Supplementary Information

Table of Contents

Supplementary Note 1

Analysis of predicted essential genes and transposon stability analysis

We obtained insertions throughout the *Z. mobilis* main chromosome (about 2 Mbp) and its five plasmids (pZM4001 thru pZM4005), suggesting that our TN5-based transposon was randomly inserting into the genome, as expected (Supplementary Figure 3A). Typically, the distribution of transposon insertions can be used to infer essential genes, and usually there is a strong bias for insertions in non-essential genes (Christen *et al*, 2011). We identified 336 putative essential genes in *Z. mobilis* using MicrobesOnline tree orthologs (Dehal *et al*, 2010) and data from another α-proteobacterium, *Caulobacter crescentus* (Christen *et al*, 2011). We mapped "good" insertions (i.e. in the central 5-80% of the coding region) in 82% of the 336 expected-essential genes (Supplementary Table 3), which is not significantly less than the rate of 85% for predicted nonessential genes ($P = 0.10$, Fisher exact test). To our knowledge, this is the first example of a large-scale transposon mutagenesis study in bacteria with this unusual distribution of insertion sites.

Because it is highly unlikely that *Z. mobilis* has no essential genes, we used a combination of transposon stability analysis and comparative genome hybridization (CGH) to understand this unusual insertion bias (Supplementary Figure 4). We developed a transposon stability assay to study 28 mutants that had been single colony purified from our mutant collection. Three of these 28 transposon mutants had insertions in *leuS*, *ftsZ*, and *rpoB*, which are predicted essential genes, and the remaining 25 were putative hydrolysate tolerance genes identified in this study (Supplementary Table 8). First, we used PCR primers P1_ZMOxxxx and P2_ZMOxxxx flanking each transposon insertion site (Supplementary Figure 4A) and found that some mutants were "mixed" and had both a wild type and mutant copy of the gene of interest (Supplementary Figure 4B). The identity of these bands was confirmed by TOPO cloning (Invitrogen) and sequence analysis. To perform the transposon stability assay, each mutant was streaked out from a -80^{o}C freezer stock for single colonies on a ZRMG + KAN plate. A single colony was then used to streak on a ZRMG plate without antibiotics, and also used to inoculate a liquid culture (ZRMG + KAN) for gDNA isolation and PCR analysis using the primers described above. Ten single colonies were then isolated from the ZRMG plate without

antibiotic and each streaked again for single colonies on a ZRMG + KAN plate (Plate 1) and also on a ZRMG plate without antibiotics (Plate 2). We scored each strain as the number of Kan^R isolates on Plate 1 (Supplementary Table 8). The ten colonies from the ZRMG without antibiotic were examined by PCR using P1 and P2 primers, and colonies that grew on the kanamycin plate were also examined. We found that mixed mutants (two bands by PCR) after growth without kanamycin selection, at some frequency, could lose their transposon leaving only a single band by PCR corresponding to the wild type gene (Supplementary Figure 4C,D). Thus, we identified two classes of transposon insertions in our *Z. mobilis* library*:* stable Kan^R mutants (Supplementary Figure 4E,F), and mixed, unstable Kan^R mutants. Out of 28 mutants tested, we identified 12 mixed mutants, which included three predicted essential genes (*rpoB, ftsZ, leuS*) and four genes from a single operon (*ZMO1429-ZMO1432*). We suspect that most of the mixed mutants in our *Z. mobilis* library are in essential genes.

Comparative genome hybridization

To understand the difference between the stable and mixed classes further, we performed comparative genome hybridization using custom designed 12-plex Nimblegen tiling arrays (array design and data can be found at: http://genomics.lbl.gov/supplemental /zm4hydrolysate/). Genomic DNA was isolated from 12 samples, which included 5 predicted essential transposon mutants (Supplementary Table 6), and wild type *Z. mobilis* as a control. Samples were labeled, hybridized, washed, and scanned according to the manufacturer's protocols (http://www.nimblegen.com/products/cgh/index.html). Data was analyzed using custom R scripts (available upon request) in order to generate an overview and local plot of the genome for each of the 12 samples (see Supplementary Figure 4G-L for some example CGH plots). We found no evidence of partial or local genome duplication for any of the 12 strains, suggesting that whole genome duplication was the most likely explanation that would explain the presence of both wild type and mutant gene copies in our mixed transposon mutants.

We hypothesize that *Z. mobilis* is normally polyploid and that transposon insertions can be recovered in both non-essential and essential genes (Supplementary Figure 5A). Random assortment of chromosomes after a single transposon insertion in a non-essential gene can lead to a "homozygous" stable Kan^R mutant (Supplementary Figure 5C). For an essential gene, or a gene that might have a severe fitness defect when deleted, there exists a "mixed" state that is stable only in the presence of kanamycin selection (Supplementary Figure 5B). In the absence of drug selection, random assortment can lead to loss of the transposon, and reversion back to wild type (Supplementary Figure 5D). In addition, mutants in predicted essential genes had a lower abundance in our start pool (Supplementary Figure 3B) making it difficult to measure changes in their fitness. In this paper, we did not pursue mixed mutants or predicted essential mutants; instead we focused our follow-up studies on hydrolysate tolerance genes that had a stable Kan^R transposon insertion.

Supplementary Figure Legends

Supplementary Figure 1. Cluster dendrogram of hydrolysate samples based on inhibitor concentrations. Hydrolysate samples were clustered using their inhibitor concentrations (Supplementary Table 1) and Pearson correlation. Samples are labeled according to field location and plant type. Illinois field locations are: Orr (ORR), Havana (HAV), Brownstown (BTN), Fairfield (FRF). Three separate clusters are indicated: Miscanthus (M_HAV, M_BTN, M_ORR), Switchgrass (S_FRF, S_BTN, S_ORR) and Batch miscanthus (M_batch1, M_batch2). The two batch hydrolysate samples were prepared at 200^{o}C instead of 180^{o}C, have higher concentrations of inhibitors, and form a third independent cluster.

Supplementary Figure 2. Growth inhibition for eight different hydrolysate samples. *Z. mobilis* was grown in ZRMG rich media supplemented with different plant hydrolysates (% v/v) as indicated. We performed at least three biological replicates. For each plant hydrolysate, representative data from a single experiment is shown. Illinois field locations are: Brownstown (BTN), Fairfield (FRF), Orr (ORR), and Havana (HAV). All experiments were performed in 0% (control), 10% hydrolysate (HZ) and 20% hydrolysate (HZ), as indicated in the legend, except for batch 2 hydrolysate (8% HZ). All eight hydrolysate samples inhibited growth of *Z. mobilis*. (A) switchgrass (BTN) (B) switchgrass (FRF) (C) switchgrass (ORR) (D) miscanthus (BTN) (E) miscanthus (HAV) (F) miscanthus (ORR) (G) miscanthus batch 1 (H) miscanthus batch 2

Supplementary Figure 3. Genome-wide barcoded transposon collection in *Z. mobilis* ZM4. (A) Histogram of transposon insertions along the five plasmids (pZZM401, pZZM402, pZZM403, pZZM404, pZZM405) and the main chromosome (about 2 Mbp). (B) Histogram of normalized barcode intensities for transposon mutants in predicted essential versus non-essential genes. Mutants in predicted essential genes (TRUE) have a significantly lower abundance than other mutants (FALSE) after recovering the up-pool or dn-pool from the freezer.

Supplementary Figure 4. Identification of two classes of transposon mutants. (A) PCR analysis of each transposon insertion was performed using genome specific P1 and P2 flanking primers (see Supplementary Table 6). Only the mutant allele (*ZMOxxxx::TN5*) was detected in stable mutants. Both the wild type (*ZMOxxxx*) and mutant alleles were detected in mixed mutants. (B) PCR analysis of nine mutants using flanking P1 and P2 primers. Six mixed mutants (*ftsZ::TN5*, *leuS::TN5*, *rpoB::TN5*, *ZMO0200::TN5*, *ZMO1430::TN5*, *ZMO1431::TN5*) and three stable mutants (*ZMO0759::TN5*, *ZMO1490::TN5*, *ZMO1723::TN5*) are shown. (C) PCR analysis of a mixed mutant *ZMO0200::TN5*. In the absence of kanamycin selection only the wild type allele remains. PCR of the starting strain from the freezer stock "F" and 10 colonies after the TN stability assay are shown. (D) PCR analysis of the mixed mutant in the predicted essential gene *rpoB::TN5*. In the absence of kanamycin selection only the wild type allele remains. PCR of the starting strain from the freezer stock "F" and 10 colonies after the TN stability assay are shown. (E-F) PCR analysis of the stable mutant *ZMO1490::TN5*. Only the mutant allele is present with and without kanamycin selection. Ten colonies after the TN stability assay and the freezer stock "F" are shown. (G-L) Comparative genome hybridization (CGH) results for four predicted essential transposon

mutants. CGH plots of the main chromosome are shown for: *ftsZ::TN5*, *glyS::TN5*, *leuS::TN5*, and *secA::TN5*. The green lines in (H) and (I) are smoothed CGH plots and the purple arrows in (H) are predicted genes in the putative prophage region. No large-scale genome duplications were observed. Two more detailed plots are also shown for *ftsZ::TN5*. (H) CGH results for the putative prophage region (*ZMO1934-ZMO1944*) for *ftsZ::TN5*. This region appeared to vary in copy number in our tested strains (compare peak at position ~2,000,000 in panels G, J, K, and L). (I) CGH results for the local region surrounding *ftsZ::TN5*. No local gene duplications were detected. The full CGH dataset can be found online (http://genomics.lbl.gov/Supplementary /zm4hydrolysate/).

Supplementary Figure 5. Model for unstable heterozygosity in *Z. mobilis*. (A) Transposon insertion (Tn) in an essential gene (ess) is possible if wild type copies of the gene are also present. Both wild type and mutant copies of the gene can be detected by PCR confirming a heterozygous state (as in Supplemental Figure 4). (B) After random assortment of chromosomes, in the presence of kanamycin, only the mixed mutant class survives for insertions in an essential gene. (C) For non-essential transposon insertions (non), random assortment of chromosomes in the absence of kanamycin selection results in a homozygous stable state. (D) Mixed mutants in essential genes are unstable and can lose the transposon in the absence of kanamycin selection (see Supplemental Figure 4C,D). Please note that although we illustrate this model with 3 copies of the chromosome, the data presented in this paper cannot be used to determine the exact number of chromosomes that cells of *Z. mobilis* can tolerate and maintain.

Supplementary Figure 6. Validation of our *Z. mobilis* pooled fitness assay. (A) Scatterplot of *Z. mobilis* gene fitness data in rich media versus in minimal media. Negative gene fitness values indicate that the gene is important for growth in that condition. Auxotrophs are indicated in color according to the graph legend, according to predicted TIGR functions (http://www.jcvi.org/cgi-bin/tigrfams/index.cgi): Met. (methionine), Aromatic A.A. (aromatic amino acids), A.A. synthesis (amino acid synthesis), Other (non-auxotrophic genes). (B) Scatterplot of *Z. mobilis* gene fitness data in minimal media versus in minimal media supplemented with casamino acids. Many of the auxotrophs are rescued, as indicated by significantly improved fitness (i.e. closer to zero). (C) Scatterplot of *Z. mobilis* fitness data in minimal media versus in minimal media supplemented with methionine. Six methionine auxotrophs are rescued, indicated by blue "x" symbols. The *metC* gene (*ZMO0327*) is indicated by an arrow. (D) Rescue of the *metC::TN5* mutant by growth on minimal media supplemented with methionine or complementation with a plasmid expressing the *metC* gene (pMetC). pJS71 is the vector backbone used to construct pMetC and is used as a negative control. (E) Scatterplot of strain fitness data for 6% (v/v) batch 1 miscanthus hydrolysate in the dn-pool versus in the up-pool. Strain fitness values are plotted only for identical strains that are present in both pools. This example plot demonstrates that mutants present in both the up-pool and dn-pool show highly correlated fitness values: r(Strain) = 0.948. (F) Operon correlation across 37 hydrolysate experiments. Typical operon correlations are > 0.5. (G) Quality metric plot (strain versus operon correlation) for all *Z. mobilis* fitness experiments. In a typical fitness experiment, strain and operon correlations were about 0.8 and 0.5, respectively. Experiments with quality metrics below these values were repeated to

ensure that their low values were not due to technical error. Fitness experiments with low quality metric scores that could not be repeated were discarded from our dataset. Out of 210 experiments, 18 were discarded. Rich media (ZRMG); Hydrolysate, ZRMG supplemented with plant hydrolysate; Synthetic hydrolysate, ZRMG supplemented with a mixture of hydrolysate components; Components, ZRMG supplemented with individual inhibitors present in hydrolysate.

Supplementary Figure 7. Gene fitness heatmap for 44 tolerance genes across all hydrolysate samples, synthetic mixtures, and baseline media experiments. The non-averaged gene fitness values (log_2 ratios) are colored according to the color bar. The specific conditions and experiment numbers (in parentheses) are indicated along the Y-axis (see Supplementary Table 5). Systematic gene names (*ZMOxxxx*) are indicated along the X-axis, and are color coded according to the legend. Some of the putative tolerance genes, such as the cytochrome c related (green in legend), have variable gene fitness values in different hydrolysate experiments; however, for this work we used the average value based on 37 hydrolysate fitness experiments. The source of this variability for this class of genes is not known.

Supplementary Figure 8. Structure of inhibitors found in fitness clusters 1-10 (A-J). Each fitness cluster (1-10) is numbered according to Figure 2. Chemicals that are in the same fitness cluster only differ by a single functional group. For each fitness cluster indicated, the chemical structures are shown, and the relevant functional group that differs is also indicated (OH, ester, ethyl, methyl, or C=C)

Supplementary Figure 9. Transposon mutants in *Z. mobilis* tolerance genes have severe growth defects in plant hydrolysate. Growth curves are shown for mutant strains in five tolerance genes that we identified as important for growth in hydrolysate (HZ). Experiments were performed in a 96-well format using a microplate reader. Data shown is the average of 5-6 biological replicates. (A-E) Mutants are indicated on each plot by systematic gene name *ZMOxxxx::TN5*. Blue lines are the control experiment, grown in rich media without hydrolysate. Red lines indicate growth in the presence of rich media supplemented with 10% (v/v) batch 1 hydrolysate. (F) Growth curves for wild type *Z. mobilis* with and without hydrolysate.

Supplementary Figure 10. Complementation of fitness defects in hydrolysate and methylglyoxal. (A-F) Single tolerance genes were expressed from a replicative plasmid using either the P_{bad} or P_{gap} promoters, and grown with and without 10% batch 1 hydrolysate (HZ) or 1.5 mM methylgloxal (MG), respectively. P_{bad} expression was induced with 2% arabinose. P_{gap} is a constitutive promoter. Control strains used an empty plasmid construct, pJS71. Experiments were performed in 96-well format and the data shown is the average of 2 biological replicates. (A-D) The *ZMO0100*, *ZMO1722*, and *ZMO0759* complemented strains grow in rich media supplemented with 10% (v/v) batch 1 hydrolysate as well as the wild type control (compare arrows in B,C,D with arrow in A). (E, F) The *ZMO0759* complemented strain grows in rich media supplemented with 1.5 mM methylglyoxal as well as the wild type control (compare arrow in F with arrow in E).

Supplementary Figure 11. Comparison of growth inhibition in batch miscanthus hydrolysate (batch 1 and batch 2) and synthetic hydrolysate mixtures (SYN-10 and SYN-37) in *Z. mobilis* and *S. cerevisiae*. (A) Growth inhibition plot of *Z. mobilis* in different miscanthus and synthetic hydrolysate concentrations (% v/v). Growth rate was determined by calculating the slope between 0.2 and 0.3 OD_{600} for experiments using *Z. mobilis* UP and DN pools. (B) Plot showing the time to mid-log phase in *S. cerevisiae* in different miscanthus and synthetic hydrolysate concentrations (% v/v). In this case, time to mid-log phase was calculated instead of the growth rate. *S. cerevisiae* inhibition in hydrolysate was measured by the length of the lag phase because once the cells begin exponential phase, their growth rates are similar to rich media without inhibitor. The time to mid-log phase was calculated from pool experiments using the *S. cerevisiae* homozygous deletion pool and by determining the time (hrs) at which $(OD_{600}min+OD_{600}max)/2$ was observed. In each plot, the colored lines represent the best fit from linear regression. For both organisms, the linear regressions indicate that the real hydrolysate material is more inhibitory than the synthetic mixtures. Data for 30% or higher is not shown for genuine hydrolysate, because cultures grown in 30% or higher in genuine hydrolysate failed to double. This also confirms that genuine hydrolysate is more inhibitory than our synthetic hydrolysates. To verify that genuine hydrolysates were significantly more inhibitory than synthetic hydrolysates, we used ANOVA to test whether the slopes were different. Specifically, we regressed slope or time-to-mid-log-phase versus concentration and an indicator variable "synth" (1 if synthetic, 0 if genuine). We tested if adding the term synth * concentration to this regression significantly improved the fit. In both organisms, this term was statistically significant ($P < 10^{-5}$ in *Z. mobilis*, $P < 10^{-15}$ in *S. cerevisiae*).

Supplementary Figure 12. Scatterplots of average gene fitness data for *Z. mobilis* and *S. cerevisiae* in synthetic mixtures and real hydrolysates. Tolerance genes are colored according to the legends. (A) Scatterplot of average gene fitness in SYN-10 (average of 4 experiments) versus average gene fitness in hydrolysate (average of 37 experiments) for the *Z. mobilis* transposon mutant pools. (B) Scatterplot of average gene fitness in SYN-10 versus average gene fitness in SYN-37 (average of 4 experiments) for *Z. mobilis*. (C) Scatterplot of average gene fitness in SYN-10 (average of 4 experiments) versus average gene fitness in batch 1 miscanthus hydrolysate (average of 6 experiments) for the *S. cerevisiae* homozygous deletion mutant pool. (D) Scatterplot of average gene fitness in SYN-10 versus average gene fitness in SYN-37 (average of 6 experiments) for *S. cerevisiae*. See Supplemental Table 5 for the list of experiments used to calculate average gene fitness values.

Supplementary Figure 13. Identification of 99 putative hydrolysate tolerance genes in *S. cerevisiae*. Scatterplot of average gene fitness data in YPD rich media (average of 3 replicates) versus average gene fitness in batch 1 hydrolysate (average of 6 experiments). Putative tolerance genes ($fitness_{hydrolysate} < -1$ and $fitness_{hydrolysate} < fitness_{rich} - 1$) are indicated on the plot and color coded according to their predicted functional categories. OCA stands for “Oxidant-induced Cell cycle Arrest”. The dashed black lines indicate the cutoff used to select tolerance genes. See Table II and

Supplementary Dataset 3 for the full list of 99 tolerance genes identified in this study and their fitness values.

Supplementary Figure 14. Analysis of the P_{bad} and P_{gap} promoters in *Z. mobilis* using GFP fluorescence. (A) Diagram of the GFP reporter construct used. The P_{bad} and P_{gap} promoters were fused to the second amino acid (AA2) of "superfolder" GFP (GFP). (B) Relative promoter activity for constitutive P_{gap} and titration of P_{bad} promoter using arabinose induction. Samples were collected in log phase and GFP fluorescence quantified for wild type + P_{gap}-*GFP* and wild type + P_{bad}-*GFP*. Normalized fluorescence (RFU/OD_{600}) is reported relative to an empty plasmid control strain (wild type + pJS71). Data shown is the average of three biological replicates and error bars indicate standard deviation.

Supplementary Figure 15. Analysis of *Z. mobilis* overexpression strains by western blot. (A-D) Western blot data for P_{bad} overexpression strains (wild type + P_{bad}-*ZMOxxxx*), as indicated. Expression constructs had an N-terminal FLAG tag allowing relative protein levels to be determined by western blot. Approximate molecular weights are indicated in kDa. Ten strains (indicated by red text) were selected for complementation studies (e.g. see Supplemental Figure 10), including four strains that were also used for batch fermentation studies (Supplemental Figure 16).

Supplementary Figure 16. Batch fermentation profiles for three *Z. mobilis* overexpression strains in rich media + 8% batch 2 hydrolysate (% v/v). Data shown is the average of two biological replicates. The overexpression strains do not have improved fermentation performance relative to the control strain. Glucose (g/L), ethanol (g/L), and OD_{600} are indicated on each plot. (A) WT + pJS71 plasmid control strain (B) WT + P_{bad}-*ZMO0100* (C) WT + P_{bad}-*ZMO0760* (D) WT + P_{bad}-*ZMO1722*.

Supplementary Figure 17. Identification of eleven genes important for growth in hydrolysate under anaerobic conditions. Scatterplot of average gene fitness in anaerobic rich media (n = 4) versus average gene fitness in anaerobic rich media supplemented with batch 2 hydrolysate (n = 2, 15% and 20% batch 2 experiments were averaged for this analysis, see Supplementary Table 5). Tolerance genes are colored according to the legend. Using our standard criterion ($fitness_{hydrolysate} < -1$ and $fitness_{hydrolysate} < fitness_{rich} - 1$), eleven anaerobic hydrolysate tolerance genes were identified. Four of these genes were previously identified in our aerobic studies (*ZMO0759*, *ZMO0100*, *ZMO0101*, *ZMO1490*). Seven additional genes were identified that are important for growth in hydrolysate under anaerobic conditions (bold face black type).

Supplementary Figure 18. Inhibition of wild type *Z. mobilis* growth by methylglyoxal. (A) Growth curve (n = 20, median values plotted) for *Z. mobilis* in rich media (ZRMG), and rich media supplemented with 0.56 mM methylglyoxal (the amount that is present in batch 2 miscanthus hydrolysate). (B) Growth rate was determined by calculating the slope between 0.2 and 0.3 OD_{600}. Boxplot of the growth rates for growth

in ZRMG and ZRMG + 0.56 mM methylglyoxal. Methylglyoxal inhibits growth of wild type *Z. mobilis* ($P < 10^{-5}$, unpaired t-test, n = 20).

Supplementary Table Legends

Supplementary Table 1. Composition of plant hydrolysates and synthetic hydrolysate mixtures determined by GC/MS and LC-RID. The concentrations of 43 compounds (39 potential inhibitors and four sugars) are listed for eight hydrolysate samples and two synthetic mixtures (SYN-10, SYN-37). For each compound, we indicate the average (AVG) concentration (μg/mL or mg/mL as indicated) across 3 measurements and the standard deviation (STDEV). Methylglyoxal concentration was determined only for batch 2 hydrolysate. SYN-10 represents a mixture of the ten most abundant inhibitors found in batch 1 hydrolysate without any added sugars. The concentrations of four compounds in SYN-10 (vanillin, syringaldehyde, vanillic acid, and furoic acid) were higher than expected. SYN-37 is a mixture of 37 inhibitors identified in batch 1 and includes four sugars. SYN-37 does not include 2-hydroxybenzoic acid, because it was not detected in batch 1 miscanthus hydrolysate. The concentrations of two compounds in SYN-37 (benzoic acid and cellobiose) were higher than expected. Six hydrolysate samples were named after their feedstock (MG = miscanthus, SG = switchgrass) and field location in Illinois (ORR = Orr, HAV = Havana, BTN = Brownstown, FRF = Fairfield). Two additional samples contained mixed miscanthus material, and are called batch 1 and batch 2. A summary list of the components used to make SYN-10 and SYN-37 can be found on the second worksheet.

Supplementary Table 2. List of mutants in our *Z. mobilis* barcoded transposon pools. Each pool (up-pool and dn-pool) has 3716 mutants. For each mutant in the pool, systematic gene name (*ZMOxxxx*), VIMSS ID (www.microbesonline.org), RefSeq annotation, Pool (up or dn), 96-well poolplate and plate position, TagModule number (Oh *et al*, 2010), and insertion site position are indicated.

Supplementary Table 3. List of putative essential genes in *Z. mobilis*. Putative essential genes were identified in *Z. mobilis*, based on essential genes identified in a recent *Caulobacter crescentus* study (Christen *et al*, 2011). Of 336 putative essential *Z. mobilis* genes, we identified transposon insertions in 288 genes. The VIMSS ID numbers can be used to identify genes on MicrobesOnline (www.microbesonline.org).

Supplementary Table 4. List of 54 amino acid synthesis genes in *Z. mobilis*. 53 of these genes were annotated with a role in amino acid synthesis using TIGRFAMs (http://www.jcvi.org/cgi-bin/tigrfams). One additional amino acid synthesis gene was annotated by hand, using our gene fitness data, which encodes *metZ* (*ZMO0676*). Most of these genes are important for growth in minimal media; addition of casamino acids restores their gene fitness values closer to zero, as would be expected for rescue of an auxotrophic mutant. The average gene fitness values in rich media, in minimal media, in minimal media supplemented with casamino acids, and in minimal media supplemented with methionine are listed.

Supplementary Table 5. List of experiments used to compile fitness datasets for *Z. mobilis* and *S. cerevisiae*. The experiment number, condition, class, operon correlation, strain correlation, and used (experiment used or averaged), are listed for 192 *Z. mobilis* pool experiments. This list includes 189 experiments described in the main text, plus 3 additional anaerobic hydrolysate experiments (496, 497, 498). "Used" indicates which experiments were used to determine average gene fitness for that condition in all of our analyses. We averaged gene fitness values for replicate experiments (e.g. rich, n = 24) and inhibitors at the same concentration (e.g. 10 mM 4-hydroxybenzoic acid, n =2). For hydrolysate, SYN-10, SYN-37, and the seven additional conditions such as methylglyoxal, we averaged across all concentrations. The experiment number and condition are listed for 21 *S. cerevisiae* homozygous diploid pool experiments on the second worksheet. For all of the yeast conditions (YPD, ZRMG, batch 1, SYN-10, SYN-37), we averaged gene fitness values across all concentrations.

Supplementary Table 6. Strains, Plasmids, Primers used in this study. Primers used for PCR verification of transposon insertion sites, pENTR clones for systematic overexpression and complementation of *Z. mobilis* tolerance genes, strains used for single mutant follow-up and comparative genome hybridization (CGH) studies, and promoter testing constructs are listed.

Supplementary Table 7. Linear regression model results. List of significant components identified, list of coefficients from the R function *lm*, and results of the ANOVA F test (P value, and F value) for Model-16 (rich + 16 components), Model-17 (rich + 17 components), and Model-24 (rich + 24 components). Note: Because the variables in the regression are all highly correlated with each other, the regression coefficients do not have any biological meaning. For example, consider the problem of fitting gene fitness in hydrolysate using gene fitness data from two similar components X and Y. Since X and Y are correlated, we can get a good fit using either X, Y, or a mixture of the two. Multiple regression will choose coefficients that give the best fit, but these coefficients tell us nothing about the relative importance of X versus Y. Next, consider a case where X and Y are similar stresses in which the same pathways are important for fitness, but to differing extents. If the pathways that are slightly more important on Y than on X tend to be slightly less important in hydrolysate, then the best fit will be something like be X - Y/10. This illustrates why the coefficients can be negative. For example, 5 of the components have negative coefficients in the regression, but for all of these components, fitness is positively correlated with fitness in hydrolysate (all $R > 0.6$). The R value for each of the components in Model-17 is listed in the last column of the table.

Supplementary Table 8. Transposon stability results. 28 transposon mutants were studied using a transposon stability assay (see Supplementary Note 1). The number of Kan^R colonies (out of 10 tested) are indicated for two replicate experiments. These experiments were used to classify our transposon insertions as stable (10/10 Kan^R) or mixed (<10 / 10 Kan^R).

Supplementary Dataset 1. Complete gene fitness datasets for *Z. mobilis* and *S. cerevisiae*. Averaged data and complete (non-averaged) datasets are in separate worksheets. Experiment number (See Supplemental Table 5) is indicated for each condition. Data can be viewed using MeV (http://www.tm4.org/mev/about) or the *Z. mobilis* data can be browsed at MicrobesOnline (http://www.microbesonline.org/).

Supplementary Dataset 2. Gene fitness data for 44 *Z. mobilis* hydrolysate tolerance genes identified in this study. Fitness values listed are from the averaged dataset for rich media (ZRMG, average of 24 replicates), rich media supplemented with hydrolysate (average of 37 experiments), and two synthetic hydrolysate mixtures (SYN-37, average of 4 experiments and SYN-10, average of 4 experiments).

Supplementary Dataset 3. Gene fitness data for 99 *S. cerevisiae* tolerance genes identified in this study. Average gene fitness values for the rich media condition (YPD, average of 3 replicates), ZRMG (rich media used to prepare synthetic mixtures, average of 2 experiments), YPD supplemented with batch 1 miscanthus hydrolysate (average of 6 experiments), and two synthetic hydrolysate mixtures (SYN-37, average of 6 experiments and SYN-10, average of 4 experiments) are listed. Ten of these genes are dubious ORFs; however, all of them likely affect the function of overlapping genes, as indicated.

Supplementary References

Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiro L (2011) The essential genome of a bacterium. *Mol Syst Biol* **7:** 528

Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* **38:** D396-400

Larsen RA, Wilson MM, Guss AM, Metcalf WW (2002) Genetic analysis of pigment biosynthesis in *Xanthobacter autotrophicus* Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. *Arch Microbiol* **178:** 193-201

Oh J, Fung E, Price MN, Dehal PS, Davis RW, Giaever G, Nislow C, Arkin AP, Deutschbauer A (2010) A universal TagModule collection for parallel genetic analysis of microorganisms. *Nucleic Acids Res* **38:** e146

Skerker JM, Shapiro L (2000) Identification and cell cycle control of a novel pilus system in *Caulobacter crescentus*. *EMBO J* **19:** 3223-3234

Figure S1

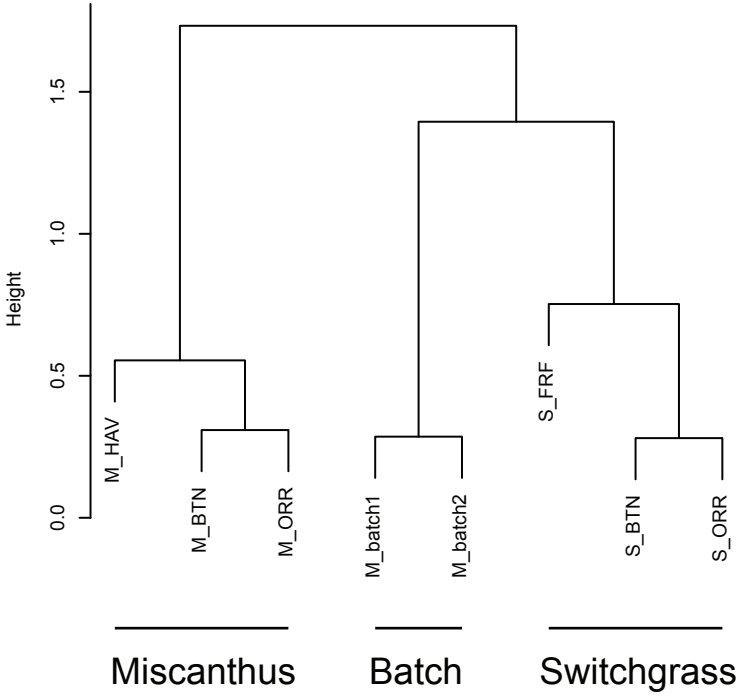

Height
0.0
0.5
1.0
1.5
M_HAV
M_BTN
M_ORR
M_batch1
M_batch2
S_FRF
S_BTN
S_ORR
Miscanthus
Batch
Switchgrass

Figure S2

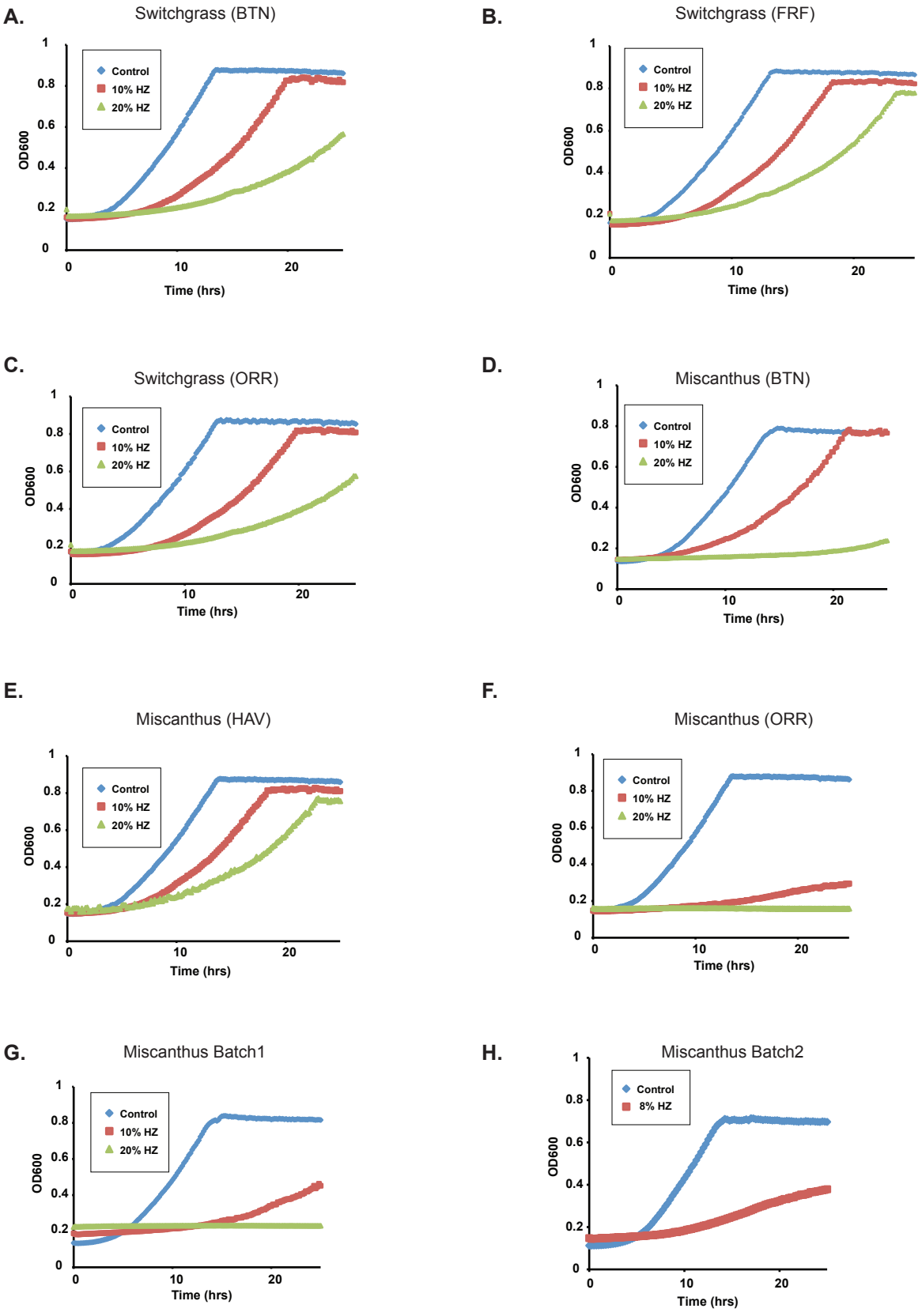
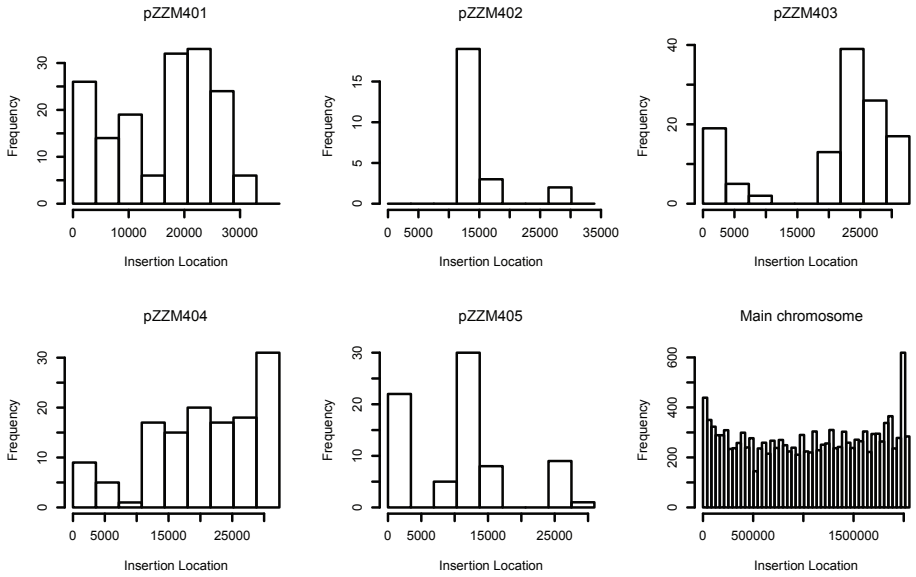

A.
Switchgrass (BTN)
B.
Switchgrass (FRF)
C.
Switchgrass (ORR)
D.
Miscanthus (BTN)
E.
Miscanthus (HAV)
F.
Miscanthus (ORR)
G.
Miscanthus Batch1
H.
Miscanthus Batch2
Control
10% HZ
20% HZ
8% HZ
OD600
Time (hrs)
0
0.2
0.4
0.6
0.8
1
10
20

Figure S3

A.


pZZM401
pZZM402
pZZM403
pZZM404
pZZM405
Main chromosome
Frequency
Insertion Location

B.

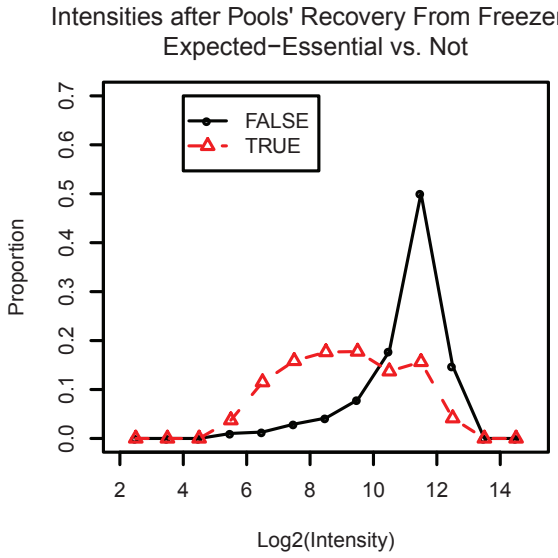

Intensities after Pools' Recovery From Freezer
Expected−Essential vs. Not
FALSE
TRUE
Proportion
Log2(Intensity)

Figure S4

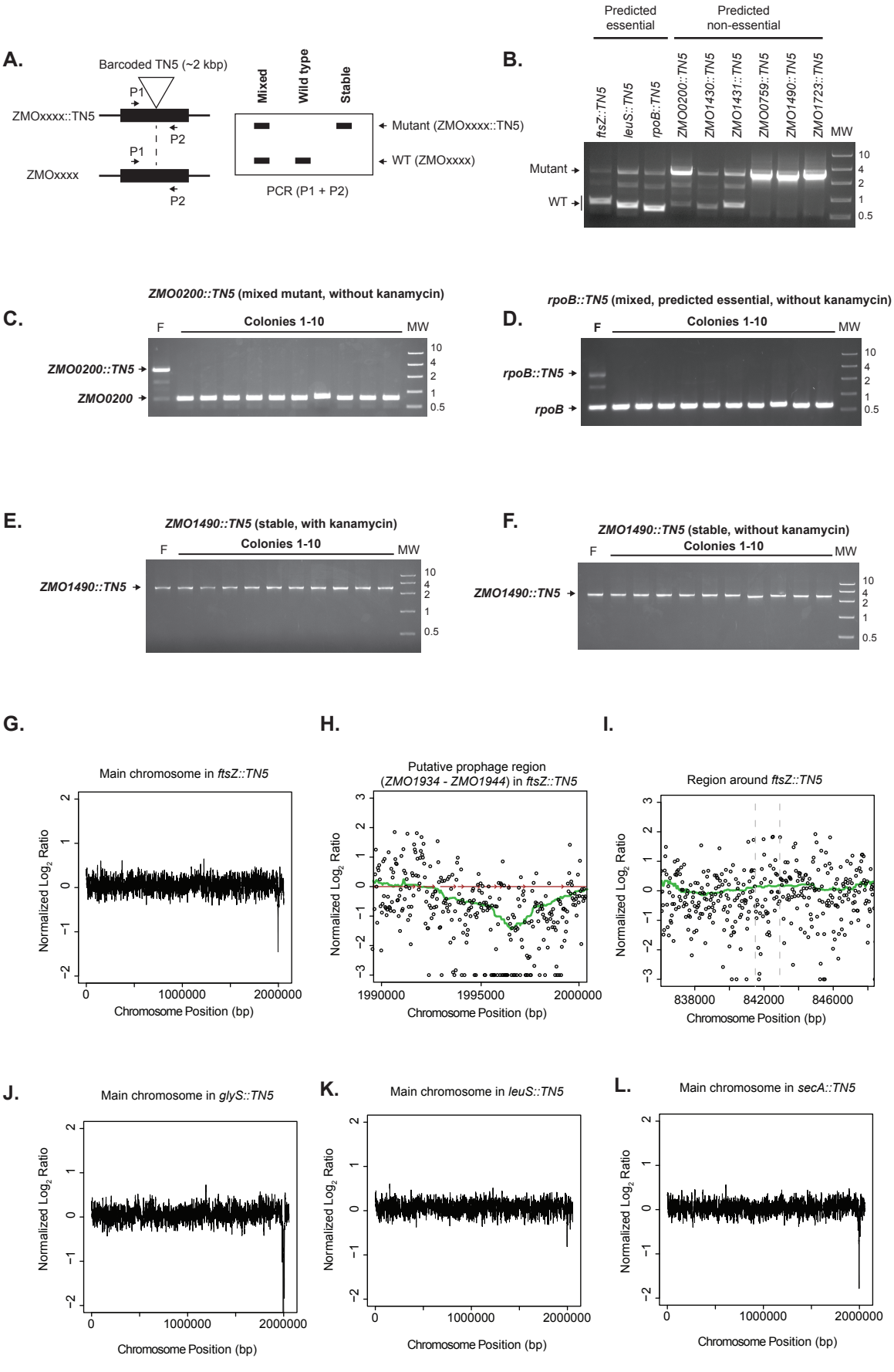

A.
Barcoded TN5 (~2 kbp)
P1
ZMOxxxx::TN5
P2
P1
ZMOxxxx
P2
Mixed
Wild type
Stable
Mutant (ZMOxxxx::TN5)
WT (ZMOxxxx)
PCR (P1 + P2)
B.
Predicted essential
Predicted non-essential
ftsZ::TN5
leuS::TN5
rpoB::TN5
ZMO0200::TN5
ZMO1430::TN5
ZMO1431::TN5
ZMO0759::TN5
ZMO1490::TN5
ZMO1723::TN5
MW
Mutant
WT
10
4
2
1
0.5
C.
ZMO0200::TN5 (mixed mutant, without kanamycin)
F
Colonies 1-10
MW
ZMO0200::TN5
ZMO0200
D.
rpoB::TN5 (mixed, predicted essential, without kanamycin)
E.
ZMO1490::TN5 (stable, with kanamycin)
F.
ZMO1490::TN5 (stable, without kanamycin)
G.
Main chromosome in ftsZ::TN5
H.
Putative prophage region (ZMO1934 - ZMO1944) in ftsZ::TN5
I.
Region around ftsZ::TN5
Normalized Log2 Ratio
Chromosome Position (bp)
J.
Main chromosome in glyS::TN5
K.
Main chromosome in leuS::TN5
L.
Main chromosome in secA::TN5

Figure S5

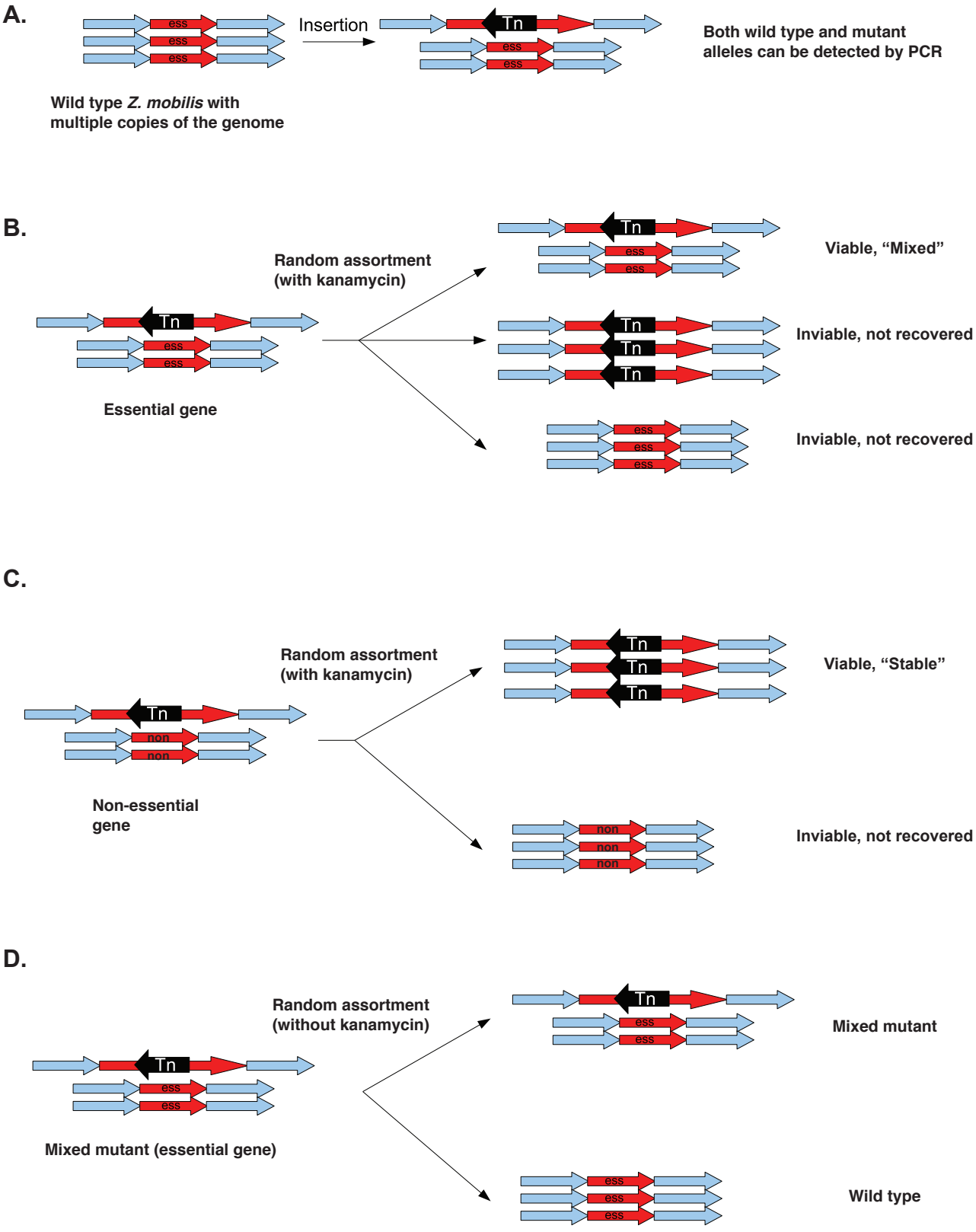

A.
ess
ess
ess
Insertion
Tn
ess
ess
Both wild type and mutant alleles can be detected by PCR
Wild type Z. mobilis with multiple copies of the genome
B.
Random assortment (with kanamycin)
Tn
ess
ess
Viable, "Mixed"
Tn
Tn
Tn
Inviable, not recovered
ess
ess
ess
Inviable, not recovered
Essential gene
C.
Random assortment (with kanamycin)
Tn
non
non
Viable, "Stable"
Tn
Tn
Tn
Inviable, not recovered
non
non
non
Non-essential gene
D.
Random assortment (without kanamycin)
Tn
ess
ess
Mixed mutant
Tn
ess
ess
Wild type
ess
ess
ess
Mixed mutant (essential gene)

Figure S6

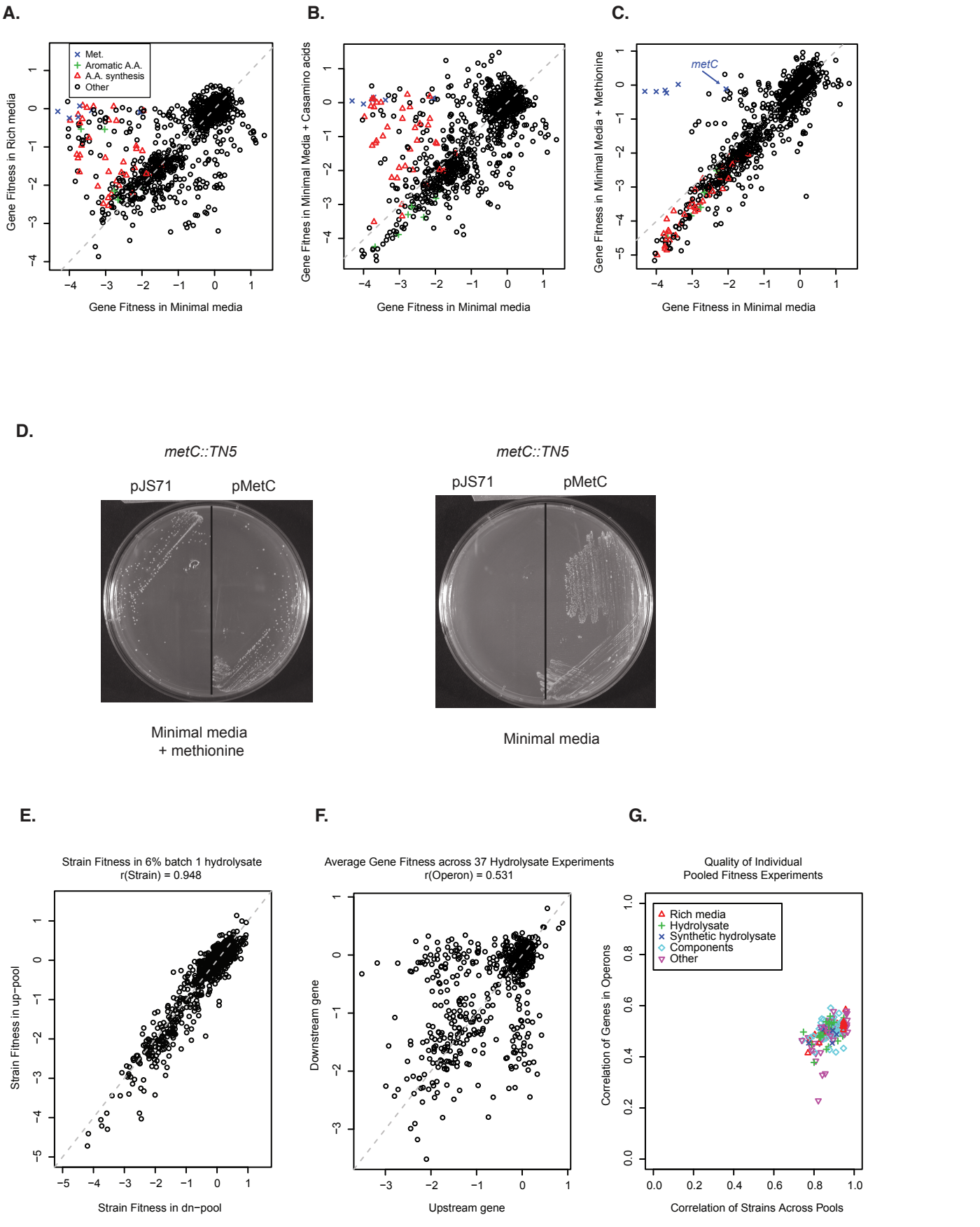

A.
Met.
Aromatic A.A.
A.A. synthesis
Other
Gene Fitness in Rich media
Gene Fitness in Minimal media
B.
Gene Fitnes in Minimal Media + Casamino acids
Gene Fitness in Minimal media
C.
metC
Gene Fitness in Minimal Media + Methionine
Gene Fitness in Minimal media
D.
metC::TN5
pJS71
pMetC
Minimal media
+ methionine
metC::TN5
pJS71
pMetC
Minimal media
E.
Strain Fitness in 6% batch 1 hydrolysate
r(Strain) = 0.948
Strain Fitness in up−pool
Strain Fitness in dn−pool
F.
Average Gene Fitness across 37 Hydrolysate Experiments
r(Operon) = 0.531
Downstream gene
Upstream gene
G.
Quality of Individual
Pooled Fitness Experiments
Rich media
Hydrolysate
Synthetic hydrolysate
Components
Other
Correlation of Genes in Operons
Correlation of Strains Across Pools

Figure S7

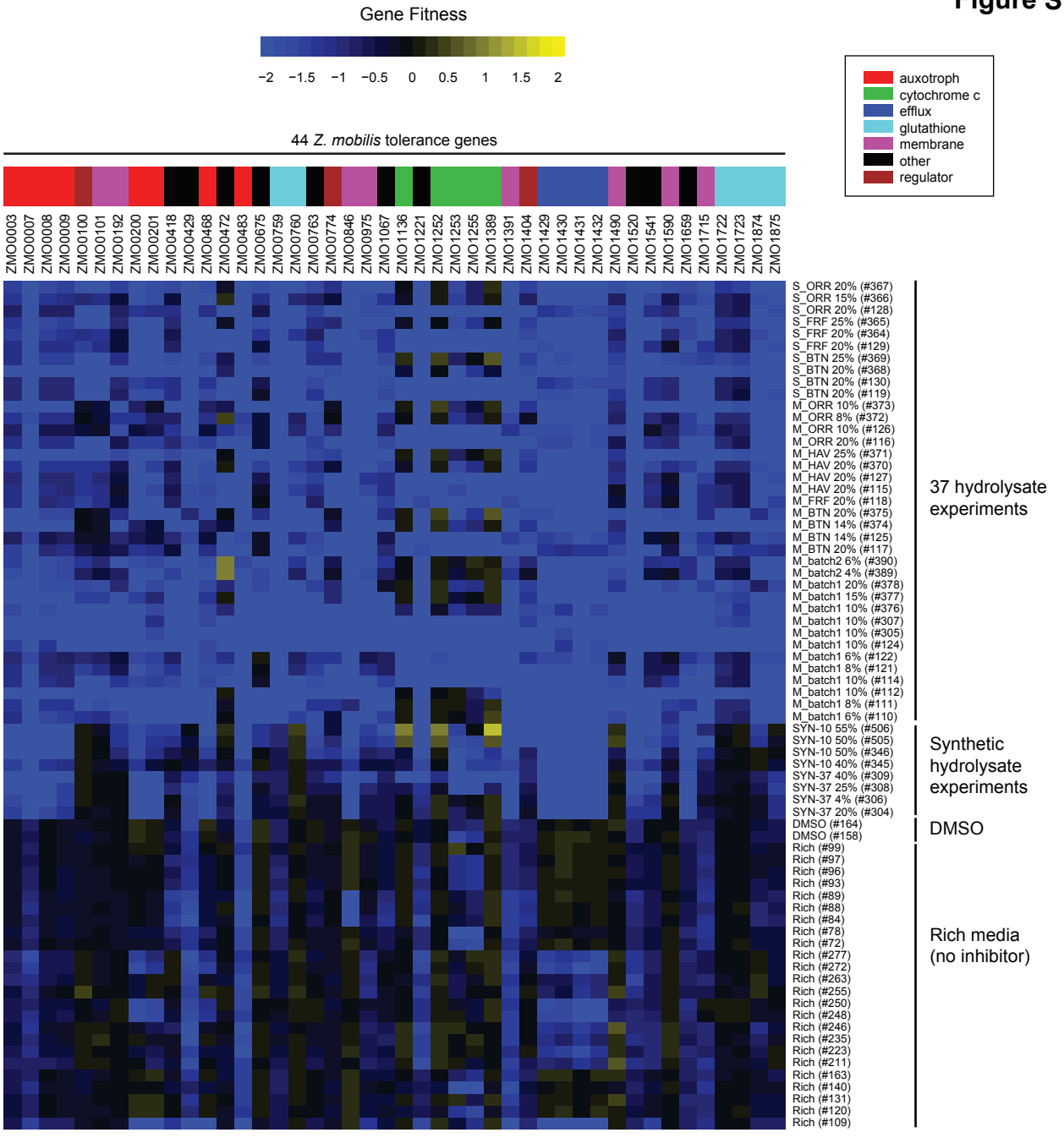

Gene Fitness
−2 −1.5 −1 −0.5 0 0.5 1 1.5 2
auxotroph
cytochrome c
efflux
glutathione
membrane
other
regulator
44 Z. mobilis tolerance genes
ZMO0003 ZMO0007 ZMO0008 ZMO0009 ZMO0100 ZMO0101 ZMO0192 ZMO0200 ZMO0201 ZMO0418 ZMO0429 ZMO0468 ZMO0472 ZMO0483 ZMO0675 ZMO0759 ZMO0760 ZMO0763 ZMO0774 ZMO0846 ZMO0975 ZMO1067 ZMO1136 ZMO1221 ZMO1252 ZMO1253 ZMO1255 ZMO1389 ZMO1391 ZMO1404 ZMO1429 ZMO1430 ZMO1431 ZMO1432 ZMO1490 ZMO1520 ZMO1541 ZMO1590 ZMO1659 ZMO1715 ZMO1722 ZMO1723 ZMO1874 ZMO1875
S_ORR 20% (#367)
S_ORR 15% (#366)
S_ORR 20% (#128)
S_FRF 25% (#365)
S_FRF 20% (#364)
S_FRF 20% (#129)
S_BTN 25% (#369)
S_BTN 20% (#368)
S_BTN 20% (#130)
S_BTN 20% (#119)
M_ORR 10% (#373)
M_ORR 8% (#372)
M_ORR 10% (#126)
M_ORR 20% (#116)
M_HAV 25% (#371)
M_HAV 20% (#370)
M_HAV 20% (#127)
M_HAV 20% (#115)
M_FRF 20% (#118)
M_BTN 20% (#375)
M_BTN 14% (#374)
M_BTN 14% (#125)
M_BTN 20% (#117)
M_batch2 6% (#390)
M_batch2 4% (#389)
M_batch1 20% (#378)
M_batch1 15% (#377)
M_batch1 10% (#376)
M_batch1 10% (#307)
M_batch1 10% (#305)
M_batch1 10% (#124)
M_batch1 6% (#122)
M_batch1 8% (#121)
M_batch1 10% (#114)
M_batch1 10% (#112)
M_batch1 8% (#111)
M_batch1 6% (#110)
SYN-10 55% (#506)
SYN-10 50% (#505)
SYN-10 50% (#346)
SYN-10 40% (#345)
SYN-37 40% (#309)
SYN-37 25% (#308)
SYN-37 4% (#306)
SYN-37 20% (#304)
DMSO (#164)
DMSO (#158)
Rich (#99)
Rich (#97)
Rich (#96)
Rich (#93)
Rich (#89)
Rich (#88)
Rich (#84)
Rich (#78)
Rich (#72)
Rich (#277)
Rich (#272)
Rich (#263)
Rich (#255)
Rich (#250)
Rich (#248)
Rich (#246)
Rich (#235)
Rich (#223)
Rich (#211)
Rich (#163)
Rich (#140)
Rich (#131)
Rich (#120)
Rich (#109)
37 hydrolysate experiments
Synthetic hydrolysate experiments
DMSO
Rich media (no inhibitor)

Figure S8

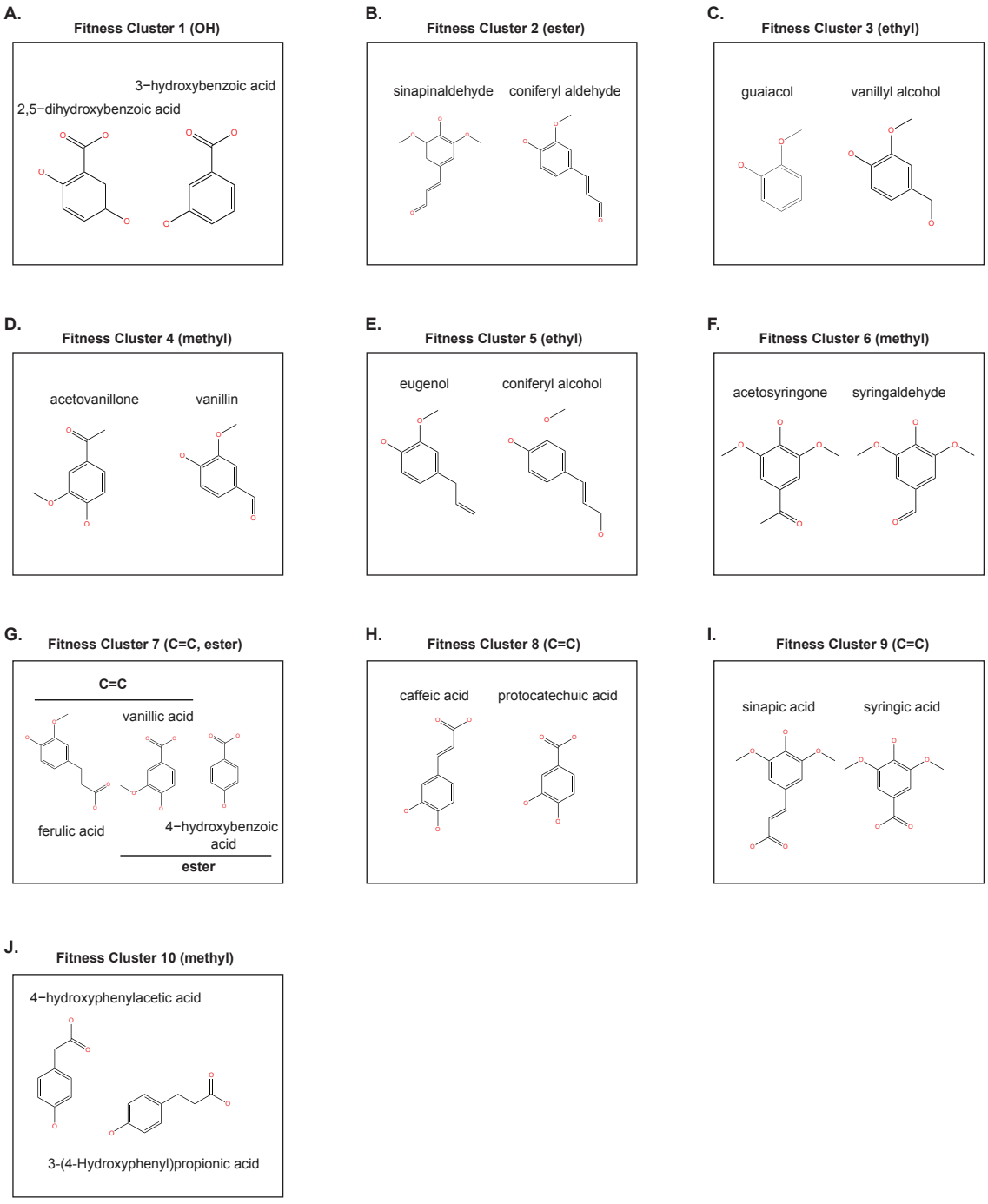

A.
Fitness Cluster 1 (OH)
3−hydroxybenzoic acid
2,5−dihydroxybenzoic acid
B.
Fitness Cluster 2 (ester)
sinapinaldehyde
coniferyl aldehyde
C.
Fitness Cluster 3 (ethyl)
guaiacol
vanillyl alcohol
D.
Fitness Cluster 4 (methyl)
acetovanillone
vanillin
E.
Fitness Cluster 5 (ethyl)
eugenol
coniferyl alcohol
F.
Fitness Cluster 6 (methyl)
acetosyringone
syringaldehyde
G.
Fitness Cluster 7 (C=C, ester)
C=C
vanillic acid
ferulic acid
4−hydroxybenzoic acid
ester
H.
Fitness Cluster 8 (C=C)
caffeic acid
protocatechuic acid
I.
Fitness Cluster 9 (C=C)
sinapic acid
syringic acid
J.
Fitness Cluster 10 (methyl)
4−hydroxyphenylacetic acid
3-(4-Hydroxyphenyl)propionic acid

Figure S9

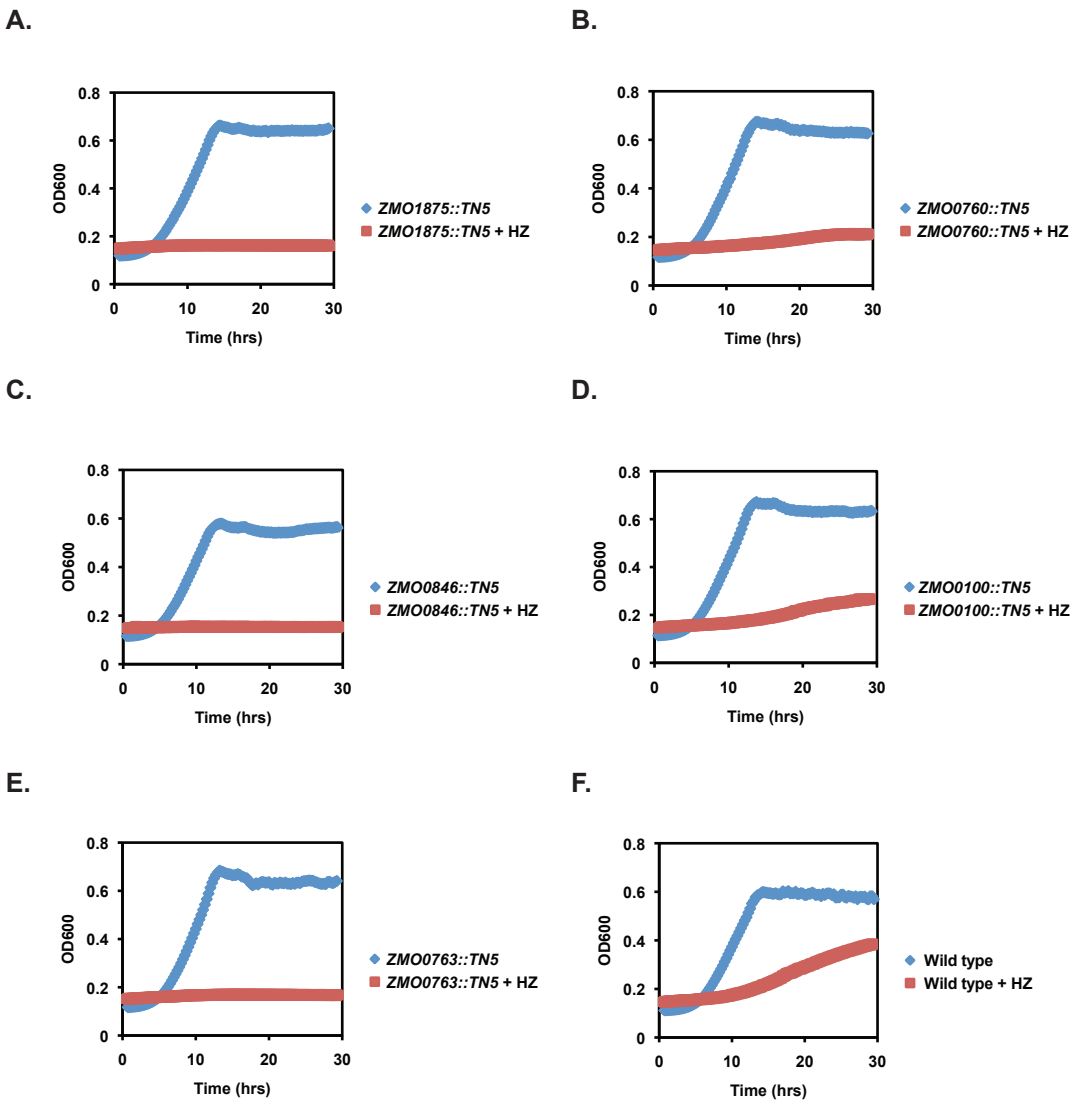

A.
B.
C.
D.
E.
F.
OD600
Time (hrs)
0
0.2
0.4
0.6
0.8
10
20
30
ZMO1875::TN5
ZMO1875::TN5 + HZ
ZMO0760::TN5
ZMO0760::TN5 + HZ
ZMO0846::TN5
ZMO0846::TN5 + HZ
ZMO0100::TN5
ZMO0100::TN5 + HZ
ZMO0763::TN5
ZMO0763::TN5 + HZ
Wild type
Wild type + HZ

Figure S10

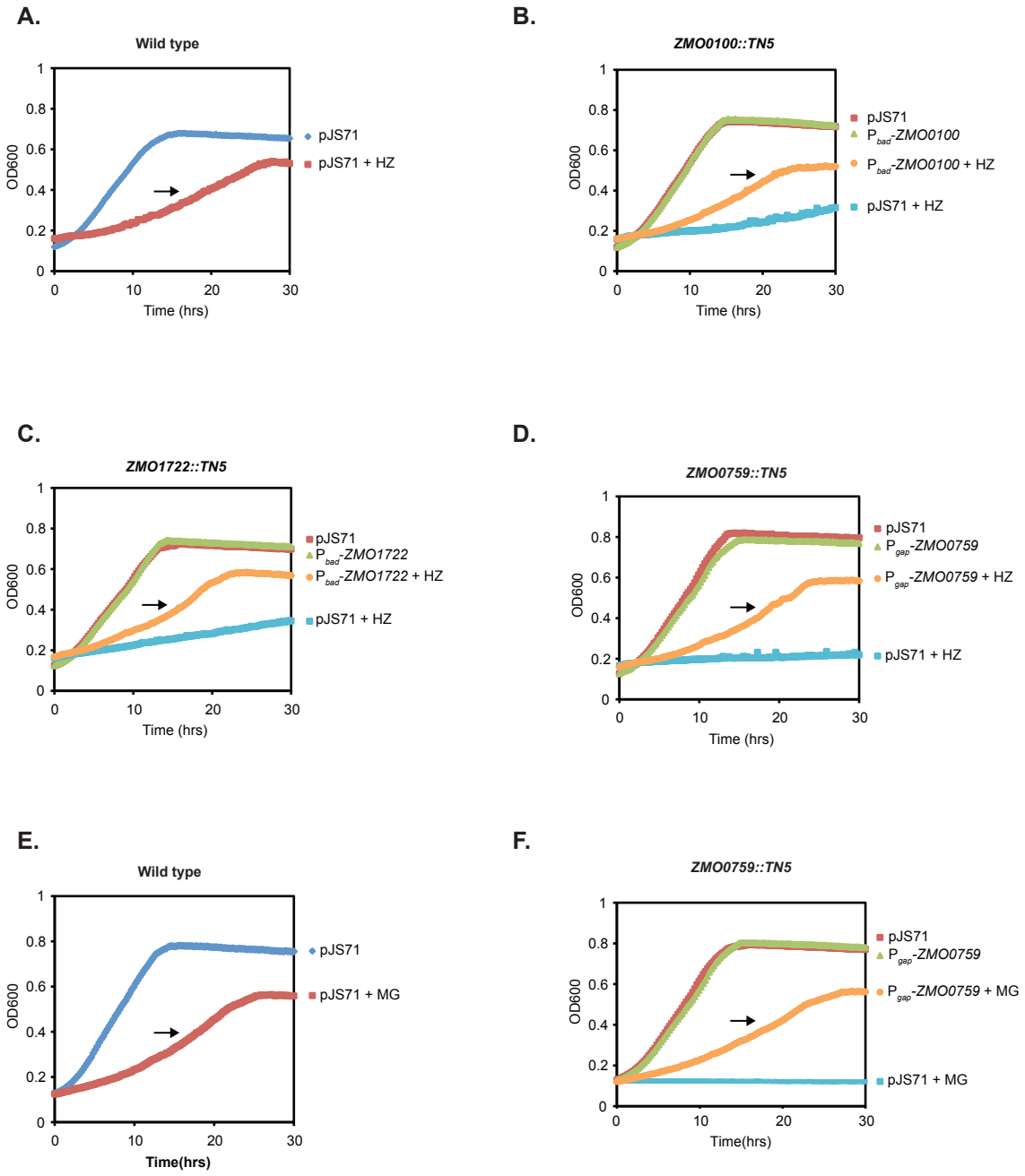
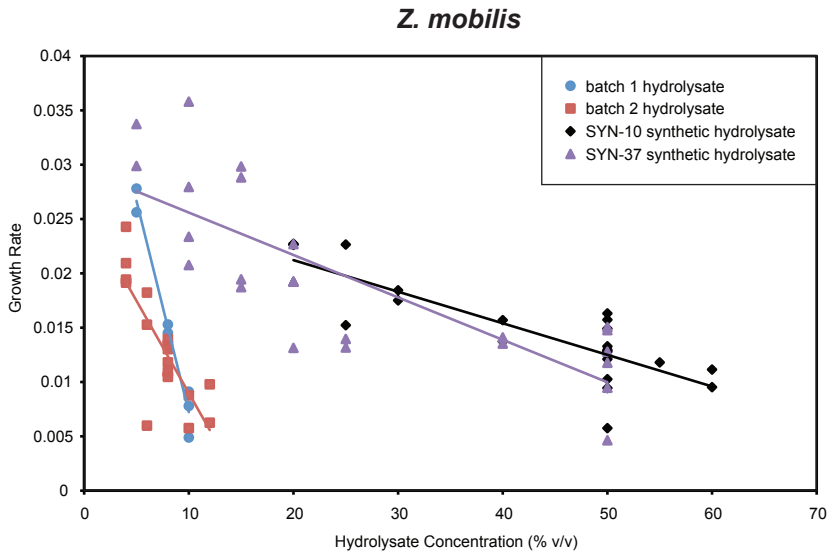

A.
Wild type
OD600
Time (hrs)
pJS71
pJS71 + HZ
B.
ZMO0100::TN5
pJS71
Pbad-ZMO0100
Pbad-ZMO0100 + HZ
pJS71 + HZ
C.
ZMO1722::TN5
pJS71
Pbad-ZMO1722
Pbad-ZMO1722 + HZ
pJS71 + HZ
D.
ZMO0759::TN5
pJS71
Pgap-ZMO0759
Pgap-ZMO0759 + HZ
pJS71 + HZ
E.
Wild type
pJS71
pJS71 + MG
Time(hrs)
F.
ZMO0759::TN5
pJS71
Pgap-ZMO0759
Pgap-ZMO0759 + MG
pJS71 + MG

Figure S11

A.



B.

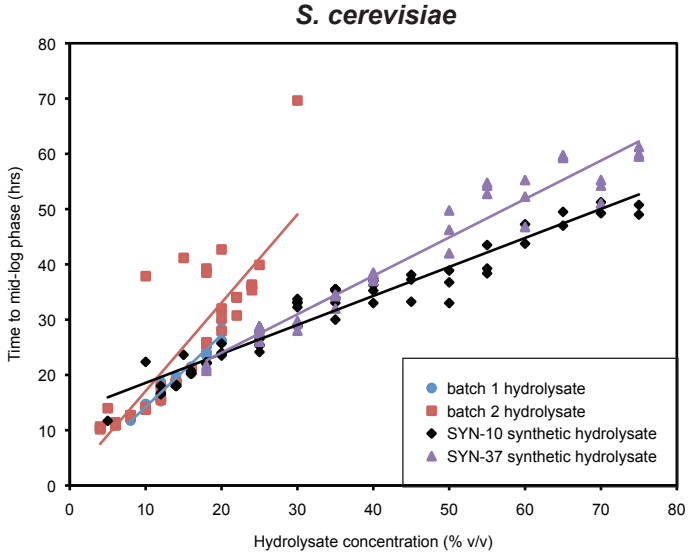


Figure S12

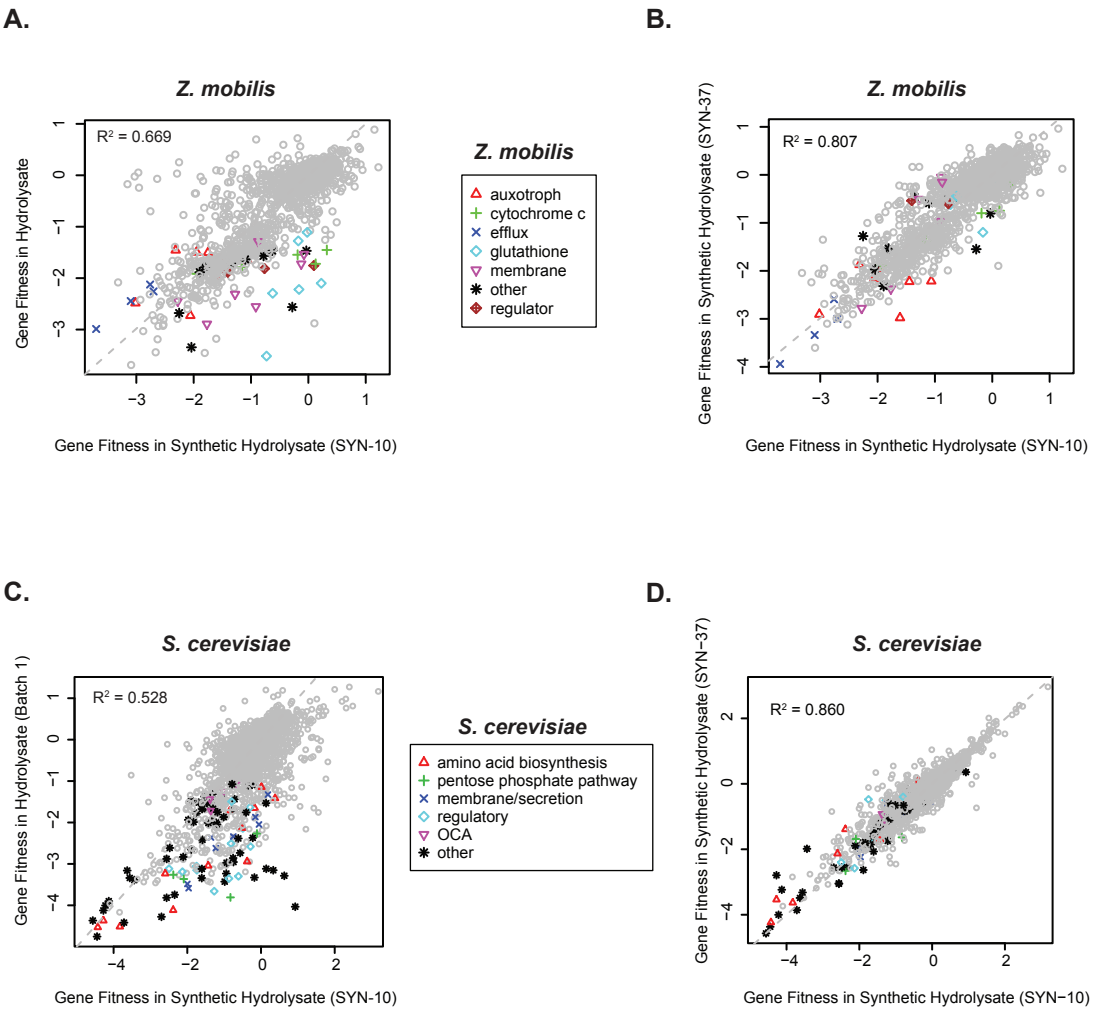

A.
Z. mobilis
R² = 0.669
Gene Fitness in Hydrolysate
Gene Fitness in Synthetic Hydrolysate (SYN-10)
Z. mobilis
auxotroph
cytochrome c
efflux
glutathione
membrane
other
regulator
B.
Z. mobilis
R² = 0.807
Gene Fitness in Synthetic Hydrolysate (SYN-37)
Gene Fitness in Synthetic Hydrolysate (SYN-10)
C.
S. cerevisiae
R² = 0.528
Gene Fitness in Hydrolysate (Batch 1)
Gene Fitness in Synthetic Hydrolysate (SYN-10)
S. cerevisiae
amino acid biosynthesis
pentose phosphate pathway
membrane/secretion
regulatory
OCA
other
D.
S. cerevisiae
R² = 0.860
Gene Fitness in Synthetic Hydrolysate (SYN−37)
Gene Fitness in Synthetic Hydrolysate (SYN−10)

Figure S13

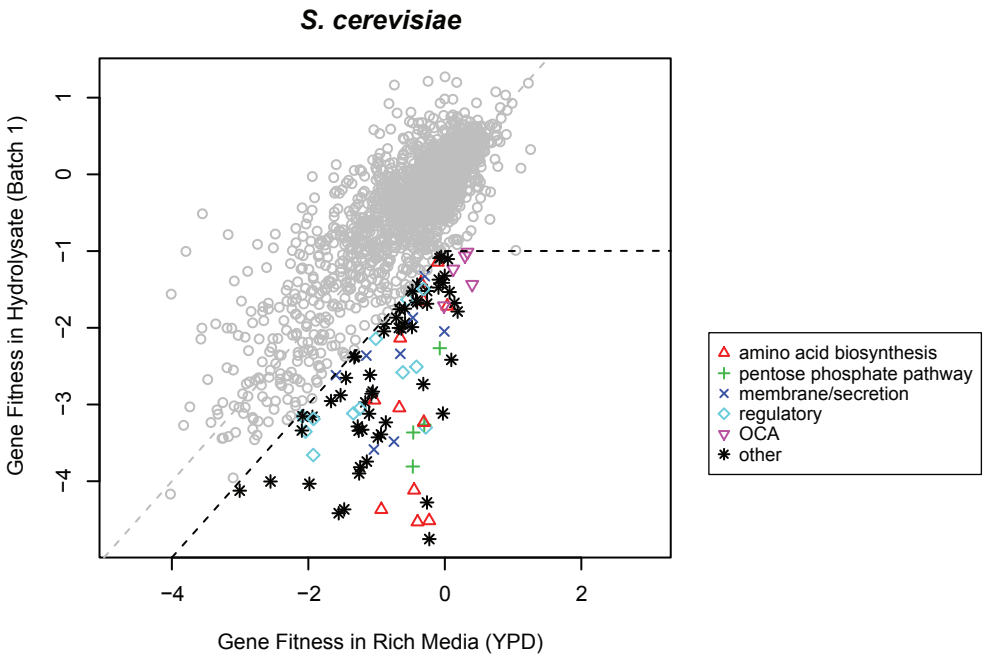

S. cerevisiae
Gene Fitness in Hydrolysate (Batch 1)
Gene Fitness in Rich Media (YPD)
1
0
−1
−2
−3
−4
−4
−2
0
2
amino acid biosynthesis
pentose phosphate pathway
membrane/secretion
regulatory
OCA
other

Figure S14

A.

B.

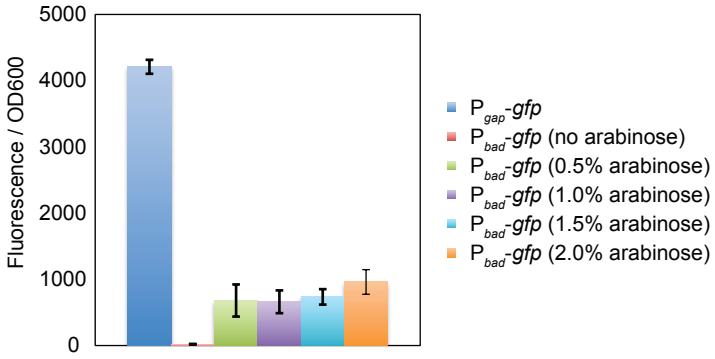


Figure S15

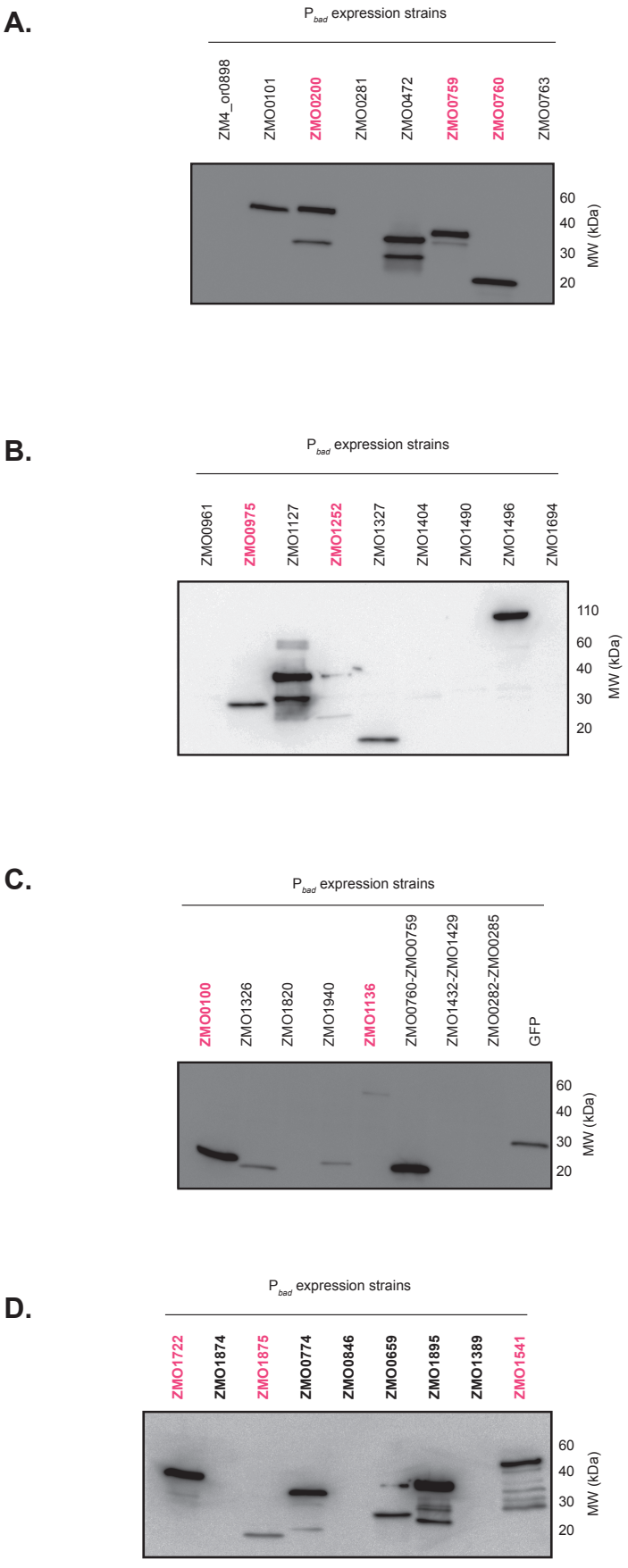
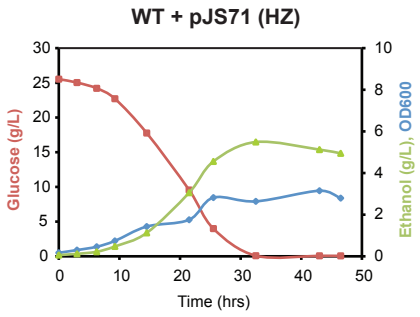
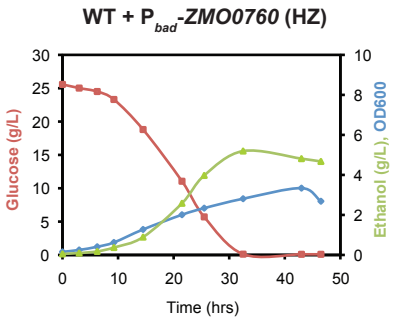

A.
P bad expression strains
ZM4_or0898
ZMO0101
ZMO0200
ZMO0281
ZMO0472
ZMO0759
ZMO0760
ZMO0763
60
40
30
20
MW (kDa)
B.
P bad expression strains
ZMO0961
ZMO0975
ZMO1127
ZMO1252
ZMO1327
ZMO1404
ZMO1490
ZMO1496
ZMO1694
110
60
40
30
20
MW (kDa)
C.
P bad expression strains
ZMO0100
ZMO1326
ZMO1820
ZMO1940
ZMO1136
ZMO0760-ZMO0759
ZMO1432-ZMO1429
ZMO0282-ZMO0285
GFP
60
40
30
20
MW (kDa)
D.
P bad expression strains
ZMO1722
ZMO1874
ZMO1875
ZMO0774
ZMO0846
ZMO0659
ZMO1895
ZMO1389
ZMO1541
60
40
30
20
MW (kDa)

Figure S16

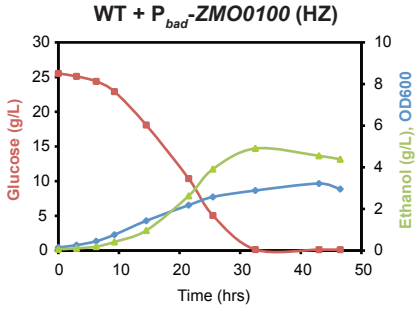
A.



C.



B.



D.

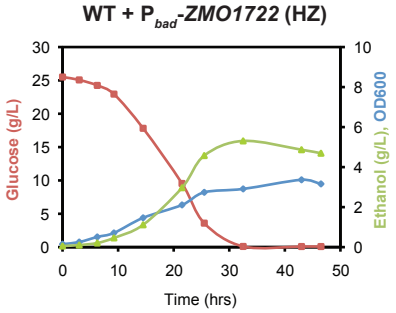


Figure S17

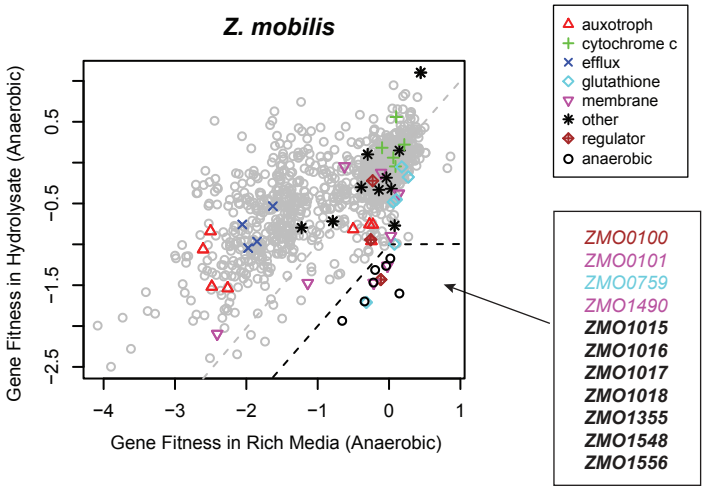
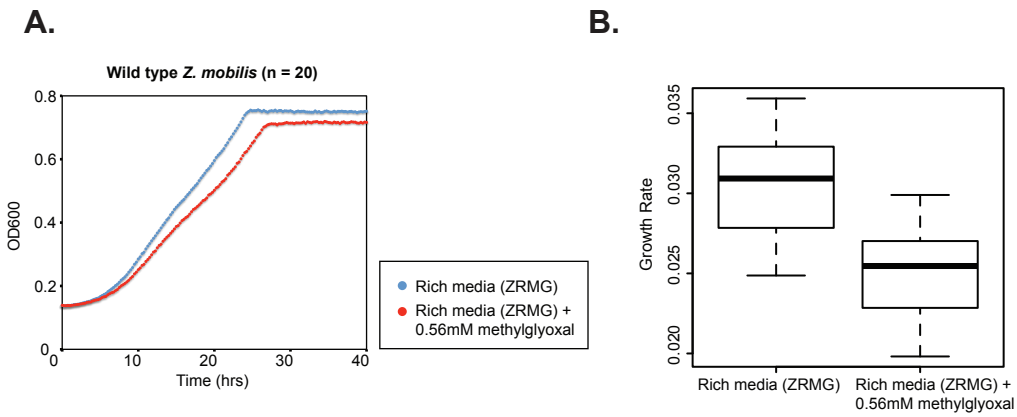

Z. mobilis
Gene Fitness in Hydrolysate (Anaerobic)
0.5
−0.5
−1.5
−2.5
−4
−3
−2
−1
0
1
Gene Fitness in Rich Media (Anaerobic)
auxotroph
cytochrome c
efflux
glutathione
membrane
other
regulator
anaerobic
ZMO0100
ZMO0101
ZMO0759
ZMO1490
ZMO1015
ZMO1016
ZMO1017
ZMO1018
ZMO1355
ZMO1548
ZMO1556

Figure S18


A.
Wild type Z. mobilis (n = 20)
OD600
Time (hrs)
Rich media (ZRMG)
Rich media (ZRMG) + 0.56mM methylglyoxal
B.
Growth Rate
Rich media (ZRMG)
Rich media (ZRMG) + 0.56mM methylglyoxal