min, 35% to 100% B in 5 min, followed by a hold at 100% B for 10 min. Full MS precursor ion scans were performed at 120,000 resolution in the *m/z* range 375-1375, the most abundant doubly or multiply charged precursors were isolated within 0.7 m/z window using the quadrupole with a "top speed" duty cycle of 3 s and dynamic exclusion within 10 ppm for 45 seconds. Precursors were fragmented using CID at 35% collision energy and the MS2 spectra were recorded in the ion trap with "Rapid" scan rate, "Auto" scan range and the maximum injection time of 50 ms.

Proteomic Data Analysis

Proteome Discoverer version 2.4 (Thermo Fisher Scientific) was used for protein identification and quantification. Reference databases for *Kluyveromyces marxianus* (strain DMKU3-1042 / BCC 29191 / NBRC 104275), *Saccharomyces cerevisiae* ATCC 204508 / S288c, *Scheffersomyces (Pichia) stipitis* (strain ATCC 58785 / CBS 6054 / NBRC 10063 / NRRL Y-11545) and *Schizosaccharomyces pombe* (strain 972 / ATCC 24843) were downloaded from Uniprot (November 2019) and supplemented with the common proteomic contaminant sequences; in the label-free workflow 2, the databases were additionally supplemented with 48 UPS protein sequences.

The following parameters were common between the workflows: Mascot version 2.5.1 (Matrix Science, London, UK) was used as a database search engine; 5 ppm precursor tolerance and 0.6 Da fragment tolerance; trypsin with 1 allowed missed cleavage as an enzyme rule; oxidation on methionine as a variable modification; methylthiolation on cysteine as a fixed modification; Percolator for PSM validation with the strict FDR threshold of 1%.

Additional parameters for the TMT workflow 1: TMT as a fixed modification on peptide N-termini and lysine; TMT 10plex Lot. No UF282325 as a quantification method with the reporters identified within 40 ppm in the MS3 spectra; quantification based on unique peptides and reporter S/N, with the correction for isotopic impurities, the average reporter S/N threshold of 10 and SPS match threshold of 0%; reporter normalization on total peptide amount.

Additional parameters for the label-free workflow 2: precursor mass recalibration enabled with the initial precursor match tolerance of 20 ppm; Minora Feature detection was used for precursor ion quantification; intensities at peak maxima for all unique and shared peptides were used for protein quantification; chromatographic retention time (RT) alignment was enabled with the max RT shift of 10 min; feature linking and mapping was enabled with the minimal S/N of 5. Protein intensities from 3 injection replicates were averaged. Modified intensity-based absolute quantification (IBAQ) approach (2) was used to estimate the absolute protein amounts in the pooled reference sample: protein intensity was divided by the number of identified peptides to yield the IBAQ intensity for each protein; linear regression was calculated for the relationships between the known log10-transformed protein amount and measured log10-transformed IBAQ abundance on the spiked UPS2 standard proteins; the resulting regression model and measured IBAQ abundances of yeast proteins were used to estimate their protein amounts.

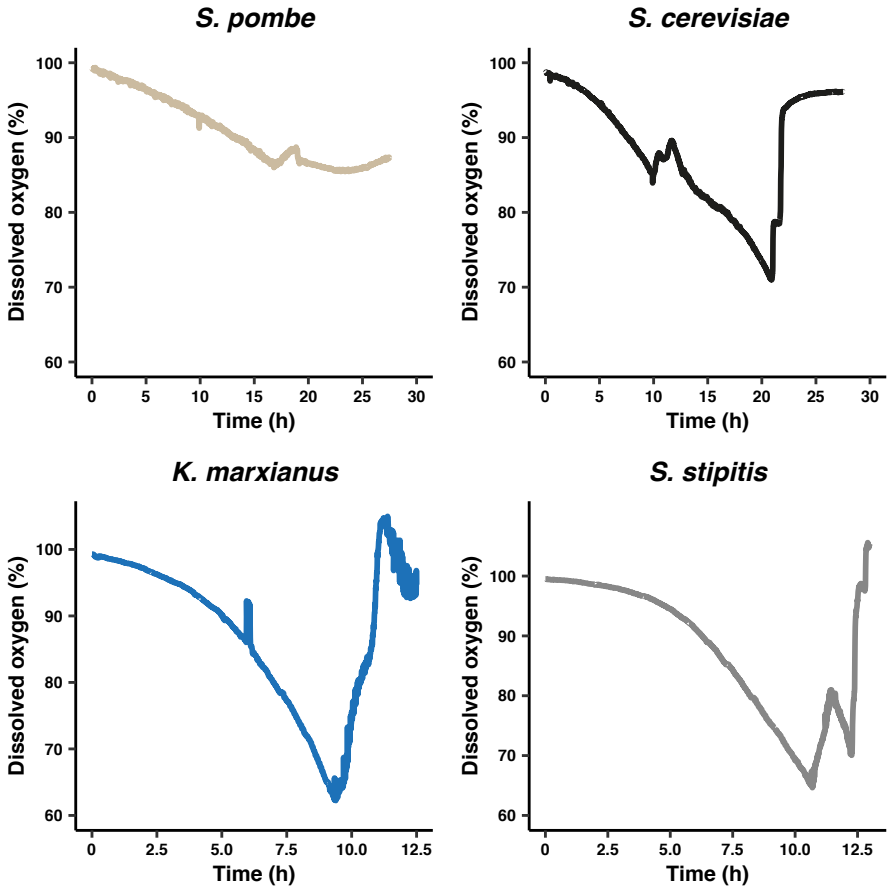


Fig. S1: % Dissolved oxygen (DO) during the batch cultivations of the four yeasts. Mean values of biological triplicates are shown.

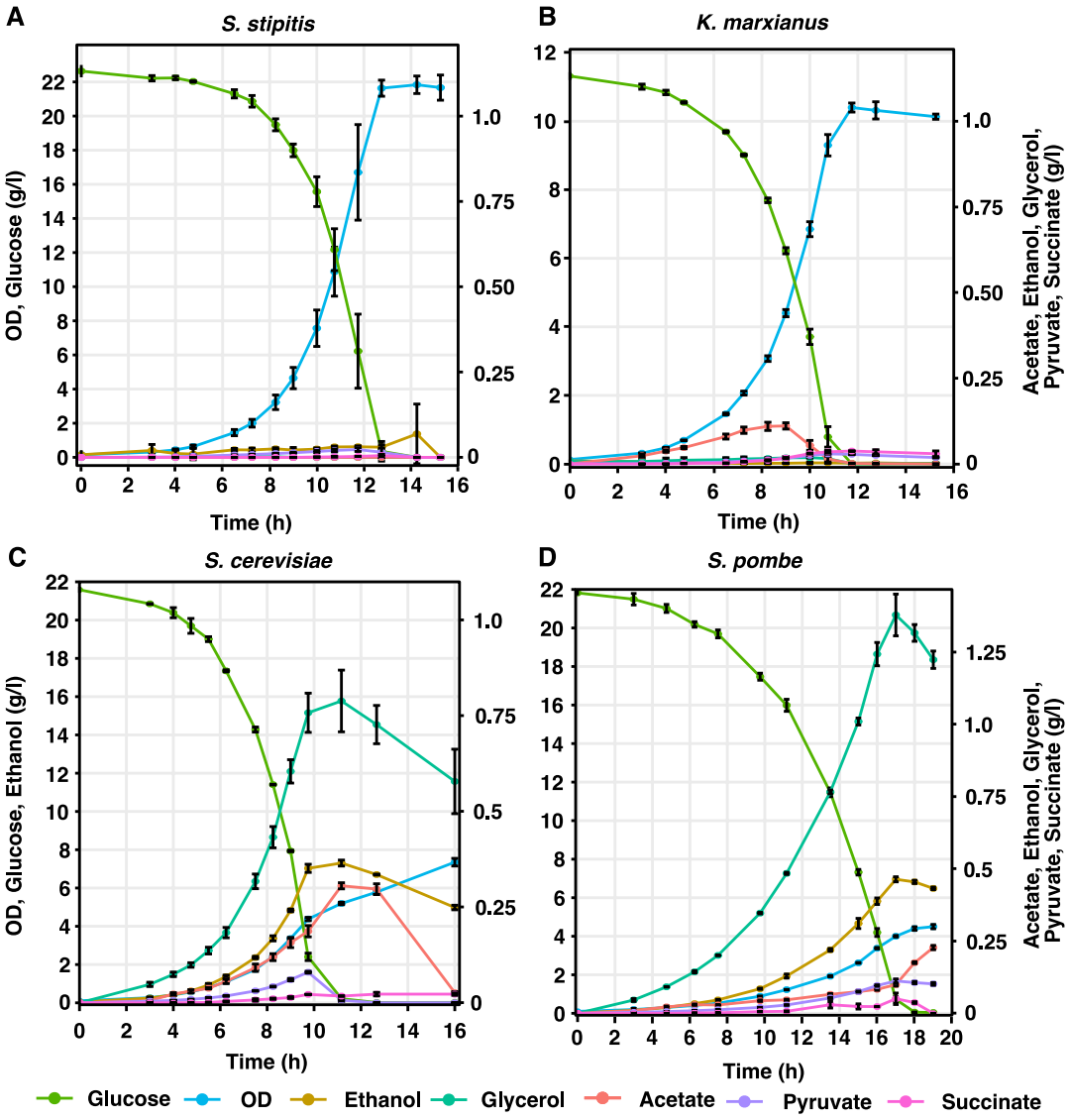


Fig. S2. Cell density (OD_{600}) and extracellular concentrations of glucose, ethanol, glycerol, acetate, pyruvate and succinate during the course of the batch cultivation for **A** *S. stipitis*, **B** *K. marxianus*, **C** *S. cerevisiae*, **D** *S. pombe*. Mean values ± standard deviation (SD) of biological triplicates are shown.

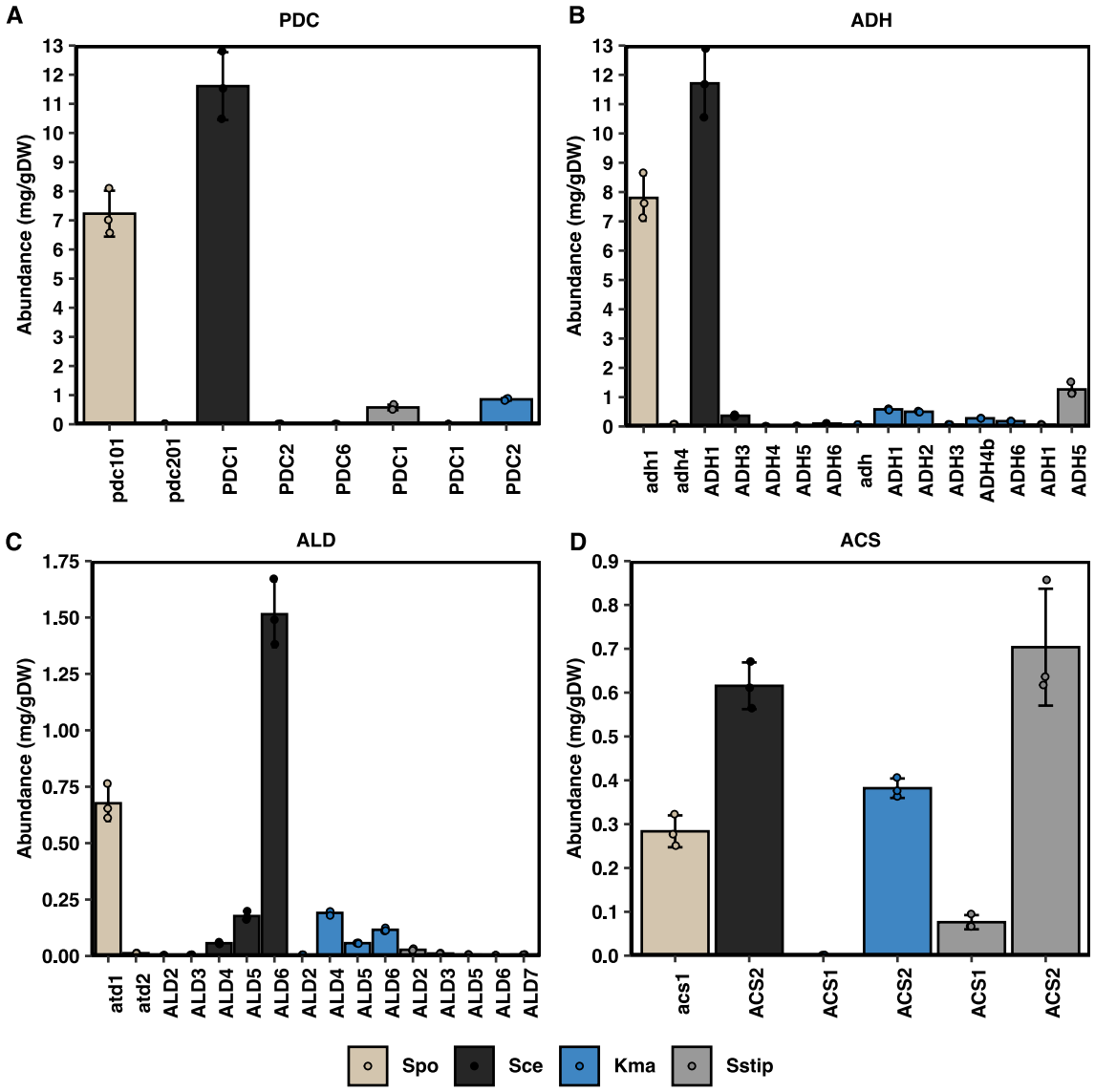


Fig. S3. Absolute abundance of proteins around the pyruvate branch point. **A** Pyruvate decarboxylase (PDC), **B** alcohol dehydrogenase (ADH), **C** aldehyde dehydrogenase (ALD), **D** acetyl-CoA synthetase. Bars represent mean values ± SD of biological triplicates. Spo, *S. pombe*; Sce, *S. cerevisiae*; Kma, *K. marxianus*; Sstip, *S. stipitis*.

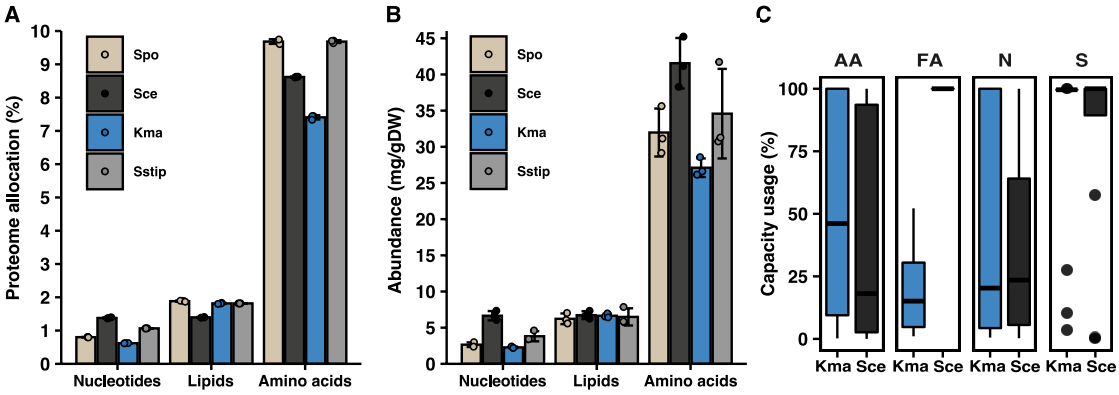


Fig. S4. Proteomics data for biosynthetic processes. A Proteome allocation of selected biosynthetic processes. B As in a, but with summed absolute abundances of the processes. Data in a and b are presented as mean ± SD of biological triplicates. C Capacity usage, calculated as the model-predicted enzyme levels divided by the experimentally measured enzyme levels, of selected biosynthetic processes. Spo, *S. pombe*; Sce, *S. cerevisiae*; Kma, *K. marxianus*; Sstip, *S. stipitis*; AA, amino acid biosynthesis; FA, fatty acid biosynthesis; N, nucleotide biosynthesis; S, sterol biosynthesis.

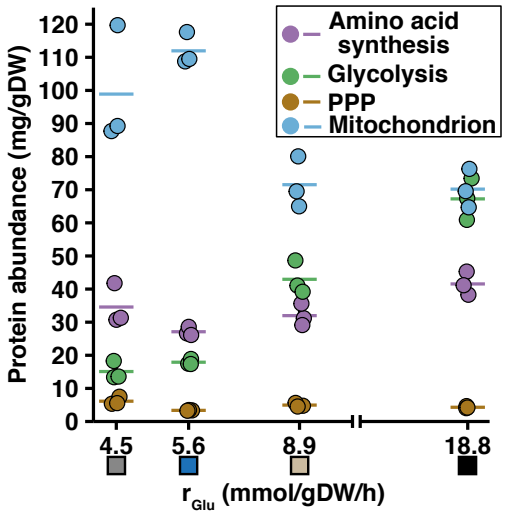


Fig. S5. Summed absolute protein abundance of selected processes plotted against the glucose uptake rate. Colored boxes represent the organisms, as indicated in Supplementary Figure 3. Mean values ± SD of biological triplicates are shown.

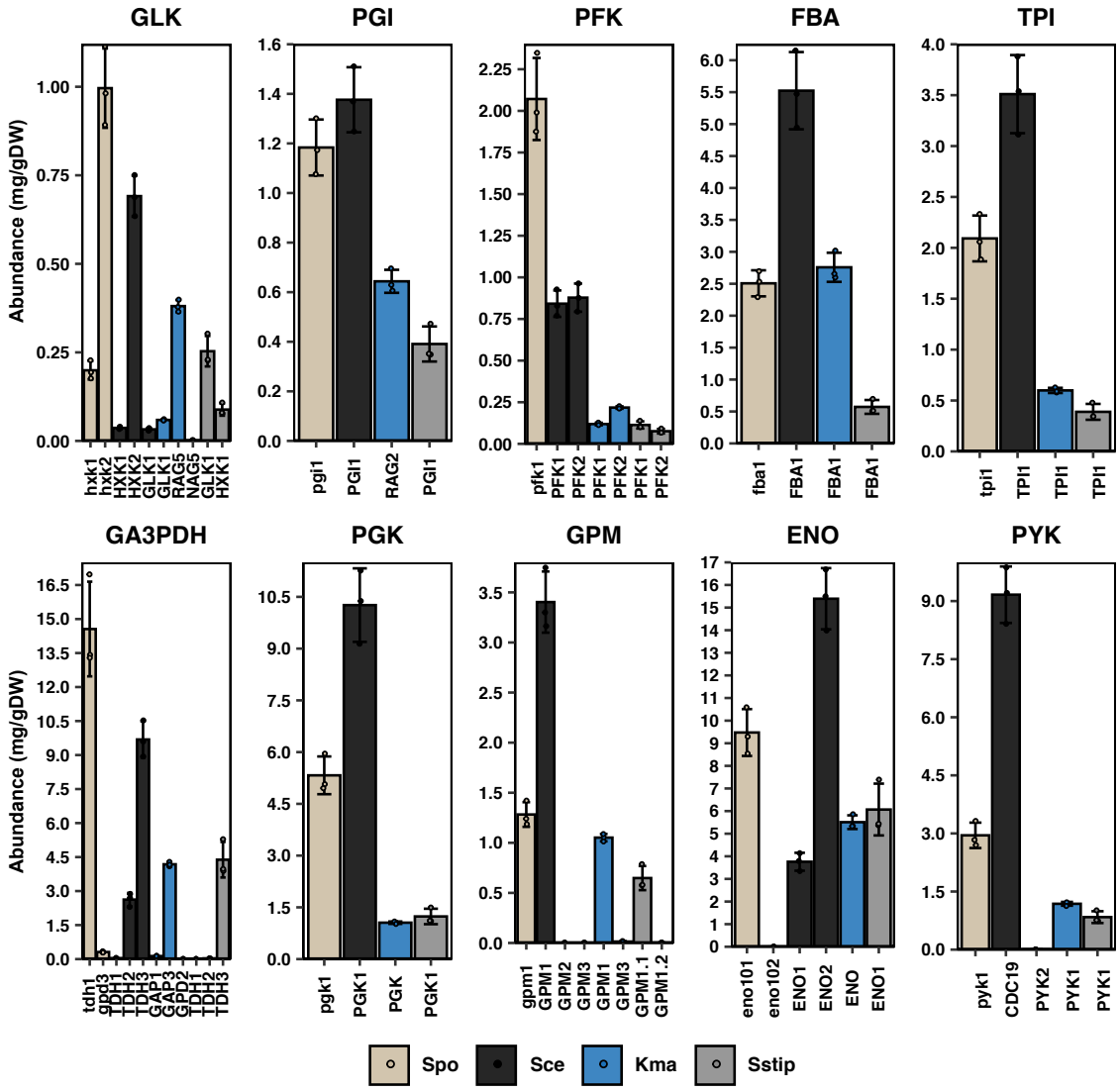


Fig. S6: Absolute abundance of glycolytic proteins. Mean values of biological triplicates are shown. Spo, *S. pombe*; Sce, *S. cerevisiae*; Kma, *K. marxianus*; Sstip, *S. stipitis*; GLK, glucose phosphorylation (including hexokinase and glucokinase); PGI, phosphoglucose isomerase; PFK, phosphofructokinase; FBA, fructose bisphosphate aldolase; TPI, triose phosphate isomerase; GA3PDH, glyceraldehyde-3-phosphate dehydrogenase; PGK, 3-phosphoglycerate kinase; GPM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase.

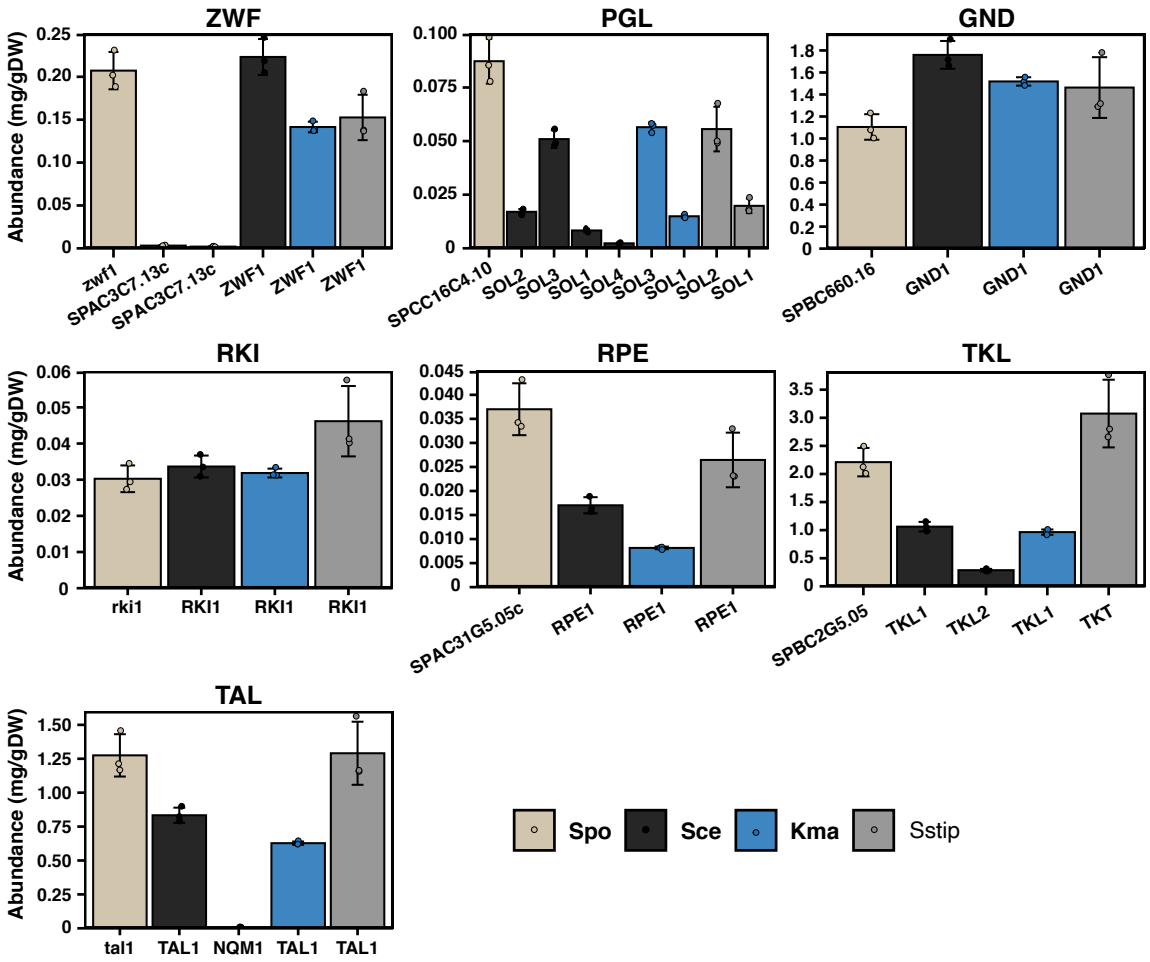


Fig S7. Absolute abundance of proteins of the pentose phosphate pathway. Mean values ± SD of biological triplicates are shown. Spo, *S. pombe*; Sce, *S. cerevisiae*; Kma, *K. marxianus*; Sstip, *S. stipitis*; ZWF, glucose-6-phosphate dehydrogenase; PGL, 6-phosphogluconolactonase; GND, 6-phosphogluconate dehydrogenase; RKI, ribose-5-phosphate ketol-isomerase; RPE, D-ribulose-5-phosphate 3-epimerase; TKL, transketolase; TAL, transaldolase.

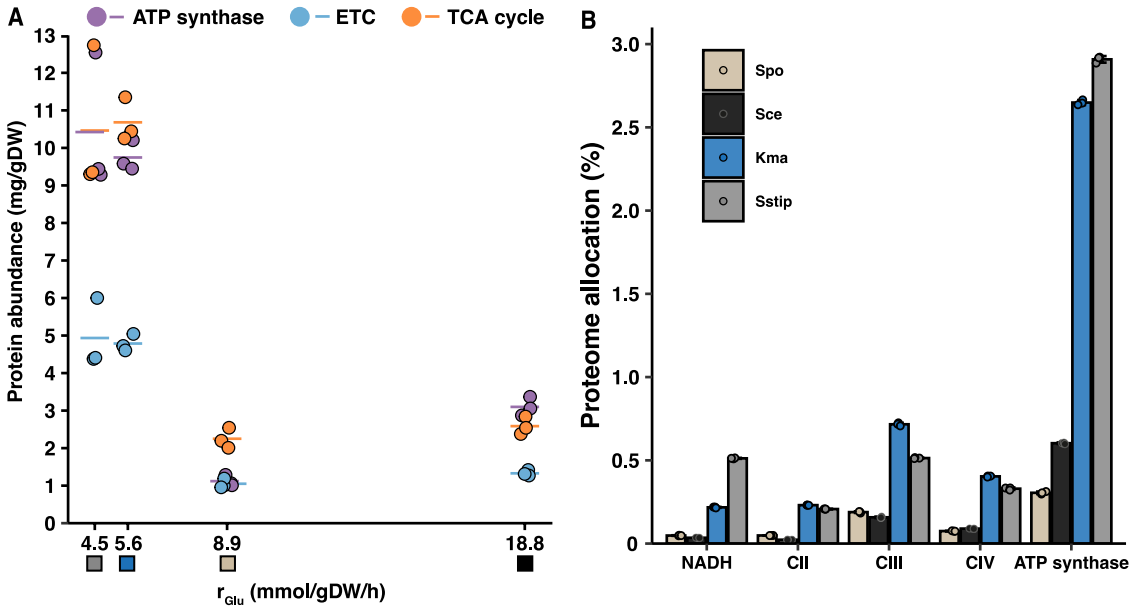


Fig. S8. **A** Summed absolute protein abundance of ATP synthase, electron transport chain (ETC) complexes and TCA cycle plotted against glucose uptake rate. Mean values of biological triplicates are shown. Colored boxes represent the organisms, as specified in panel b. **B** Proteome allocation, in % of total cellular protein, of the components of the ETC and ATP synthase. Mean values ± SD of biological triplicates are shown. Spo, *S. pombe*; Sce, *S. cerevisiae*; Kma, *K. marxianus*; Sstip, *S. stipitis*.

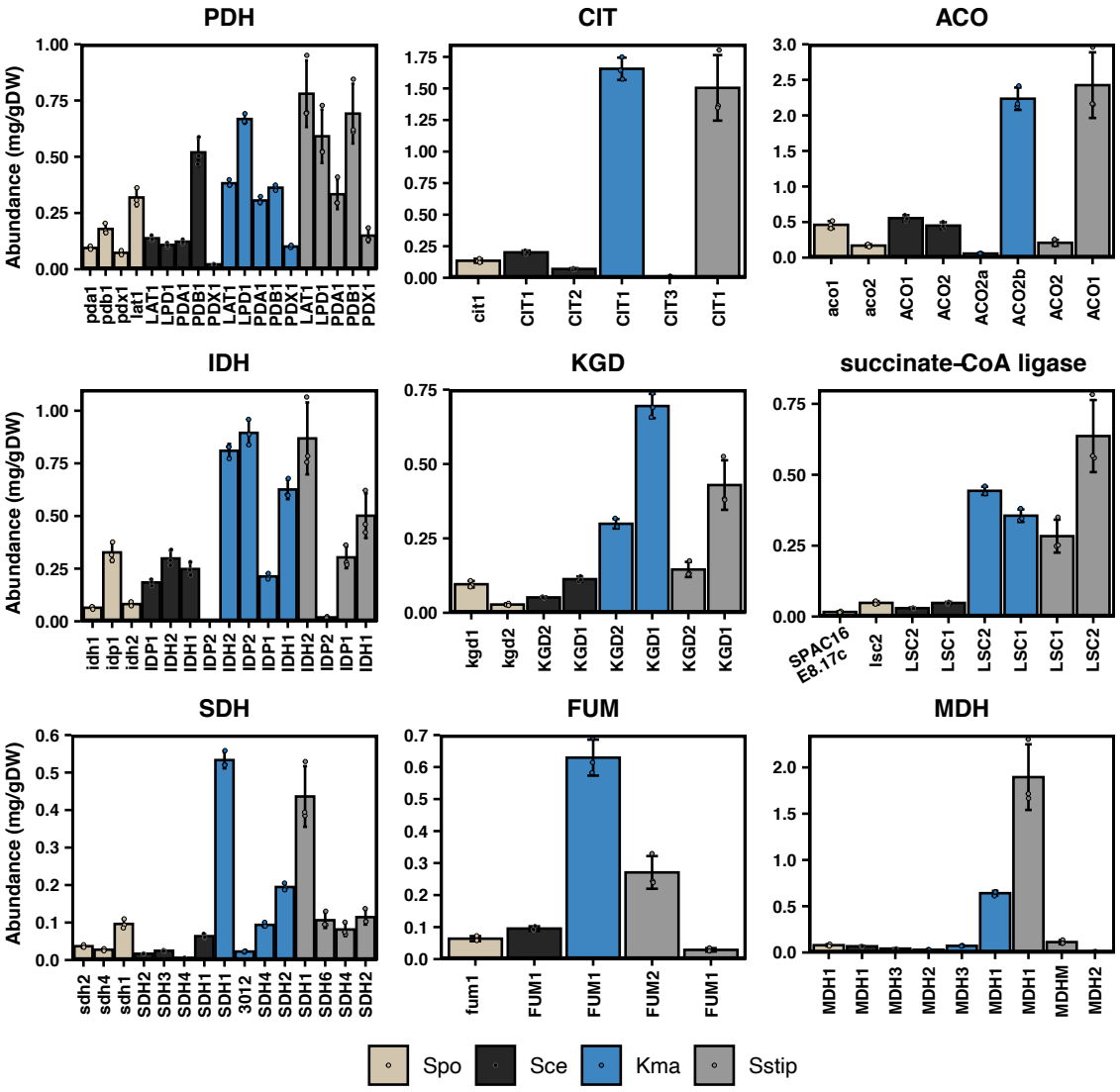


Fig. S9. Absolute abundance of proteins of the tricarboxylic acid (TCA) cycle. Mean values ± SD of biological triplicates are shown. Spo, *S. pombe*; Sce, *S. cerevisiae*; Kma, *K. marxianus*; Sstip, *S. stipitis*; PDH, pyruvate dehydrogenase; CIT, citrate synthase; ACO, aconitase; IDH, isocitrate dehydrogenase; KGD, 2-oxoglutarate dehydrogenase; SDH, succinate dehydrogenase; FUM, fumarase; MDH, malate dehydrogenase.

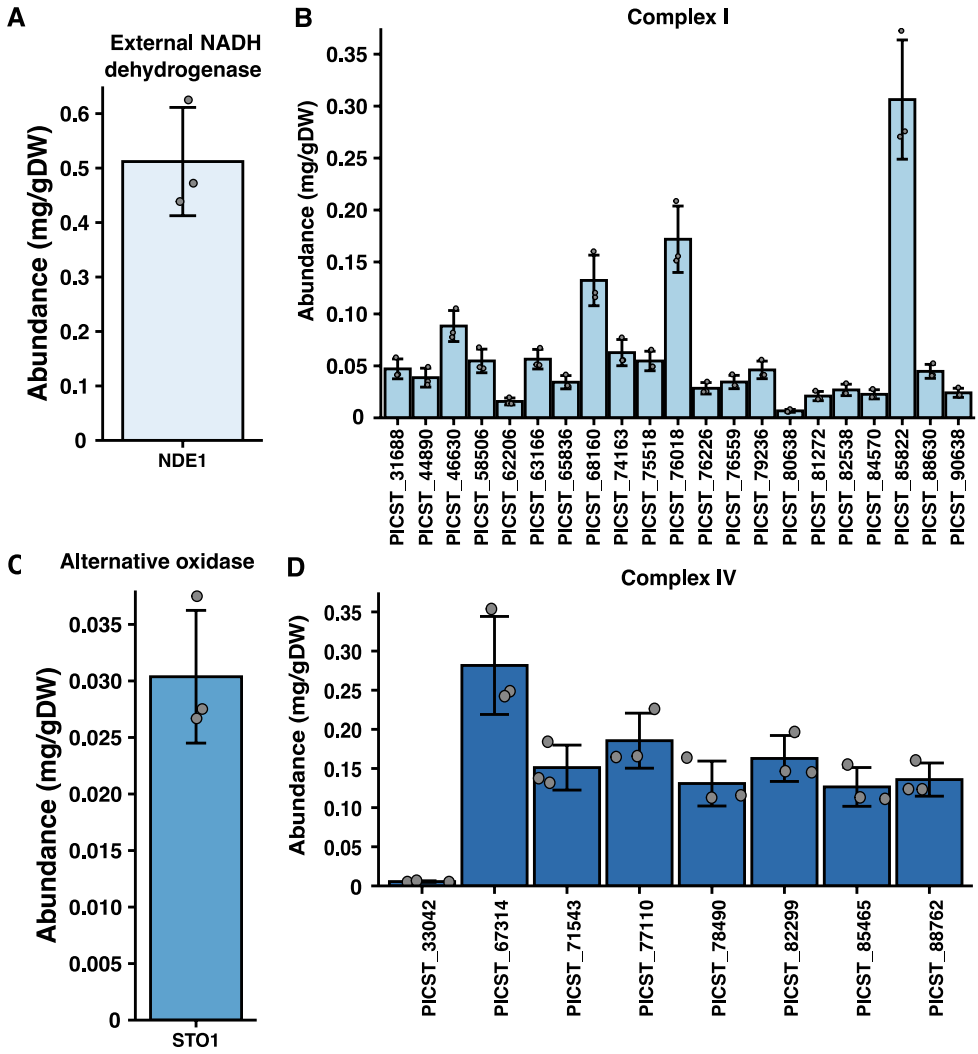


Fig. S10: NADH dehydrogenases and oxidases of the electron transport chain of *S. stipitis*. Absolute protein abundance of **A** of the External NADH dehydrogenase, NDE1, **B** the subunits of respiratory chain complex I, **C** alternative oxidase, STO1, and **D** the subunits of respiratory chain complex IV (cytochrome c oxidase). Mean values ± SD of biological triplicates are shown.

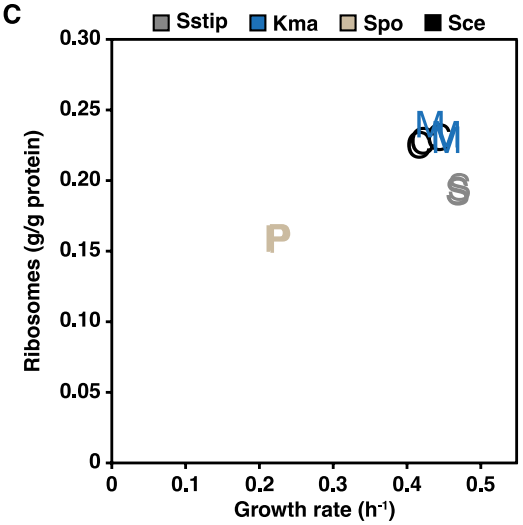
A

	Sce	Spo	Kma	Sstip
# RP (Uniprot)	137	143	78	86
# RP detected	101	110	76	84
% RP detected	74	77	97	98

B

	Sce	Spo	Kma	Sstip
# RP subunits (paralogs binned)	78	81	78	86
# RP subunits detected	77	78	76	84
% RP subunits detected	99	96	97	98

C



D

	Sce	Spo	Kma	Sstip
# MRPs (Uniprot)	77	71	66	73
# MRPs detected	75	64	65	72
% MRPs detected	97	90	98	99

Fig. S11. Overview of ribosomal proteins (RPs) in the four yeasts. **A** Number of cytosolic ribosomal proteins (RPs), as well as the number of RPs detected in this study. **B** Ribosomal subunits, binning paralogs, as well as detected subunits. **C** Absolute abundance of RPs, expressed as the fraction of the total cellular protein. **D** Number of mitochondrial ribosomal proteins (MRPs) and the number of MRPs detected.

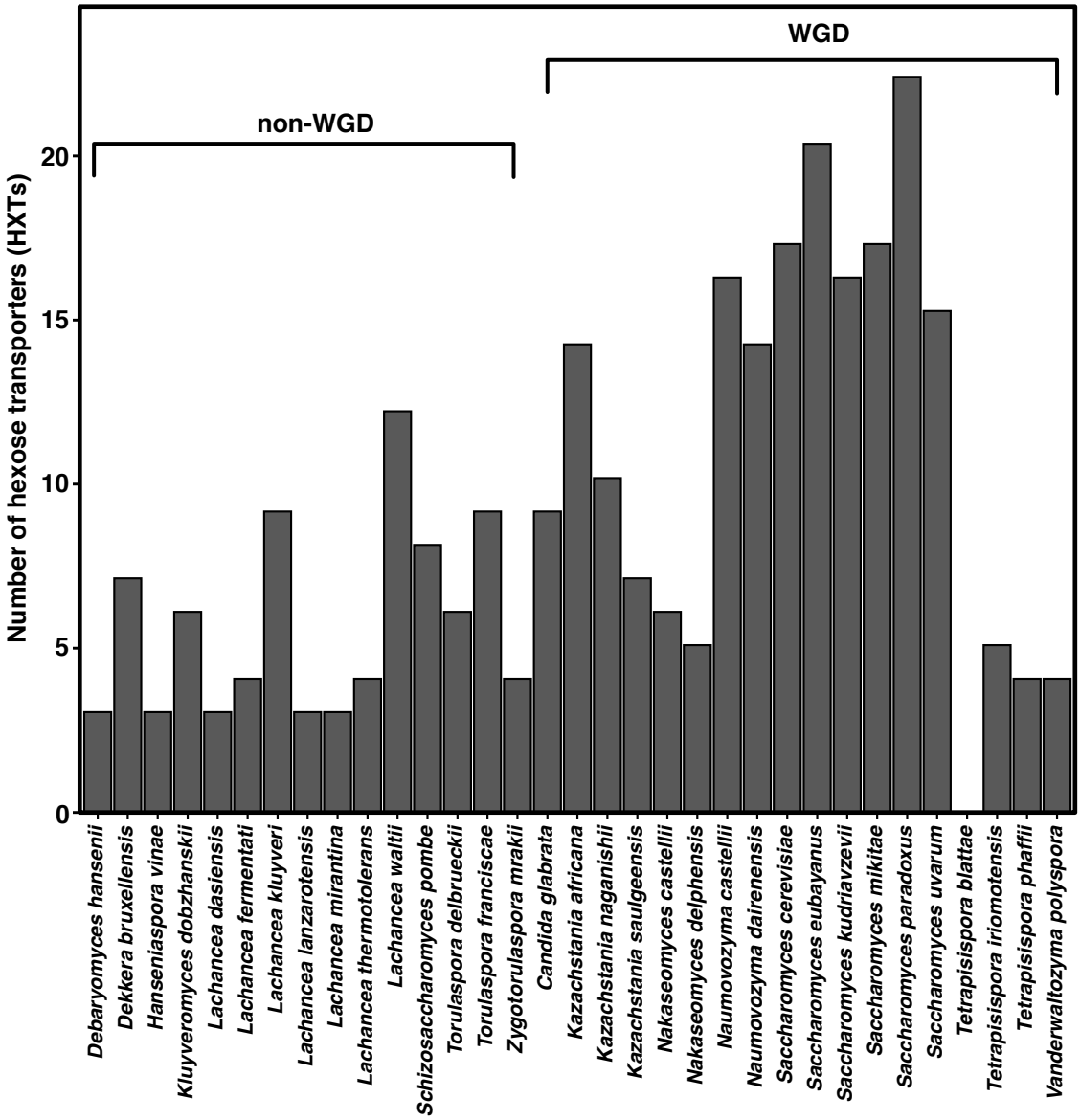


Fig. S12. Comparison of the number of hexose transporters in Crabtree-positive yeasts from non-WGD and WGD genera. WGD, whole genome duplication.

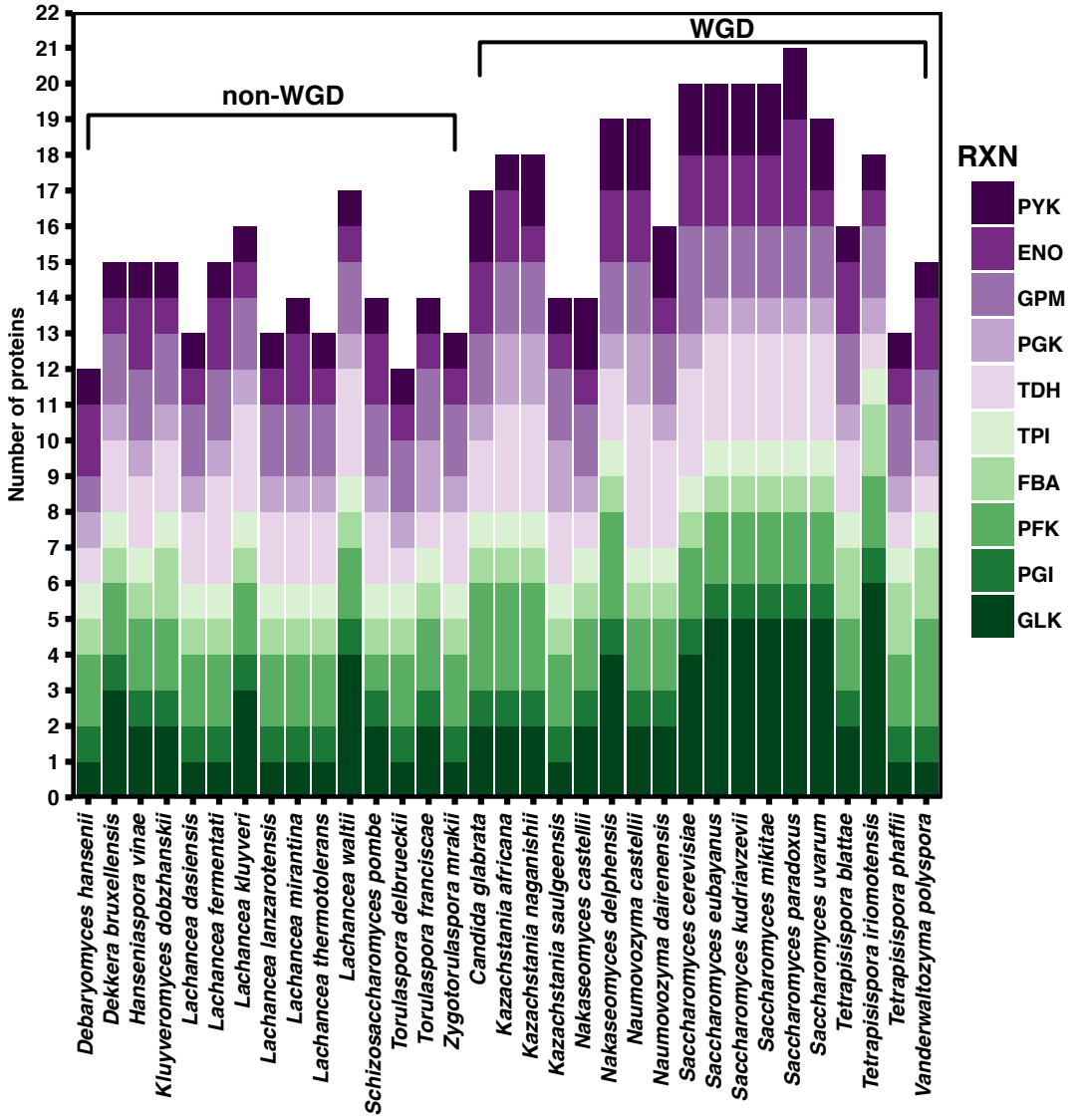


Fig. S13. Comparison of the number of glycolytic protein paralogs in Crabtree-positive yeasts from non-WGD and WGD genera. WGD, whole-genome duplication; GLK, glucose phosphorylation (including hexokinase and glucokinase); PGI, phosphoglucose isomerase; PFK, phosphofructokinase; FBA, fructose bisphosphate aldolase; TPI, triose phosphate isomerase; GA3PDH, glyceraldehyde-3-phosphate dehydrogenase; PGK, 3-phosphoglycerate kinase; GPM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase.

Table S1. Physiological parameters of the four yeasts. Data shown are mean values ± standard deviation of biological triplicates. n.d. not detected

Parameter	***S. cerevisiae***	***S. pombe***	***K. marxianus***	***S. stipitis***
μ (h^{-1})	0.424 ± 0.041	0.216 ± 0.01	0.441 ± 0.012	0.466 ± 0.002
Y_{sx} (g DW g^{-1} glucose)	0.125 ± 0.002	0.135 ± 0.005	0.441 ± 0.017	0.575 ± 0.015
r_{Glu} (mmol g_{DW}^{-1} h^{-1})	-18.79 ± 2.011	-8.945 ± 0.762	-5.558 ± 0.119	-4.500 ± 0.100
r_{Eth} (mmol g_{DW}^{-1} h^{-1})	26.97 ± 3.67	11.185 ± 0.493	n.d.	n.d.
r_{Gly} (mmol g_{DW}^{-1} h^{-1})	1.469 ± 0.093	1.185 ± 0.0995	n.d.	n.d.
r_{Ace} (mmol g_{DW}^{-1} h^{-1})	0.566 ± 0.082	0.106 ± 0.0160	0.328 ± 0.118	n.d.
r_{Pyr} (mmol g_{DW}^{-1} h^{-1})	0.163 ± 0.020	0.0942 ± 0.0084	0.034 ± 0.002	0.015 ± 0.005
r_{Suc} (mmol g_{DW}^{-1} h^{-1})	0.0365 ± 0.002	0.0312 ± 0.0079	0.0327 ± 0.004	n.d.

Dataset S1 (data provided in a separate file). Absolute protein abundances for *Saccharomyces cerevisiae.*

Dataset S2 (data provided in a separate file). Absolute protein abundances for *Schizosaccharomyces pombe.*

Dataset S3 (data provided in a separate file). Absolute protein abundances for *Kluyveromyces marxianus.*

Dataset S4 (data provided in a separate file). Absolute protein abundances for *Scheffersomyces stipitis.*

SI References

1. J. R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nature Methods* **6**, 359–362 (2009).
2. B. Schwanhäusser, Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).