Systems approach to refining genome annotation

Jennifer L. Reed*, Trina R. Patel*, Keri H. Chen*, Andrew R. Joyce*†, Margaret K. Applebee‡, Christopher D. Herring*, Olivia T. Bui*, Eric M. Knight*, Stephen S. Fong§, and Bernhard O. Palsson*¶

Departments of *Bioengineering and ‡Chemistry and Biochemistry, and †Program in Bioinformatics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412; and §Department of Chemical and Life Science Engineering, Virginia Commonwealth University, P.O. Box 843028, Richmond, VA 23284-3028

Edited by David S. Eisenberg, University of California, Los Angeles, CA, and approved September 28, 2006 (received for review April 25, 2006)

Genome-scale models of *Escherichia coli* K-12 MG1655 metabolism have been able to predict growth phenotypes in most, but not all, defined growth environments. Here we introduce the use of an optimization-based algorithm that predicts the missing reactions that are required to reconcile computation and experiment when they disagree. The computer-generated hypotheses for missing reactions were verified experimentally in five cases, leading to the functional assignment of eight ORFs (*yjjLMN*, *yeaTU*, *dctA*, *idnT*, and *putP*) with two new enzymatic activities and four transport functions. This study thus demonstrates the use of systems analysis to discover metabolic and transport functions and their genetic basis by a combination of experimental and computational approaches.

constraint-based | flux balance analysis | functional genomics | metabolic reconstruction | systems biology

Current genome annotations include a substantial fraction of ORFs with unknown function (1, 2); methods are needed to provide insight into the possible function of these genes, without the need for screening individual gene products across a multitude of possible activities. Metabolic and regulatory networks are reconstructed from genome annotations and scientific literature to integrate and represent our current knowledge of network components and interactions (3). Dual-perturbation methods have been developed to study regulatory networks (4), which can be used to reconcile model predictions and experimental data, thus leading to possible iterative model refinements and experimentally testable hypotheses (4, 5). Iterative model-building can be systematized through the use of computational algorithms (6). Such an approach is presented here, and it consists of four steps. First, computational analysis identified discrepancies between model predictions and growth phenotyping data by using a reconstructed genome-scale *Escherichia coli* metabolic network. Second, an algorithm then identified enzymatic and transport reactions that likely were missing from the current metabolic reconstruction that could reconcile model predictions and experimental observations. Third, ORFs that might be responsible for these missing activities then were identified by using literature searches, sequence-homology searches (7), context-based homology methods (8, 9), and in some cases unpublished microarray data. Fourth, experimental verification of the algorithm's predictions then were carried out by evaluating growth phenotypes of single-deletion strains available in the Keio collection (10) and gene-expression measurements. Here we present a comprehensive combined computational and experimental approach to analyze phenotypic data and genome annotation information in a global manner to uncover individual ORF function.

Results

Growth phenotyping data (11), available from Biolog (Hayward CA; www.biolog.com), were used to identify missing reactions from the reconstructed genome-scale metabolic network of *E. coli* MG1655 (*i*JR904) (12). Using a flux balance model of *E. coli*, we identified 54 minimal medium conditions with different carbon or nitrogen sources that support growth experimentally in the Biolog data but for which the reconstructed metabolic network could not explain a growth state. Four of these discrepancies (D-allose, glucose 1-phosphate, fructose 1-phosphate, and L-cysteine) could be reconciled with available literature for *E. coli* metabolism. For the remaining discrepancies, we then used an algorithm (step two in the procedure described above; for details see *Materials and Methods*) to identify the minimal number of reactions [from a universal database of known metabolic transformations (13, 14)] that needed to be added to the reconstructed network so that cellular growth could be computed by the model (Fig. 1). Additional solutions also were found, which could have as many or more missing reactions as prior solutions, by repeating the algorithm and excluding previous solutions.

The algorithm identified metabolic and/or transport activities whose addition would allow the reconstructed network to reproduce growth in 26 of the 50 identified environments (see Table 2, which is published as supporting information on the PNAS web site, for details). The algorithm found 10 or more solutions for 14 of these 26 cases and a single unique solution for 11 of these 26 cases (for D-malate, there were two possible solutions; see text below). Eight of the 26 cases only require a transporter for the metabolite of interest, indicating that the enzymatic transformations needed are already present in the reconstruction.

A subset of the 26 minimal medium conditions was chosen for further experimentation to verify the existence of the predicted missing transformations and to identify the genes responsible for them. Fourteen minimal medium conditions were initially screened, but we were unable to confirm growth in five of the minimal medium conditions (xanthine and xanthosine as nitrogen sources and lyxose, α-ketobutyrate, and α-hydroxybutyrate as carbon sources). *E. coli* grew slowly on agmatine and citrulline as nitrogen sources and showed a 3-day lag period when grown on glyoxylate as a carbon source (data not shown). We describe the remaining six cases in more detail, and the results are summarized in Table 1.

Transport Activities. *In silico* growth of *E. coli* on propionate and 5-keto-D-gluconate requires the addition of a transporter for these metabolites to the reconstructed metabolic network because there are currently no existing annotated transporters for these compounds in *E. coli*. Eight genes with predicted or known transport

Author contributions: J.L.R., T.R.P., and B.O.P. designed research; J.L.R., T.R.P., K.H.C., A.R.J., M.K.A., C.D.H., O.T.B., E.M.K., and S.S.F. performed research; J.L.R., T.R.P., A.R.J., and M.K.A. analyzed data; and J.L.R. and B.O.P. wrote the paper.

Conflict of interest statement: B.O.P. and the University of California at San Diego have a financial interest in Genomatica, Inc. Although this grant has been identified for conflict of interest management based on the overall scope of the project and its potential to benefit Genomatica, Inc., the research findings included in this publication may not necessarily relate directly to the interests of Genomatica, Inc.

This article is a PNAS direct submission.

Abbreviation: μ, growth rate.

¶To whom correspondence should be addressed at: Department of Bioengineering, University of California at San Diego, 9500 Gilman Drive, Mail Code 0412, La Jolla, CA 92093-0412. E-mail: bpalsson@bioeng.ucsd.edu.

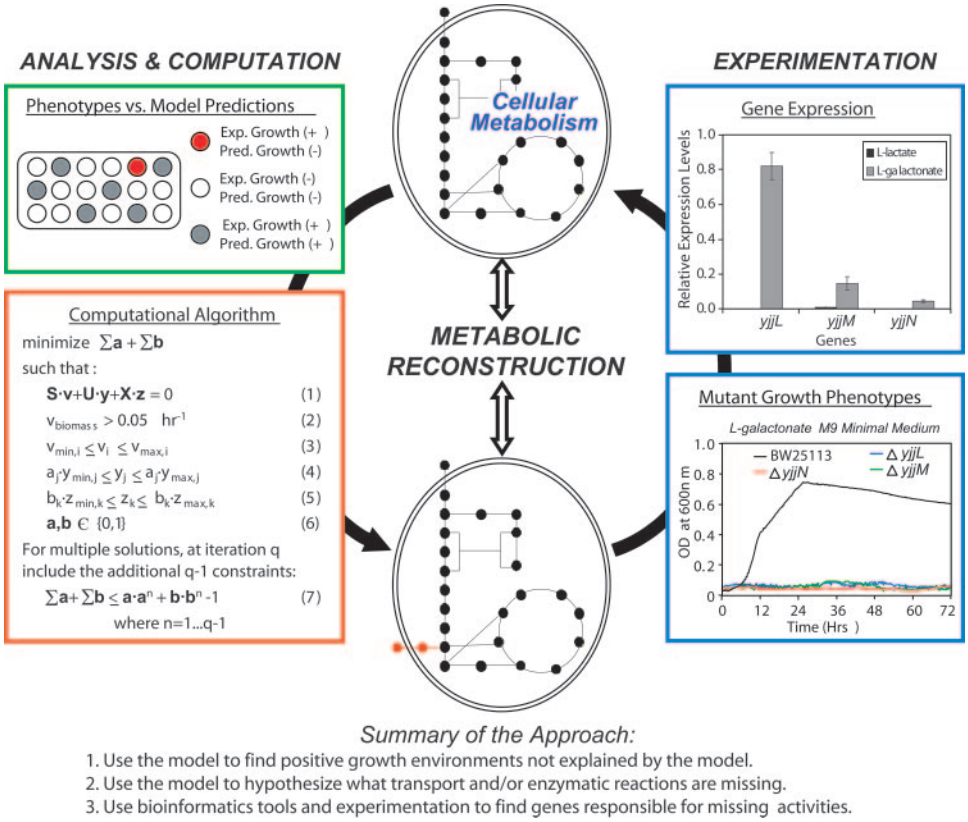

Fig. 1. Systematic procedure for identifying missing reactions and ORF assignments. With a metabolic network reconstruction, growth predictions can be made and compared with experimental growth phenotypes (green box). For medium conditions where the microorganism grows and the model does not predict growth (red wells), reactions from a universal database can be identified by an optimization algorithm (red box) that, if added to the current reconstruction, allow for growth predictions by the model (see *Materials and Methods* and *Supporting Text* for algorithm details). Experimental testing of mutant strains can be used to identify genes responsible for the added enzymatic activities, and subsequent studies of gene-expression and biochemical characterizations can be used to further support conclusions (blue boxes). Data shown in the blue boxes is for the genes identified as being involved in L-galactonate catabolism: *yjjL* (putative transporter), *yjjM* (putative regulator), and *yjjN* (putative oxidoreductase). Reported gene-expression levels are relative to the internal control gene, *acpP*.

activities were investigated by evaluating growth phenotypes of single-gene deletion strains (see Table 3, which is published as supporting information on the PNAS web site, for list of mutants screened). Of eight single-gene deletion mutants screened, only the following showed reduced growth rates: Δ*putP* on propionate and Δ*idnT* on 5-keto-D-gluconate (see Fig. 3, which is published as

Table 1. Summary of experimental results for growth on various carbon sources

				RT-PCR*	
Carbon source	Missing functions	Genes	Mutant phenotypes	Control	Carbon source
L-galactonate	Transporter	*yjjL*	Lethal	8.9×10^{-5}	0.82
	Oxidoreductase	*yjjN*	Lag,† μ	1.1×10^{-5}	4.4×10^{-2}
	Regulator	*yjjM*	Lethal	9.4×10^{-3}	0.14
D-malate	Transporter	*dctA*	Lag, μ	0.12	0.62
	Dehydrogenase	*yeaU*	Lethal	1.9×10^{-3}	1.1
	Regulator	*yeaT*	Lag,† μ	6.5×10^{-3}	1.5
Propionate	Transporter	*putP*	μ	0.13	0.32
5-keto-D-gluconate	Transporter	*idnT*	μ	3.0×10^{-2}	7.51
Thymidine	Transporter	*tolC*	Yx/s, μ	3.6×10^{-2}	5.0×10^{-2}

Lag, no growth was observed for a period of 2 days; lethal, no growth was observed for a period of 3 days; Yx/s, overall biomass yields per substrate were substantially reduced; μ, growth rates were substantially reduced (see text for reported values).

*RT-PCR levels are the mean relative expression levels to the internal control gene *acpP* for the control condition (see *Materials and Methods* for details) and the respective carbon source condition. See Table 4 for the means and standard deviations.

†For Δ*yjjN* with L-galactonate and Δ*yeaT* with D-malate as a carbon source, no growth was observed in the Bioscreen C analysis, but growth was observed after 2 days in Erlenmeyer flasks, which is likely attributable to aeration differences within the different culture conditions.

MICROBIOLOGY

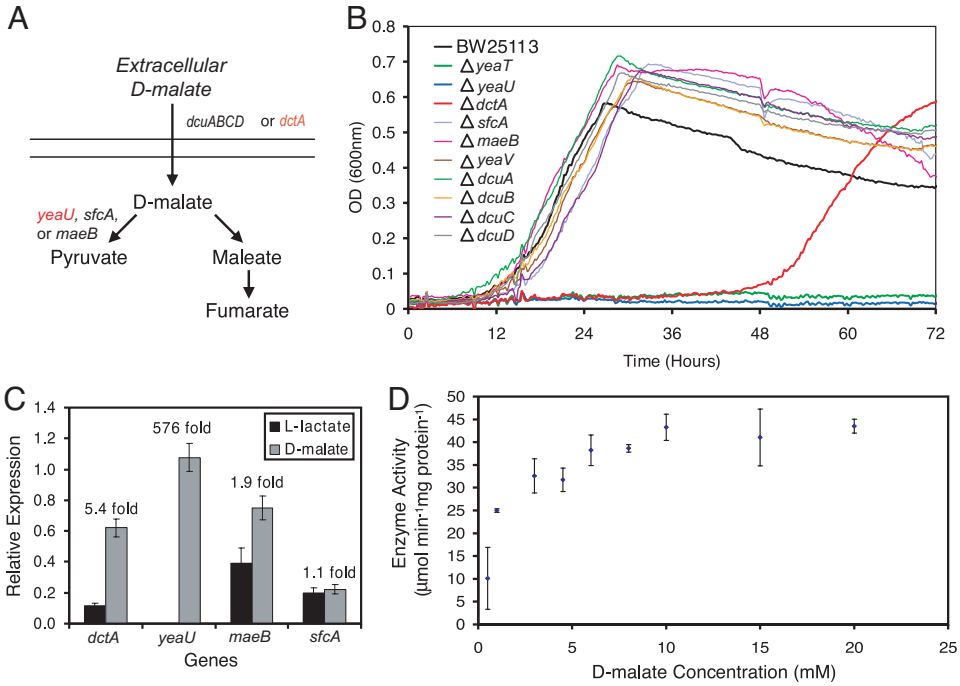

Fig. 2. Utilization of D-malate as a sole carbon source. (*A*) The two pathways that are predicted by the algorithm and the corresponding genes whose mutants were screened for altered phenotypes (red genes indicate that corresponding mutant strains showed altered growth). (*B*) Growth phenotype screens for the parental strain (BW25113), three mutants with altered phenotypes (Δ*yeaU*, Δ*dctA*, *and* Δ*yeaT*), and other screened mutants for growth on D-malate. (*C*) Expression levels were determined for BW25113 grown on L-lactate and D-malate by using RT-PCR. The values reported are the expression levels relative to the internal control gene *acpP* and the fold expression changes (D-malate to L-lactate) across the two conditions. (*D*) Results from kinetic assays of overexpressed YeaU protein show enzymatic activity levels as a function of increasing concentrations of D-malate.

supporting information on the PNAS web site). Additionally, both *putP* and *idnT* show increased expression levels when grown on propionate and 5-keto-D-gluconate minimal medium, respectively (Table 1). Although IdnT is a known D-gluconate and L-idonate transporter, a prior study has shown that 5-keto-D-gluconate inhibits D-gluconate transport (15). PutP is a proline transporter that belongs to the same family of transporters as a propionate transporter (MctP) from *Rhizobium leguminosarum* (16), and the deletion of *putP* reduces the growth rate (μ) on propionate ($\mu = 0.02 \pm 0.002$ h^{-1} for Δ*putP* as compared with 0.06 ± 0.002 h^{-1} for the parental strain). This combined evidence suggests that *putP* and *idnT* are responsible for the transport of propionate and 5-keto-D-gluconate, respectively.

Growth on Thymidine. The Biolog experimental data indicates that *E. coli* also can grow on thymidine as a sole carbon source but not as a nitrogen source. The *i*JR904 metabolic network includes a reaction cleaving thymidine into deoxy-ribose-1-phosphate and thymine. Although the sugar can be converted into central metabolic intermediates, the network has no way of dissipating thymine, resulting in a no growth prediction. Accordingly, the simplest solution identified by the algorithm calls for the secretion of thymine (some of the other solutions use the thymine reductive pathway coupled with secretion of pathway intermediates). HPLC analysis of the batch culture medium was in agreement with this hypothesis, because thymidine consumption was accompanied by corresponding thymine secretion into the growth medium (see Fig. 4, which is published as supporting information on the PNAS web site). A recent study has shown that *E. coli* strains with *ntrB*(Con) mutations can use thymine as a nitrogen source at room temperature (but not at 37°C and not in the parental strain) (17). The HPLC results are in agreement with this study because we only observed secretion of thymine in cells grown at 37°C, which lack the *ntrB*(Con) mutation. Only 1 mutant (Δ*tolC*), of the 29 screened mutants, exhibited a substantially reduced growth rate and biomass yield, but the mutant still secreted thymine [an independent screen of the Δ*tolC* mutant found it had defective growth on 19 of 38 evaluated carbon sources (18)]. It remains unclear which transporter is responsible for transport of thymine across the inner membrane and if *tolC* plays a role, but our data indicate that a secretion mechanism must exist for thymine.

D-Malate Catabolism. Transport and enzymatic reaction(s) are needed to reconcile D-malate catabolism; the computational algorithm developed here identified the decarboxylation of D-malate into pyruvate as the simplest explanation (Fig. 2*A*). Three mutants (Δ*dctA*, Δ*yeaT*, and Δ*yeaU*) showed altered growth on D-malate. The Δ*dctA* mutant was only able to grow after a 2-day lag period ($\mu = 0.23 \pm 0.006$ h^{-1} for Δ*dctA* as compared with 0.35 ± 0.006 h^{-1} for the parental strain), whereas the Δ*yeaU* and Δ*yeaT* mutants showed no growth over 3 days (Fig. 2*B*) in the Bioscreen C (a 200-well growth-monitoring system). The Δ*yeaT* mutant was able to grow after 2 days in Erlenmeyer flasks but at a much slower growth rate ($\mu = 0.01 \pm 0.001$ h^{-1}, based on two replicates). The eventual growth of the Δ*dctA* mutant indicates that another transporter eventually compensates for the *dctA* deletion (Fig. 2); in fact, *dcuA* and *dcuB* show increased expression levels in the Δ*dctA* strain grown on D-malate (see Table 4, which is published as supporting information on the PNAS web site). In addition to mutant phenotyping data, both *dctA* and *yeaU* show increased expression levels during growth on D-malate as compared with L-lactate (Fig. 2*C*), providing further evidence for their involvement in D-malate catabolism. DctA is one of four dicarboxylic acid transporters that are known to transport L-malate. Based on the combined evidence provided by mutant growth phenotypes and gene-expression data, *dctA* appears to be responsible for the transport of D-malate as well. YeaU has significant homology

with tartrate dehydrogenase from *Pseudomonas putida*, which has been shown to oxidize D-malate into pyruvate (19). Biochemical assays of the partially purified *E. coli* YeaU protein showed that the enzyme does oxidize D-malate by using NAD as an electron acceptor (Fig. 2*D*). YeaT is annotated as a putative regulator and, based on its chromosomal location and lethal phenotype, it appears to positively regulate *yeaU*. This conclusion was further supported by RT-PCR measurements of DNA isolated by chromatin immunoprecipitation (ChIP) from a *yeaT*-8x myc-tagged strain grown on D-malate, which shows enrichment for the upstream region of *yeaU* (see Table 5, which is published as supporting information on the PNAS web site). Our experimental data thus shows that D-malate catabolism involves the *dctA*, *yeaT*, and *yeaU* gene products.

Galactonate γ-Lactonate and Galactonate Catabolism. The Biolog data indicates that both D- and L-galactonate γ-lactone can support growth as sole carbon sources; however, the hydration of galactonate γ-lactone into galactonate occurs nonenzymatically (at pH 7; data not shown), making it difficult to assess growth solely on galactonate γ-lactone. D-Galactonate catabolism has been well characterized in the literature (20) and already is represented in the metabolic reconstruction. Currently, L-galactonate is not present in the KEGG database (14), so the algorithm in its current implementation cannot resolve the observed phenotype; however, the catabolism of L-galactonate has been reported in *E. coli*, and the pathway has been identified (21). The genes responsible for the transport and oxidation of L-galactonate have not been found and subsequently were targeted in this study.

From Affymetrix gene-expression data of MG1655 grown on L-galactonate (data not shown), we were able to identify two genes (*yjjL* and *yjjN*) that were strongly up-regulated compared with growth on other carbon sources. Knockout mutants of these two genes and the putative regulator (*yjjM*) located between these loci were unable to grow on L-galactonate in a Bioscreen C analysis. In batch growth experiments conducted in flasks, the Δ*yjjL* and Δ*yjjM* strains still were unable to grow; however, the Δ*yjjN* strain did grow after a 2-day lag phase but at a slower growth rate than the parental strain ($\mu = 0.36 \pm 0.02$ h^{-1} for BW25113 and 0.17 ± 0.01 h^{-1} for Δ*yjjN*). Subsequent analysis of expression levels for all three genes by using RT-PCR showed that their expression is strongly up-regulated on L-galactonate as compared with L-lactate (data shown in Fig. 1 and Table 1). From these results, we conclude that the putative transporter (*yjjL*) and putative oxidoreductase (*yjjN*) are likely to be responsible for the missing transport and enzymatic (L-galactonate oxidoreductase) activities and that *yjjM* regulates their gene expression.

Discussion

The results from this study demonstrate that the synergy between experimentation and *in silico* modeling allows for the discovery of the roles individual genes play in an organism's behavior. In this study, a metabolic model was used to evaluate growth phenotypes and when discrepancies were found the model was used to generate testable hypotheses concerning the presence of metabolic and transport functions. The algorithm presented here can be used to identify missing metabolic and transport reactions based on growth phenotyping data, which, as shown here, can subsequently lead to the identification of gene functions. This algorithmic method can be used to direct the search for enzymes and transporters within a genome, and a variety of experimental methods can be used to find and confirm the roles of the responsible genes (7–9, 22). Phenotypic screens of single-deletion mutants have been carried out (18, 23), and these will be important for narrowing down likely gene candidates as well as pointing out alternate metabolic routes [requiring the identification of synthetic lethal gene interactions (24)]. Because predictions could be made for only 26 of the 50 cases in this study, a limitation to the computational approach in its current implementation is that the missing reactions must belong to the universal database. As such, undiscovered metabolic reactions will not be identified; however, the inclusion of computational methods that calculate novel and thermodynamically feasible metabolic transformations (25) will expand the scope of this study.

We have shown that the presented approach (combining data analysis, hypothesis generation, and experimental testing) can be used to systematically discover missing gene functions in microbial organisms. Although these efforts have centered on *E. coli*, the approach is likely to be more useful for less characterized organisms where new growth phenotypes are observed and many more genes with unknown functions are present.

Materials and Methods

Computational Methods. For each minimal medium growth condition (supplemented with a carbon or nitrogen source) the algorithm calculates the minimum number of reactions that need to be added from **U** (matrix of known metabolic reactions) and **X** (a matrix of exchange reactions) to allow for growth. The optimization objective and constraints are detailed in Fig. 1 (and discussed in greater depth in *Supporting Text*, which is published as supporting information on the PNAS web site). The computational algorithm is a hybrid of those developed for calculating the minimum number of reactions required for cellular growth and for designing strains for metabolic engineering purposes (13, 26). The matrices **S**, **U**, and **X**, respectively, contain the stoichiometric coefficients for the reactions in the current metabolic reconstruction (*i*JR904) (12), the universal database (13) [derived from KEGG (14)], and the exchange reactions allowing intracellular metabolites to be added to or depleted from the system. The two binary vectors, **a** and **b**, will have only nonzero entries if the corresponding reaction(s) need to be added to the network to enable growth from **U** and **X**, respectively. The vectors **v**, **y**, and **z** contain the fluxes through the *i*JR904, universal database, and exchange reactions, respectively. The algorithm was used iteratively, with each iteration excluding previous solutions, to identify up to 15 possible solutions (13, 26). Mutants were identified for phenotypic screens based on literature searches, BLAST sequence-homology searches, context-based homology searches [via the Prolinks database (9)], available Biolog results for *E. coli* mutants [ASAP database (23)], and unpublished gene-expression microarray data (for L-galactonate).

Strains and Culture Conditions. The Keio collection of in-frame single-gene deletion strains (10) was the source for nearly all of the mutant strains used in this study. The individual strains from the collection were obtained from Andrei Osterman at The Burnham Institute (La Jolla, CA). The parental strain for all of the mutants was BW25113 (27) and was obtained from the *E. coli* Genetic Stock Collection (New Haven, CT). A Δ*yjjM* mutant was generated and used in this study because the Keio collection mutant lacks the predicted promoter for *yjjN* (28). To generate the Δ*yjjM* mutant, the same method used to make the knockout collection (27) was used but with the following primers: 5′-TTAATGAGCATAACGCGTGTTCTCATTAATGGATCGAATCATTGATTGTTTAGCTGAATTATTCCGGGGATCCGTCGACC-3′ and 5′-ATGTACAACATCAGCCGCACCACAGTGCGTCATATTCTCAGCCACTTACGCGAATGCGCGGTGTAGGCTGGAGCTGCTTC-3′. All mutants for genes listed in Table 1 were confirmed with PCR. A *yeaT*-8x myc-tagged strain used for ChIP experiments was generated from the BW25113 parental strain by using pBOP508 as described previously (29).

All strains were screened for growth by using the Bioscreen C platereader system (Growth Curves USA, Piscataway, NJ) by monitoring optical density at 600 nm over 72 h. For strains that showed altered phenotypes relative to the parental strain (such as lethal phenotypes, prolonged lag periods, reduced biomass

yields, or reduced growth rates), additional batch experiments were conducted in triplicate in 250-ml Erlenmeyer flasks because aeration rates can be limiting in the Bioscreen C. For most nonlethal strains with reduced growth rates (Δ*yjjN*, Δ*tolC*, Δ*dctA*, Δ*idnT*, and Δ*putP*), Erlenmeyer flask experiments were conducted after the removal of the antibiotic resistance gene (for methods see ref. 27). Growth rates are reported for flask experiments as mean ± standard deviation for three replicates, with the exception of duplicates for Δ*yeaT*. Strains were precultured overnight in 2 g/liter glucose-supplemented M9 minimal medium (strains tested for their ability to use propionate as a carbon source were precultured in LB medium), resuspended in M9 minimal medium with no carbon source, and transferred to a new medium with a starting OD (at 600 nm) equal to 0.05. All carbon sources were tested in M9 minimal medium (6.8 g of Na_2HPO4, 3 g of KH_2PO_4, 0.5 g of NaCl, 1 g of NH_4Cl, 2 ml of 1M $MgSO_4$, and 100 μl of 1 M $CaCl_2$ per liter) with a concentration of 2 g/liter. Nitrogen sources (2 g/liter) were screened in W-salts minimal medium (30) containing succinate (2 g/liter).

HPLC Analysis of Medium. HPLC analysis of the growth medium was conducted for parental and mutant strains growing on thymidine as a sole carbon source. Cells were filtered from the medium by using a 0.22-μm filter. An aminex HPX-87H anion-exchange column (Bio-Rad Laboratories, Hercules, CA) then was used to separate medium components on a Waters HPLC system. A 5 M degassed solution of sulfuric acid was used as the mobile phase with a flow rate of 0.5 ml/min. Each sample was run for 35 min, and standards for thymine and thymidine showed that they eluted off the column at 25 and 30 min, respectively. UV absorbances were measured at 260 nm and 210 nm.

Enzymatic Assays. An expression vector containing the *yeaU* gene was generated with the pCRT7 TOPO TA expression kit (Invitrogen, Carlsbad, CA) by using the manufacturer's recommended protocols. Transcription of the gene was induced by the addition of 75 mM isopropyl β-D-thiogalactoside (IPTG). Cells were lysed, and the overexpressed protein in crude extract was partially purified by using an Amicon μm 30,000-molecular-weight cut-off centrifugal filter (Millipore, Billerica, MA). Kinetic assays were carried out at room temperature as previously reported (at pH 8 instead of 8.4) (31) with D-malate concentrations between 0.5 and 20 mM. All tests were additionally performed on control lysates from cells harboring a reverse-insert vector, where the *yeaU* gene was inserted in the backward orientation.

RT-PCR Measurements of Expression Levels. BW25113 precultures were grown in 250-ml Erlenmeyer flasks until reaching mid-log phase and then transferred in triplicate into fresh media. Cells then were harvested during mid-log phase (OD at 600 nm between 0.2 and 0.5). Samples were RNA-stabilized by using RNAProtect Bacterial Reagent, and a RNeasy mini kit was used to isolate total RNA (both from Qiagen, Valencia, CA). cDNA was prepared from the total RNA, and QIAquick PCR Purification Kits were used to clean up the cDNA synthesis product. The resulting cDNA then was quantified and used in subsequent RT-PCR assays. Nine replicate RT-PCR measurements (three replicates for each of the triplicate cDNA samples) were made for each gene (including the reference gene, *acpP*) under the specified growth conditions. cDNA from cells grown on L-lactate was used as a control. A standard curve (generated by using different amounts of genomic DNA instead of cDNA with fixed primer concentrations) was used to determine the primer efficiencies. The relative gene-expression levels were determined by normalizing the amount of a cDNA product to the quantity of *acpP* cDNA from the same cDNA sample.

ChIP. BW25113 and YeaT-8x myc-tagged cells were grown aerobically in M9 minimal medium with D-malate (2 g/liter) at 37°C until an OD of 0.6 at 600 nm was reached. Triplicate experiments were conducted for each strain. Cells then were treated with 1% formaldehyde at room temperature for 30 min to allow cross-linking to occur. YeaT-8x myc bound DNA was isolated for all six samples by using previously described immunoprecipitation methods (29). Quantitative PCR then was carried out on these samples by using primers for the −155 to −2 region upstream of *yeaU* and the promoter region of *acpP*. Triplicate quantitative PCR measurements were made for each of the six samples.

We wish to thank Tomoya Baba and Hirotada Mori for generously providing us access to the Keio *E. coli* gene deletion strain collection and Byung-Kwan Cho for help with the ChIP experiments. Use of this collection greatly facilitated the detailed results generated in this work and was a valuable resource. These studies were supported by National Institutes of Health Grant GM57089. J.L.R. also received financial support as a Faculty Fellow at the University of California at San Diego.

1. Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, *et al*. (2006) *Nucleic Acids Res* 34:1–9.
2. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, *et al*. (1999) *Nature* 399:323–329.
3. Reed JL, Famili I, Thiele I, Palsson BO (2006) *Nat Rev Genet* 7:130–141.
4. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) *Science* 292:929–934.
5. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) *Nature* 429:92–96.
6. King RD, Whelan KE, Jones FM, Reiser PG, Bryant CH, Muggleton SH, Kell DB, Oliver SG (2004) *Nature* 427:247–252.
7. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402.
8. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO (2000) *Nature* 405:823–826.
9. Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO, Eisenberg D (2004) *Genome Biol* 5:R35.
10. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) *Mol Syst Biol* 2:2006.0008.
11. Bochner BR, Gadzinski P, Panomitros E (2001) *Genome Res* 11:1246–1255.
12. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) *Genome Biol* 4:R54.
13. Pharkya P, Burgard AP, Maranas CD (2004) *Genome Res* 14:2367–2376.
14. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) *Nucleic Acids Res* 32:D277–D280.
15. Bausch C, Peekhaus N, Utz C, Blais T, Murray E, Lowary T, Conway T (1998) *J Bacteriol* 180:3704–3710.
16. Hosie AH, Allaway D, Poole PS (2002) *J Bacteriol* 184:5436–5448.
17. Loh KD, Gyaneshwar P, Markenscoff Papadimitriou E, Fong R, Kim KS, Parales R, Zhou Z, Inwood W, Kustu S (2006) *Proc Natl Acad Sci USA* 103:5114–5119.
18. Ito M, Baba T, Mori H (2005) *Metab Eng* 7:318–327.
19. Tipton PA, Peisach J (1990) *Biochemistry* 29:1749–1756.
20. Cooper RA (1978) *Arch Microbiol* 118:199–206.
21. Cooper RA (1979) *FEBS Lett* 103:216–220.
22. Kharchenko P, Vitkup D, Church GM (2004) *Bioinformatics* 20 (Suppl 1):I178–I185.
23. Glasner JD, Liss P, Plunkett G, III, Darling A, Prasad T, Rusch M, Byrnes A, Gilson M, Biehl B, Blattner FR, Perna NT (2003) *Nucleic Acids Res* 31:147–151.
24. Measday V, Baetz K, Guzzo J, Yuen K, Kwok T, Sheikh B, Ding H, Ueta R, Hoac T, Cheng B, *et al*. (2005) Proc *Natl Acad Sci USA* 102:13956–13961.
25. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ (2005) *Bioinformatics* 21:1603–1609.
26. Burgard AP, Vaidyaraman S, Maranas CD (2001) *Biotechnol Prog* 17:791–797.
27. Datsenko KA, Wanner BL (2000) *Proc Natl Acad Sci USA* 97:6640–6645.
28. Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jimenez-Jacinto V, Santos-Zavaleta A, Bonavides-Martinez C, Collado-Vides J (2004) *Nucleic Acids Res* 32:D303–D306.
29. Cho BK, Knight EM, Palsson BO (2006) *Biotechniques* 40:67–72.
30. Smith GR, Halpern YS, Magasanik B (1971) *J Biol Chem* 246:3320–3329.
31. Kohn LD, Jakoby WB (1966) *Methods Enzymol* 9:236–240.