nature biotechnology

Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets

Hal Alper[1], Kohei Miyaoku[1,2] & Gregory Stephanopoulos[1]

Identification of genes that affect the product accumulation phenotype of recombinant strains is an important problem in industrial strain construction and a central tenet of metabolic engineering. We have used systematic (model-based) and combinatorial (transposon-based) methods to identify gene knockout targets that increase lycopene biosynthesis in strains of *Escherichia coli*. We show that these two search strategies yield two distinct gene sets, which affect product synthesis either through an increase in precursor availability or through (largely unknown) kinetic or regulatory mechanisms, respectively. Exhaustive exploration of all possible combinations of the above gene sets yielded a unique set of 64 knockout strains spanning the metabolic landscape of systematic and combinatorial gene knockout targets. This included a global maximum strain exhibiting an 8.5-fold product increase over recombinant K12 wild type and a twofold increase over the engineered parental strain. These results were further validated in controlled culture conditions.

Optimization of metabolic phenotype often requires the simultaneous rerouting of metabolic intermediates and rewiring of regulatory networks. In prior work, this optimization has been accomplished by the modification of genes with well-defined structural or regulatory roles in the context of the particular metabolic pathway being considered[1–3]. Distant genes affecting a metabolic phenotype either through redistribution of metabolite precursors or indirect kinetic and global regulatory effects have been particularly challenging to identify. Models are relatively ineffective in the search for such genes because of their inability to capture the genes' complex, nonlinear kinetic and regulatory interactions. In general, methods for identifying genetic targets are not as powerful as the molecular biological tools that are effectively used to modify such targets. These issues become more involved when one considers the possibility of multiple gene modulations[4]. In general, the complex nature of the metabolic landscape raises significant challenges in the development of an optimal search strategy because varying genetic backgrounds and culturing conditions have a profound impact on the type of gene targets identified by various strategies.

Recently, we reported on a method for the rational design of strains that identifies single and multiple gene knockout targets based on a global stoichiometric analysis. The method was applied successfully to increase lycopene production in recombinant strains of *Escherichia coli*[5]. Lycopene production was investigated in the context of the nonmevalonate[6] pathway in which cells are recombinant, expressing the *crtEBI* operon to encode for the polymerization into the 40-carbon molecule product. The pre-engineered strain used for the study contained chromosomal overexpressions of *dxs*, *idi* and *ispFD*[5] (**Fig. 1a**). There has been a significant effort to specifically engineer the isoprenoid pathway and downstream genes[7–13]; however, in the previous study[5] and this current one, we investigate genome-wide gene knockout targets. A total of seven single and multiple stoichiometric gene deletions, (Δ*gdhA*, Δ*aceE*, Δ*ytjC (gpmB)*, Δ*fdhF*, Δ*gdhA* Δ*aceE*, Δ*gdhA* Δ*ytjC*, Δ*gdhA* Δ*aceE* Δ*fdhF*), were predicted and experimentally validated to increase lycopene production through increasing the supply of precursors and cofactors that are important in the lycopene pathway[5]. These seven mutations along with the parental strain comprise the set of eight systematically designed genotypes. The left panel of **Figure 1b** depicts the methodology for identifying these systematic gene knockout targets.

Lycopene production in these systematically identified knockout strains was still below the stoichiometric maximum, presumably limited by unknown kinetic or regulatory factors that are unaccounted for in stoichiometric models. To identify additional knockout targets that affect the lycopene phenotype via regulatory, kinetic or other unknown mechanisms, we undertook a global transposon library search in the background of the pre-engineered parental strain. Screening this transposon library on glucose plates identified three gene targets that correlated with lycopene overproduction. Upon sequencing, these combinatorial targets were identified as *rssB* (also known as *hnr*), *yjfP* and *yjiD*. In the case of *yjiD*, the transposon was found to be inserted between the identified promoter region and the gene for *yjiD* and will henceforth be referred to as Δ_p*yjiD*. The right panel of **Figure 1b** shows the identity and annotated function of these selected gene targets along a representative location of the transposon insertion event. We note that none of the previously identified single stoichiometric genes surfaced in the combinatorial transposon search

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Room 56-469, Cambridge, Massachusetts 02139, USA. [2]On leave from Mitsubishi Chemical Corporation. Correspondence should be addressed to G.S. (gregstep@mit.edu).

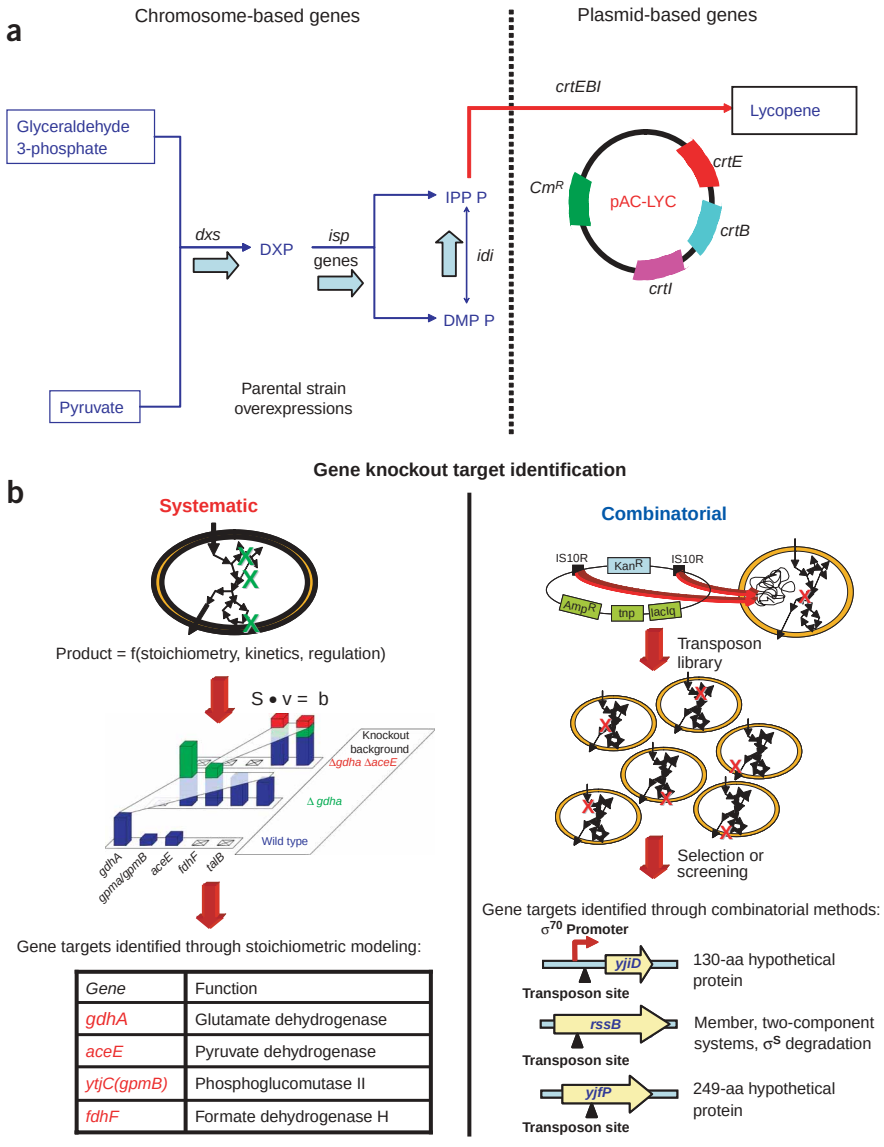
Published online 10 April 2005; doi:10.1038/nbt1083

because of the relatively high threshold of the lycopene accumulation level imposed in the selection of candidate strains. Using these three identified targets, it is possible to create a total of seven gene combinations of single, double and triple combinatorial target mutations (Δ*rssB*, Δ*yjfP*, Δ$_P$*yjiD*, Δ*rssB* Δ*yjfP*, Δ*rssB* Δ$_P$*yjiD*, Δ*yjfP* Δ$_P$*yjiD* and Δ*rssB* Δ*yjfP* Δ$_P$*yjiD*). These seven combinations along with the parental strain constitute the combinatorial strain set comprising a total of eight strains.

The previous results point to two distinct sets of stoichiometric and combinatorial gene targets. It is not clear how these targets interact when combined. To answer this question, we conducted an exhaustive study of the 64 strains comprising all combinations of the eight stoichiometric and eight combinatorial genotypes. These target genes were modified in the background of the pre-engineered recombinant *E. coli* strain. The resulting production profiles over the course of a 48-hour shake-flask fermentation process provided the information needed for the complete mapping of the lycopene metabolic landscape (**Fig. 2**).

Several interesting observations arise from the topology of this metabolic landscape. First, two global maxima exist, each with production levels around 11,000 p.p.m. (μg/g dry cell weight). The first strain contains the Δ*gdhA* Δ*aceE* Δ*fdhF* genotype, which is a purely stoichiometrically designed strain. The other maximum is Δ*gdhA* Δ*aceE* Δ$_P$*yjiD*, which is created through the combination of stoichiometric and combinatorial targets. Second, several local maximum points are present with production levels ranging from 8,400 to 9,400 p.p.m., each formed from the combination of systematic and combinatorial targets. Third, the left quadrant of the graph indicates that the combination or stacking of more than one combinatorial knockout target greatly reduces lycopene levels to below 2,000 p.p.m., and as low as only 500 p.p.m. for some constructs, which is below the production level of a recombinant wild-type *E. coli* K12 strain. Finally, visual inspection of this landscape suggests a highly nonlinear function with many local optima.

Clustering methods have been routinely applied to the analysis of microarray (and other) data to determine sets of genes that exhibit similar expression profiles[14]. Likewise, the technique of hierarchical clustering may be applied to the metabolic landscape of **Figure 2** to cluster gene knockout constructs exhibiting similar production profiles over the four time points. Presumably, strains clustering most



Gene	Function
gdhA	Glutamate dehydrogenase
aceE	Pyruvate dehydrogenase
ytjC(gpmB)	Phosphoglucomutase II
fdhF	Formate dehydrogenase H

Figure 1 Systematic and combinatorial gene knockout target identification. (**a**) Lycopene synthesis begins with the condensation of the key glycolytic intermediates, glyceraldehyde 3-phosphate and pyruvate and continues in a nearly linear pathway. In the engineered strain used in this study, the *idi*, *ispFD*, and *dxs* genes are overexpressed by chromosomal promoter replacement. To produce lycopene, a cluster of genes, *crtEBI* are expressed on a plasmid. (**b**) Systematic targets (illustrated on the left) were identified through the use of global, stoichiometric modeling (as more comprehensive models are unavailable) to identify gene knockouts which were predicted *in silico* to increase lycopene by increasing either cofactor or precursor supply[5]. Combinatorial targets (illustrated on the right) were identified through the use of transposon mutagenesis. The gene *rssB* is a response regulator responsible for recruiting the proteolysis of the stationary phase sigma factor, σ^S (encoded by *rpoS*)[21,22], which has previously been implicated in the overproduction of carotenoids[23]. The gene *yjfP* is a 249–amino acid protein which is currently not annotated, but has been putatively categorized as either a nonpeptidase homolog[24] or as a putative hydrolase (1st module)[25]. Finally, *yjiD* is a 130–amino acid protein with an unknown function[25]. For only the *yjiD* mutants, the transposon site was only found between the promoter region and the gene. These targets were combined to create the unique set of 64 mutant strains used in this study.

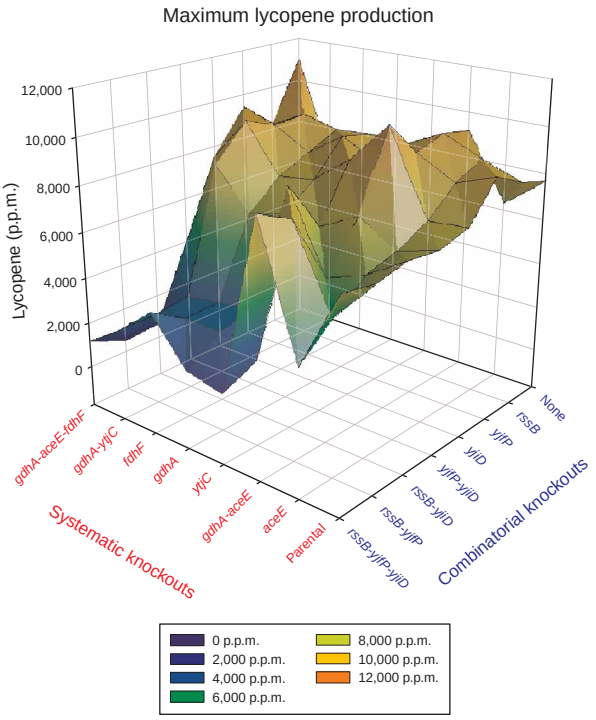


Figure 2 Visualization of the metabolic landscape. Eight systematically derived knockout genotypes were combined with eight systematically derived genotypes to create a unique collection of 64 strains. The maximum lycopene production (in p.p.m.) during the course of a 48-h shake-flask fermentation is plotted. Among the interesting features of this landscape is the presence of two global maxima (at around 11,000 p.p.m.) and several local maxima. Furthermore, certain combinations of combinatorial targets in most systematically derived genetic backgrounds resulted in a substantial decrease in lycopene production.

closely accumulate product by following similar modes of action in the mechanism of lycopene production. Upon clustering the entire set of 64 strains, two distinct organizations emerge for the two sets of gene targets previously identified.

Clustering lycopene profiles (across the four time points) for the eight stoichiometric knockout strains revealed a fairly close, stacked dendrogram (see abscissa of **Fig. 3a**). When these strains are plotted against the lycopene accumulation level, they reveal an expanding concentric bubble-plot suggesting an additive effect of accumulating gene deletions. This is in concert with the presumed mode of action in these strains, namely the increasing availability of precursors and cofactors that are needed for lycopene biosynthesis.

In contrast to **Figure 3a**, all combinatorial targets, as exemplified by *rssB*, force a split-tree shape in the dendrogram when performed in the background of each of the seven stoichiometric targets (**Fig. 3b**). Different time courses in lycopene accumulation suggest different modes of action for the effect of the combinatorial genes on this phenotype. Specifically, whereas each construct formed from the deletion of a single combinatorial target gene tends to exhibit similar behavior (increased production), the deletion of combinations of these genes yields phenotypes that are neither linear nor synergistic. In fact, double and triple knockout constructs arising from these combinatorial targets exhibit vastly different production profiles from the individual targets (**Fig. 2**). This nonlinearity suggests that the combinatorial targets are disrupting regulatory processes that are relatively incompatible, and in certain cases deleterious, when combined.

Biological differences are observed when combinatorial genes are deleted together with stoichiometric ones. Strains in cluster Y (**Fig. 3b**) all exhibit an extended lag phase, which extends to 16–18 h before reaching a typical cell density OD 3.5–4.0. In contrast, strains in cluster Z do not posses such a lag phase and exhibit a steady increase of lycopene production with time. The average, scaled

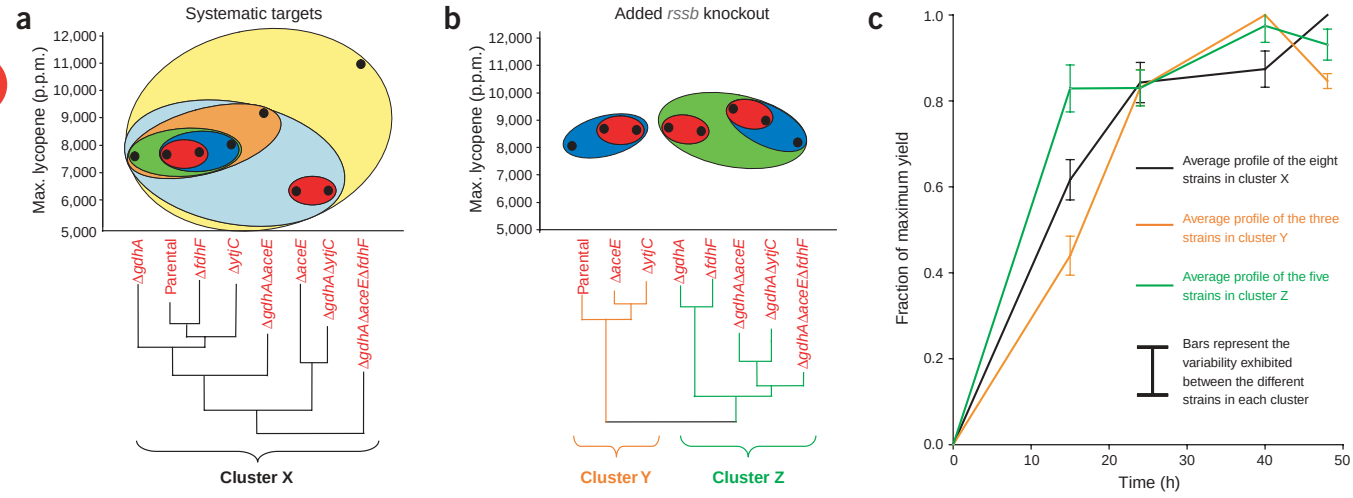


Figure 3 Clustering analysis depicting the interaction of systematic and combinatorial targets. Lycopene production profiles across the 48-h shake-flask fermentation are clustered, resulting in the dendrograms illustrated. (**a**) The purely systematic strains have a stacked dendrogram which is visually illustrated with a concentric bubble plot. Strains that are more tightly clustered have similar modes of action, thus all systematic strains seem to be additive in nature. This is further evidenced by the close clustering of Δ*fdhF* and the parental strain, as the *fdhF* single knockout was determined from the stoichiometric analysis to bring about no enhancement of lycopene production. (**b**) Conversely, the addition of any combinatorial genotype, *rssB* in this case, decouples the systematic design and causes a disjoint pattern in the dendrogram and bubble plot. This has the implication that local, metabolic gene targets are more accessible through a sequential search than global, regulatory targets, which require a simultaneous search that is sensitive to the genetic background of the strain. **c** compares the average, relative production profiles for the three clusters shown in **a** and **b**. The biological differences in the production profiles for each of these clusters are evident.

production profiles for the purely systematic cluster and the two clusters forced by an *rssB* deletion are compared in **Figure 3c**. It is noted that this branched pattern is exhibited by all strains constructed from the deletion of any combinatorial gene in the background of the stoichiometric targets, with different production profiles characterizing each of the clusters.

Drawing from this analysis, it appears that stacking (that is, deleting) combinatorial target genes upon stoichiometric ones leads to a decoupling of the stoichiometric logic. This decoupling is evident in analyzing the impact of the deletion of *rssB* or any other combinatorial gene, on the shape of the dendrogram obtained from hierarchical clustering of the lycopene accumulation profiles for the eight stoichiometric strains; it is also quantified by covariance analysis (**Supplementary Fig. 1** and **Supplementary Discussion**).

The exhaustive exploration of the combinations of stoichiometric and combinatorial targets allowed the identification of several interesting strains on the basis of their performance in batch shake-flask cultivations. To better assess the production capacity of these knockout strains, fed-batch cultivations were carried out in shake-flasks and controlled bioreactors with staged glucose feed (**Fig. 4**). Several strains were thus evaluated. Optimized shake-flask fermentations highlight the capability of the global maximum strains to produce upwards of 18,000 p.p.m. in 24–40 h (**Fig. 4**). These global maximum strains were also grown in 500-ml bioreactors with a similar glucose feeding profile and pH control and showed enhanced lycopene production producing upwards of 23,000 p.p.m. in only 60 h (data not shown). Further improvements are possible through iterative bioreactor optimization. The good correspondence between fermentor and shaker-flasks results suggests that the performance of the strains selected by the described method is transferable to larger systems.

Identification of multiple gene targets affecting a particular phenotype is an open problem. Among the complications are strong nonlinear effects, lack of accurate models capable of capturing genetic interactions and ineffective search strategies. To address these issues in the context of lycopene production, we undertook an exhaustive experimental search to investigate combinations of rationally selected genes with those identified through combinatorial methods. A number of promising strains were obtained, some of which were capable of producing upwards of 18,000 p.p.m. (or 18 mg/g dry cell weight) of lycopene in defined glucose medium using simple fed-batch conditions. This value represents a nearly fourfold increase over the parental strain when cultured in simple cultivation, a twofold increase over the pre-engineered parental strain in similar conditions and an 8.5-fold increase over recombinant wild-type K12 *E. coli* under similar conditions.

The metabolic landscape defined through this unique set of 64 knockout strains allows for several observations of importance to metabolic engineering. First, rationally selected stoichiometric gene knockout targets have the potential of generating serious contenders in the quest for maximally producing strains. We note that one of the two maximum overproducing strains resulted from the knockout of three stoichiometric genes (*gdhA*, *aceE*, *fdhF*). Additionally, the knockout of specific combinatorial genes yielded substantially enhanced phenotypes in the background of particular stoichiometric knockout genes. Second, whereas combinatorial gene targets hold greater potential than stoichiometric ones as single knockout mutants, multiple knockouts of the combinatorial gene set led to a distinct deterioration of the lycopene phenotype. Yet, it proved invaluable in the creation of some important strains in the landscape. Third, the presence of many local maxima complicates the nature of the landscape and raises questions about general sequential search strategies. Previously, sequential search strategies were found to be quite effective

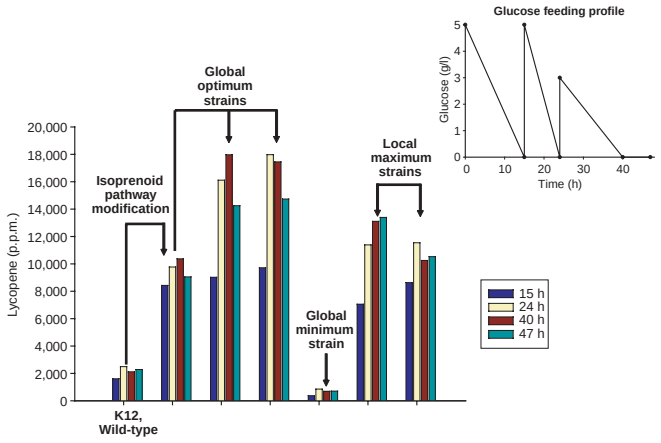


Figure 4 Behavior of selected strains in optimized culturing conditions. Selected strains from the metabolic landscape were cultured in fed-batch shake-flasks with increased M9 salts and a staged glucose feed as represented in the glucose feeding profile. Strains presented from left to right are K12 (recombinant wild-type); engineered parental strain with *dxs*, *idi*, and *ispFD* overexpressions; Δ*gdhA* Δ*aceE* Δ*fdhF*; Δ*gdhA* Δ*aceE* Δ$_P$*yjiD*; Δ*gdhA* Δ*aceE* Δ*fdhF* Δ*rssB* Δ*yjfP*; Δ*gdhA* Δ*aceE* Δ*fdhF* Δ*yjfP*; Δ*gdhA* Δ*aceE* Δ*rssB* Δ*yjfP* Δ$_P$*yjiD*. The two global maxima were capable of producing upwards of 18,000 p.p.m. in 24 to 40 h. Strain behavior was transferable to these optimized conditions. These results highlight that strains isolated on solid-media plates retained the lycopene overproduction phenotype in the course of the scale-up process.

when applied to the space of stoichiometric genes[5], which is due to their overall additive effect on phenotype. **Figure 2** suggests that this result does not hold when combinatorial genes are also included in the search space, necessitating exhaustive combinatorial searches of the type undertaken in this study. Although identification of optimal gene targets will continue to be a demanding undertaking, searches for gene targets will be significantly aided by advanced models of cell function accounting for kinetic and regulatory mechanisms.

It should be noted that the search of this study was limited to the effect of gene knockout only. Gene knockdown or overexpression adds an extra layer of complexity in the metabolic engineering of overproducing strains and could provide further drastic improvements of product overproduction phenotypes.

This study underscores some important issues optimizing phenotype. First, high-throughput screening methods combined with detailed cellular models will aid in efficient strain optimization. Second, combinatorial targets influencing global cellular function should be invoked at later stages in the strain improvement process to avoid selecting those with limited utility or incompatible modes of action. Finally, metabolic genes seem to have a linear impact in the overall cellular phenotype whereas the effect of regulatory targets is definitely nonlinear and more complex. This work serves as a case study aiming to understand the complex interaction of the genotype-phenotype space in the context of product overproduction phenotype. The lessons gained from the exhaustive exploration of systematic and combinatorial gene knockout sets can help shape future strain improvement programs as they are tested in diverse systems for divergent products.

METHODS

Strains and media. *E. coli* K12 PT5-*dxs*, PT5-*idi*, PT5-*ispFD*, provided by DuPont, was used as the lycopene expression strain when harboring the pACLYC plasmid containing the *crtEBI* operon[15]. Overexpression of *dxs*, *idi*, and

ispFD was chromosomally incorporated without an antibiotic marker through promoter delivery. Strains were grown at 37 °C with 225 r.p.m. orbital shaking in M9-minimal medium[16] containing 5 g/l D-glucose and 68 μg/ml chloramphenicol. All simple cultures were 50 ml, grown in a 250-ml flask with an 1% (vol/vol) inoculation from an overnight 5-ml culture and assayed at 15, 24, 39 and 48 h. Optimized shaker-flasks were 50-ml cultures grown in 250-ml flasks with a 1% (vol/vol) inoculation from an overnight 5-ml culture with glucose feeds of 5 g/l at 0 and 15 h and 3 g/l at 24 h. The medium for these experiments was M9-minimal medium[16] with double concentrations of all salts except $CaCl_2$ and $MgSO_4$. All experiments were performed in biological knockout replicates to validate data and calculate statistical parameters. Glucose monitoring was conducted periodically using an r-Biopharm kit to verify complete usage of glucose. Cell density was monitored spectrophotometrically at 600 nm. All PCR products were purchased from Invitrogen and used *Taq* polymerase. M9 Minimal salts were purchased from US Biological and all remaining chemicals were from Sigma-Aldrich.

Transposon library screening and sequencing. Transposon libraries were generated using the pJA1 vector[17]. Cells were transformed with between 800 and 1,600 ng of the plasmid, then diluted and plated on M9-glucose-agar plates (containing 1 mM isopropyl-β-D-thiogalactoside) with a target density of 200 colonies per 150 × 15 mm Petri dishes. Plates were incubated at 37 °C for 36 h, then allowed to sit at 22 °C. Cells identified as exhibiting increased lycopene content (more red) were isolated and cultured throughout the culturing process. The identity of promising targets were sequenced using an altered version of Thermal Asymmetric Interlaced PCR (TAIL-PCR)[18]. For the TAIL1 reaction, 1.5 μl of genomic DNA isolated using a DNA purification kit (Promega) was used as the initial template. The TAIL3 reaction was increased to 30 cycles. Kanamycin-specific primers: TAIL1, 5′-TATCAGGACATAG CGTTGGCTACCCG-3′; TAIL2, 5′-CGGCGAATGGGCTGACCGCT-3′; TAIL3, 5′-TCGTGCTTTACGGTATCGCCGCTC-3′. The degenerate primer AD1 was used as described in the reference. The product of the TAIL3 reaction was purified by a PCR cleanup kit (Qiagen) after gel visualization. This product was sequenced using the primer TAIL-seq, 5′-CATCGCCTTCTATCGCCTTCTT-3′. Gene target identity was determined through BLAST nucleotide sequence comparison. Strains identified through transposon mutagenesis were subsequently constructed by using PCR product recombination and tested for maintenance of the lycopene overproduction phenotype.

Knockout construction and verification. Gene deletions were conducted using PCR product recombination[19] using the pKD46 plasmid expressing the lambda red recombination system and pKD13 as the template for PCR (see **Supplementary Table 1** online for primer designs). Gene knockouts were verified through colony PCR. Phage transduction was used for creating multiple gene knockout strains. P1*vir* phage transduction was used to transfer knockout mutants between strains[20]. PCR primers used for knockout and verification may be found in **Supplementary Table 1** online.

Lycopene assay. Intracellular lycopene content was extracted from 1 ml of bacterial culture at the point of total glucose exhaustion. The cell pellet was washed, and then extracted in 1 ml of acetone at 55 °C for 15 min with intermittent vortexing. The lycopene content in the supernatant was quantified through absorbance at 475 nm[12] and concentrations were calculated through a standard curve. The entire extraction process was performed in reduced light conditions to prevent photobleaching and degradation. Cell mass was calculated by correlating dry cell with OD_{600} for use in p.p.m. (mg lycopene/g dry cell weight) calculations.

Hierarchical clustering routines. A complete linkage hierarchical clustering of the lycopene time profiles for the entire 8 × 8 strain matrix (containing values of the maximum lycopene production) using the Euclidean distance as the similarity metric was performed using Cluster Version 3.0. Dendrograms were visualized using Java TreeView Version 1.0.8.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We acknowledge financial support of this work by the DuPont-MIT Alliance. In particular, we would like to thank Wonchul Suh for providing the parental *E. coli* strain. We also thank Joel Moxley for providing thoughtful suggestions and Veronica Godoy for providing the initial phage stock.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 10 February; accepted 2 March 2005
Published online at http://www.nature.com/naturebiotechnology/

1. Stephanopoulos, G., Aristidou, A. & Nielsen, J. *Metabolic Engineering: Principles and Methodologies* (Academic Press, San Diego, CA, 1998).
2. Ostergaard, S., Olsson, L., Johnston, M. & Nielsen, J. Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. *Nat. Biotechnol.* **18**, 1283–1286 (2000).
3. Stafford, D.E. *et al.* Optimizing bioconversion pathways through systems analysis and metabolic engineering. *Proc. Natl. Acad. Sci. USA* **99**, 1801–1806 (2002).
4. Koffas, M.A., Jung, G.Y. & Stephanopoulos, G. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. *Metab. Eng.* **5**, 32–41 (2003).
5. Alper, H., Jin, Y.-S., Moxley, J. & Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* (in the press) doi:10.1016/j.ymben.2004.12.003 (2005).
6. Adam, P. *et al.* Biosynthesis of terpenes: studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase. *Proc. Natl. Acad. Sci. USA* **99**, 12108–12113 (2002).
7. Matthews, P.D. & Wurtzel, E.T. Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl. Microbiol. Biotechnol.* **53**, 396–400 (2000).
8. Misawa, N. & Shimada, H. Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeasts. *J. Biotechnol.* **59**, 169–181 (1997).
9. Farmer, W.R. & Liao, J.C. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol.* **18**, 533–537 (2000).
10. Smolke, C.D., Martin, V.J.J. & Keasling, J.D. Controlling the metabolic flux through the carotenoid pathway using directed mRNA processing and stabilization. *Metab. Eng.* **3**, 313–321 (2001).
11. Lee, P.C. & Schmidt-Dannert, C. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. *Appl. Microbiol. Biotechnol.* **60**, 1–11 (2002).
12. Kim, S.-W. & Keasling, J.D. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol. Bioeng.* **72**, 408–415 (2001).
13. Jones, K.L., Kim, S.-W. & Keasling, J.D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab. Eng.* **2**, 328–338 (2000).
14. Eisen, M., Spellman, P., Brown, P. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
15. Cunningham, F.X., Jr., Sun, Z., Chamovitz, D., Hirschberg, J. & Gantt, E. Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus sp* strain PCC7942. *Plant Cell* **6**, 1107–1121 (1994).
16. Maniatis, T., Fritsch, E.F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1982).
17. Badarinarayana, V. *et al.* Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat. Biotechnol.* **19**, 1060–1065 (2001).
18. Liu, Y.-G. & Whittier, R.F. Thermal asymmetric interlaced pcr: automatable amplification and sequencing of insert end fragments from pi and yac clones for chromosome walking. *Genomics* **25**, 674–681 (1995).
19. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
20. Miller, J.H. *A Short Course in Bacterial Genetics* (Cold Springs Harbor Laboratory Press, Cold Springs Harbor, NY, 1992).
21. Muffler, A., Fischer, D., Altuvia, S., Storz, G. & Hengge-Aronis, R. The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in *Escherichia coli*. *EMBO J.* **15**, 1333–1339 (1996).
22. Sandmann, G., Woods, W. & Tuveson, R.W. Identification of carotenoids in *Erwinia herbicola* and in a transformed *Escherichia coli* strain. *FEMS Microbiol. Lett.* **59**, 77–82 (1990).
23. Becker-Hapak, M., Troxtel, E., Hoerter, J. & Eisenstark, A. RpoS dependent over-expression of carotenoids from *Erwinia herbicola* in OXYR-deficient *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **239**, 305–309 (1997).
24. Rawlings, N., Tolle, D. & Barrett, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **32**, D160–D164 (2004).
25. Serres, M.H. *et al.* A functional update of the *Escherichia coli* K-12 genome. *Genome Biol.* **2**, published online 20 August 2001 (doi:10.1186/gb-2001-2-9-research0035).