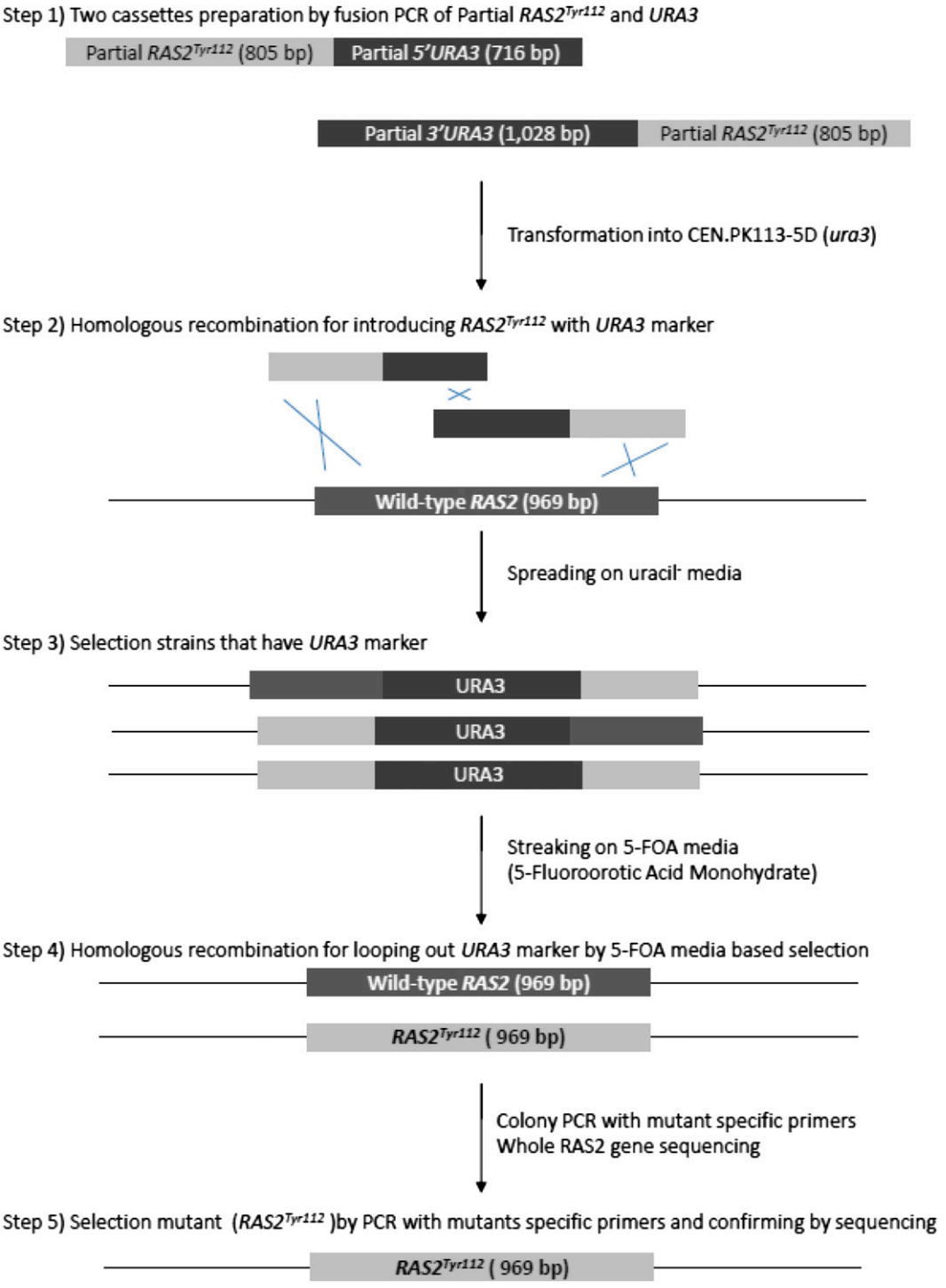
Supporting Information

Hong et al. 10.1073/pnas.1103219108

SI Materials and Methods

Site-Directed Mutant Generation and Measurements of a Maximum Growth Rate.

i) Steps for site-directed mutagenesis



ii) Primer lists for construction of site-directed mutants

For fusion of *RAS2* and *URA3*	
RAS2-5P1	CCACTCTTTATCTGACTCTTCTGC
RAS2-3P1	GGAATTGCCATGAAGCCGAA
	TTCACATTTTTACCGTTGGCAGC
For fusion of *RAS2* and *URA3*	
RAS2-5P2	GGAATTCGATATCAAGCTTATCGAT
	CCACTCTTTATCTGACTCTTCTGC
RAS2-3P2	TTCACATTTTTACCGTTGGCAGC
For selection of mutants	
RAS2-4BW (wild-type)	AATTGGAACATAGTCGGTAT**C**
RAS2-4BM (mutant)	AATTGGAACATAGTCGGTAT**A**
For sequencing of whole *RAS2* genes	
RAS2-5P	GTTTTAGCCGTTGTCTTCTCTT
RAS2-3P	GTTCTTTTCGTCTTAGCGTTTC

iii) Characterization of maximum growth rate of site-directed mutants in a flask culture

Biological triplicates of all strains were performed in flask scale. Uracil (150 mg/L) was supplied to galactose (20 g/L) minimal media. *P* value was calculated by assumption of one-tailed distribution and homoscedasticity.

Batch Fermentation and Sampling. Aerobic batch cultures were carried out in 1.0-L DasGip stirrer-pro vessels (DasGip) containing 0.8-L working volume. The same media with preculture was used, and the temperature was kept at 30 °C. The airflow rate was set to 1 L per liquid volume per minute, and the pH of the media was maintained at 5.0 by automatic addition of 2 M KOH. All conditions of temperature, agitation, air supply, pH, and off-gas analysis were controlled and monitored by the DasGip system. The concentration of dissolved oxygen was measured with an autoclavable polarographic oxygen electrode (Mettler Toledo). The off-gas from the bioreactors was analyzed for real-time determination of oxygen and carbon dioxide concentration by DasGip fedbatch pro gas analysis systems with the off gas analyzer 1 GA4 based on zirconium dioxide and two-beam infrared sensor (DasGip). Samples for measuring biomass and extracellular metabolites were obtained hourly after starting the exponential growth phase. When the concentration of carbon dioxide reached ≈1.7%, which was almost midexponential phase in the aerobic batch culture with 20 g/L galactose, samples for DNA microarray and intracellular metabolite analysis were taken.

Transcriptome Analysis. PCA was applied to normalized data from all strains by using the TM4 software (1). To find common changes and specific changes of evolved mutants, differentially expressed genes (Student's *t* test; $P < 0.01$) were identified by comparison among all evolved mutants and the other strains. Differentially expressed genes were used as input to g:Profiler to find significantly changed metabolic pathways, reactions, and GO terms by using a cumulative hyper-geometric test (2). The results were visualized by using software of MultiExperiment Viewer (Dana–Farber Cancer Institute).

Metabolome Analysis. The concentration of sugar phosphates was measured by ion chromatography (3). Quantification of redox cofactors was done by assay kits of BioVision. Free amino acids were quantified by EZ:faast amino acid analysis kit (Phenomenex) by using gas chromatography-mass spectrometry. To analyze carbohydrates and sterols, harvested cells were directly used for extraction of those metabolites; 10 mg and 100 mg of dry cell weight were used, respectively. Trehalose and glycogen were quantified by measuring the amount of glucose released after treatment with enzymes that degrade these carbohydrates (4). Ergosterol and dihydroergosterol were identified by gas chromatography-mass spectrometry (5) and quantified by high performance liquid chromatography. It was difficult to find a standard for dihydroergosterol, and we therefore quantified this by using a standard curve with ergosterol. For hierarchical clustering, all quantified metabolites were standardized to *z* score to remove different weights for each metabolite. The result of the clustering was visualized by using the software of MultiExperiment Viewer. To analyze significant changes in metabolite concentrations, a Student *t* test ($P < 0.05$) was applied between the three evolved mutants and the reference strain, respectively. Venn diagram was used to represent common and specific features of all of the evolved mutants.

Illumina/Solexa Genome Sequencing. Genome Analyzer (GA-IIx) was used with sequencing cycles of $2 \times 38 + 7$ (index) by Chrysalis 36 cycles version 4.0 as a sequencing kit. The Mapping and Assembly with Quality (MAQ) software version 0.7.1 (http://maq.sourceforge.net) was used with a maximum set at 2 mismatches in the first 24 bases on the reference sequence that was original CEN.PK113-7D genome sequence (6) available at the CENPK genome database (www.sysbio.se/cenpk). The MAQ software was then used to validate the mapping obtained by comparing the paired end of each read orientation and position. The maximal insert size was set to 400 base pairs for this process. The results from the mapping were further used for calling of single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELS) in the reads compared with the reference sequence. The raw candidates of SNPs and InDels having identical mutations in all evolved mutants were removed for further comparative analysis, because those mutations came from the ancestor strain. It was confirmed by resequencing of some of identical mutations with an ancestor strain. Filtering for SNPs was done by selecting of SNPs with Phred-like consensus quality >30 and a coverage depth >3. Some of INDELS were filtered out when the number of reads with INDELS was lower than the number of reads without INDELS. To perform annotation, GlimmerHMM was used for prediction the gene structures by training the gene models from strain CEN.PK113-7D and S288c (6). Especially, to calculate the coverage folds of gal genes in each evolved mutant, the count function of integrative genome viewer (IGV software) was used. A 200-bp window was set to estimate the average coverage folds. The results are summarized in Dataset S2.

1. Saeed AI, et al. (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.
2. Reimand J, Kull M, Peterson H, Hansen J, Vilo J (2007) g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* 35 (Web Server issue):W193–200.
3. Smits HP, Cohen A, Buttler T, Nielsen J, Olsson L (1998) Cleanup and analysis of sugar phosphates in biological extracts by using solid-phase extraction and anion-exchange chromatography with pulsed amperometric detection. *Anal Biochem* 261:36–42.
4. Parrou JL, François J (1997) A simplified procedure for a rapid and reliable assay of both glycogen and trehalose in whole yeast cells. *Anal Biochem* 248:186–188.
5. Kelly SL, et al. (1995) Purification and reconstitution of activity of Saccharomyces cerevisiae P450 61, a sterol delta 22-desaturase. *FEBS Lett* 377:217–220.
6. Otero JM, et al. (2010) Whole genome sequencing of Saccharomyces cerevisiae: From genotype to phenotype for improved metabolic engineering applications. *BMC Genomics* 11:723–740.

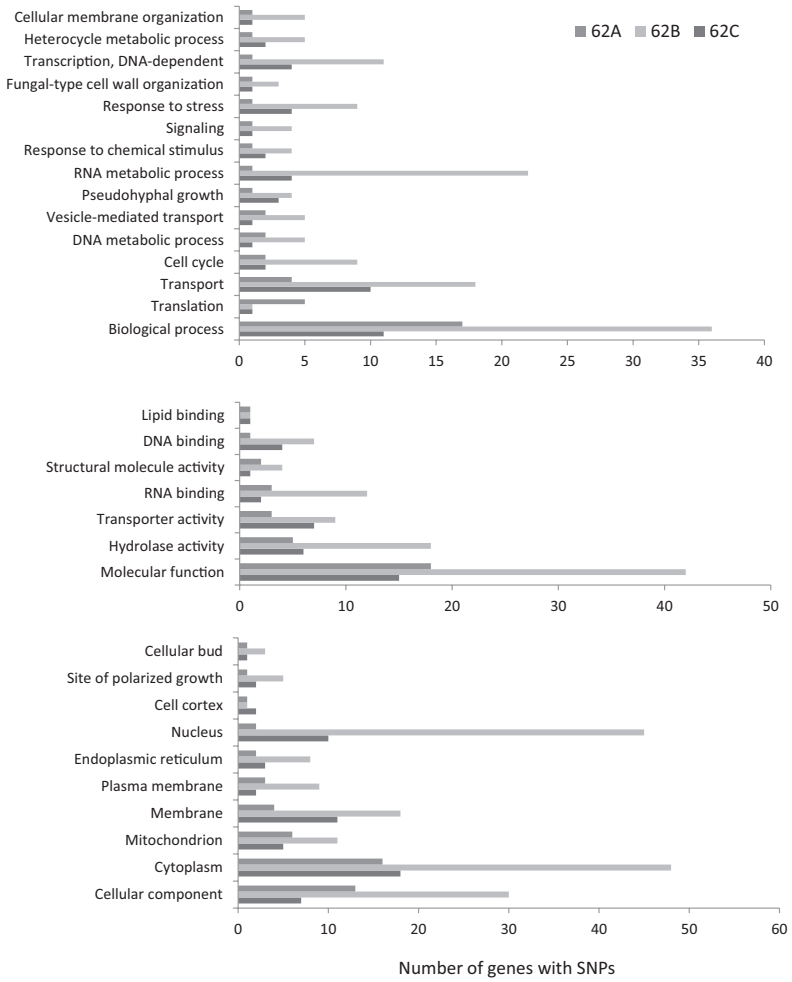


Fig. S1. Gene ontology (GO) enrichment analysis by GO slim mapper based on *S. cerevisiae* genome background.

Table S1. Characterization of phenotype in aerobic batch culture on galactose

Strains	Maximum specific growth rate (h-1)	Biomass yield (C-mol/C-mol galactose)	Ethanol yield (C-mol/C-mol galactose)	CO_2 yield (C-mol/C-mol galactose)
REF	0.21 ± 0.00	0.46 ± 0.00	0.13 ± 0.01	0.39 ± 0.00
SO16	0.17 ± 0.01	0.30 ± 0.01	0.31 ± 0.01	0.31 ± 0.00
PGM2	0.21 ± 0.01	0.35 ± 0.02	0.32 ± 0.02	0.27 ± 0.03
62A	0.26 ± 0.00	0.48 ± 0.01	0.14 ± 0.00	0.31 ± 0.05
62B	0.26 ± 0.00	0.45 ± 0.00	0.19 ± 0.01	0.29 ± 0.02
62C	0.26 ± 0.00	0.41 ± 0.01	0.26 ± 0.02	0.30 ± 0.01

Table S2. The concentration of measured metabolites

μmol/g DCW	REF-1	REF-2	SO16-1	SO16-2	PGM2-1	PGM2-2	62A-1	62A-2	62B-1	62B-2	62C-1	62C-2
Galactose (intracellular)	149.0	169.1	72.2	114.7	117.5	139.4	215.5	309.0	198.5	240.7	136.5	179.3
Galactitol	29.37	40.23	19.42	20.88	12.20	16.71	29.93	35.98	35.42	33.74	35.53	34.79
Galactose 1P	0.618	0.751	0.373	0.396	0.242	0.272	0.225	0.346	0.279	0.411	0.278	0.290
Glucose 1P	0.279	0.310	0.106	0.115	0.037	0.032	0.054	0.078	0.061	0.090	0.051	0.057
Glucose 6P	0.246	0.305	0.209	0.200	0.155	0.199	0.181	0.229	0.207	0.208	0.227	0.242
Fructose 6P	0.051	0.068	0.052	0.044	0.037	0.050	0.041	0.060	0.048	0.042	0.051	0.056
Fructose 1,6 dP	0.540	0.620	0.954	0.817	0.570	0.711	0.630	0.539	1.828	2.246	0.650	0.627
ATP	3.380	3.857	3.073	2.746	2.003	2.122	3.381	3.221	3.422	4.359	2.950	2.934
ADP	0.668	0.683	0.594	0.532	0.372	0.409	0.512	0.522	0.512	0.559	0.440	0.418
AMP	0.041	0.045	0.059	0.043	0.033	0.043	0.047	0.048	0.056	0.048	0.033	0.033
NADPH	0.167	0.256	0.126	0.202	0.139	0.177	0.090	0.097	0.183	0.133	0.099	0.180
NADP	0.277	0.306	0.197	0.231	0.173	0.266	0.287	0.299	0.235	0.248	0.220	0.318
NADH	0.275	0.301	0.212	0.201	0.115	0.207	0.186	0.223	0.149	0.094	0.210	0.253
NAD	3.989	5.057	2.814	2.752	2.360	3.566	3.956	4.032	2.585	1.517	4.076	3.396
Alanine	0.683	0.768	0.478	0.390	0.475	0.797	0.813	0.949	1.010	0.798	0.793	0.832
Glycine	0.571	0.516	0.361	0.339	0.315	0.524	0.411	0.468	0.648	0.699	0.450	0.468
Valine	0.323	0.293	0.152	0.159	0.179	0.284	0.305	0.261	0.479	0.431	0.260	0.282
Leucine	0.106	0.095	0.096	0.091	0.071	0.104	0.136	0.110	0.159	0.149	0.093	0.115
Isoleucine	0.096	0.074	0.047	0.052	0.049	0.080	0.063	0.093	0.155	0.116	0.045	0.081
Threonine	0.844	0.736	0.608	0.600	0.666	0.885	0.918	0.891	1.494	1.342	0.676	0.902
Proline	0.226	0.178	0.000	0.000	0.036	0.106	0.233	0.255	0.240	0.247	0.293	0.345
Asparagine	0.298	0.266	0.243	0.239	0.220	0.323	0.344	0.321	0.494	0.459	0.253	0.297
Aspartic Acid	2.247	1.890	1.377	1.428	1.618	1.835	2.846	2.288	2.910	2.788	2.296	2.499
Phenylalanine	0.040	0.053	0.018	0.021	0.037	0.033	0.055	0.012	0.094	0.097	0.023	0.043
Glutamic acid	7.528	7.467	4.897	4.720	5.511	7.474	8.076	7.361	10.288	9.354	7.057	7.623
Glutamine	2.432	1.685	1.075	1.553	1.008	1.823	3.765	2.921	5.588	5.611	2.949	2.912
Lysine	0.676	0.696	0.701	0.684	0.672	0.996	0.842	0.768	1.026	0.938	0.617	0.601
Histidine	0.586	0.601	0.722	0.712	0.533	0.805	0.821	0.775	0.782	0.681	0.492	0.519
Tyrosine	0.269	0.272	0.302	0.292	0.265	0.331	0.340	0.311	0.408	0.365	0.230	0.246
Trehalose	2.37	2.47	5.09	4.49	2.22	3.99	10.80	10.36	17.90	17.70	10.65	9.91
Glycogen [mg (as glucose)/g DCW]	26.33	26.22	37.04	37.34	24.22	25.67	53.20	53.04	33.18	33.00	49.68	48.05
Ergosterol	27.60	26.61	28.95	28.11	19.06	20.66	18.52	18.00	3.02	3.23	19.51	18.94
Dihydroergsterol*	1.16	1.19	1.88	1.75	0.90	0.96	0.75	0.79	32.79	37.06	0.85	0.80

*Dihydroergosterol was estimate from ergosterol standard curve.

Table S3. Overall Illumina/Solexa genome sequencing results

Sequencing parameters	62A	62B	62C
No. of reads	5,605,504	18,203,846	5,239,106
Total bases,* bp	213,009,152	691,746,148	199,086,028
Coverage fold	17	55	16
Undetermined base	158,723	86,791	171,362
Genome percent reference coverage, † %	98.7	99.3	98.6
No. of supercontigs	17	17	17
Chromosomes	16	16	16
Mitochondria	1	1	1

*38 bases per sequence read for two cycles.
†Based on genome consensus sequence length of CEN.PK113-7D of 12,155,742 base pairs.

Table S4. Detection of mutations across different genomes

Sequencing parameters	62A/others*	62B/others	62C/others
Total number of mutations	44	334	40
Total number of SNPs	21	104	29
Coding region	6	29	11
Noncoding region	15	75	18
Total number of insertions and deletions	23	230	11
Coding region	0	11	3
Noncoding region	23	219	8

*All other strains including the reference strain, CEN.PK113-7D and the other evolved strains.

Table S5. Characterization of maximum growth rate of site-directed mutants in flask culture

Strain	Genotype	U_{max}	STD (n = 3)	Increase, %	t test (P value)
CEN.PK117-5D	*MATa MAL2-8c SUC2 ura3*	0.199	0.013	—	—
RB	*MATa MAL2-8c SUC2 ura3 RAS2$^{Tyr-112}$*	0.219	0.010	10	0.05

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)