MIT Open Access Articles

Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca

Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*

Karsten Temme[a,1], Dehua Zhao[b,1], and Christopher A. Voigt[b,2]

[a]Joint Graduate Group in Bioengineering, University of California, Berkeley/University of California, San Francisco, CA 94158; and [b]Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Bacterial genes associated with a single trait are often grouped in a contiguous unit of the genome known as a gene cluster. It is difficult to genetically manipulate many gene clusters because of complex, redundant, and integrated host regulation. We have developed a systematic approach to completely specify the genetics of a gene cluster by rebuilding it from the bottom up using only synthetic, well-characterized parts. This process removes all native regulation, including that which is undiscovered. First, all noncoding DNA, regulatory proteins, and nonessential genes are removed. The codons of essential genes are changed to create a DNA sequence as divergent as possible from the wild-type (WT) gene. Recoded genes are computationally scanned to eliminate internal regulation. They are organized into operons and placed under the control of synthetic parts (promoters, ribosome binding sites, and terminators) that are functionally separated by spacer parts. Finally, a controller consisting of genetic sensors and circuits regulates the conditions and dynamics of gene expression. We applied this approach to an agriculturally relevant gene cluster from *Klebsiella oxytoca* encoding the nitrogen fixation pathway for converting atmospheric N_2 to ammonia. The native gene cluster consists of 20 genes in seven operons and is encoded in 23.5 kb of DNA. We constructed a "refactored" gene cluster that shares little DNA sequence identity with WT and for which the function of every genetic part is defined. This work demonstrates the potential for synthetic biology tools to rewrite the genetics encoding complex biological functions to facilitate access, engineering, and transferability.

Many functions of interest for biotechnology are encoded in gene clusters, including metabolic pathways, nanomachines, nutrient scavenging mechanisms, and energy generators (1). Clusters typically contain internal regulation that is embedded in the global regulatory network of the organism. Promoters and 5′-UTRs are complex and integrate many regulatory inputs (2, 3). Regulation is highly redundant; for example, it contains embedded feedforward and feedback loops (4). Regulation can also be internal to genes, including promoters, pause sites, and small RNAs (5, 6). Further, genes often physically overlap, and regions of DNA can have multiple functions (7). The redundancy and extent of this regulation makes it difficult to manipulate a gene cluster to break its control by native environmental stimuli, optimize its function, or transfer it between organisms. As a consequence, many clusters are cryptic, meaning that laboratory conditions cannot be identified in which they are active (8).

Gene clusters have been controlled from the top down by manipulating the native regulation or adding synthetic regulation in an otherwise WT background (9). For example, knocking out a repressor or overexpressing an activator has turned on clusters encoding biosynthetic pathways (10–14). When the cluster is a single operon, it has been shown that a promoter can be inserted upstream to induce expression (15). The entire echinomycin biosynthetic cluster was transferred into *Escherichia coli* by placing each native gene under the control of a synthetic promoter (16).

In engineering, one approach to reduce the complexity of a system is to "refactor" it, a term borrowed from software development whereby the code underlying a program is rewritten to achieve some goal (e.g., stability) without changing functionality (17). This term was first applied to genetics to describe the top-down simplification of a phage genome by redesigning known genetic elements to be individually changeable by standard restriction digest (18). Here, we use it to refer to a comprehensive bottom-up process to systematically eliminate the native regulation of a gene cluster and replace it with synthetic genetic parts and circuits (Fig. 1). The end product is a version of the gene cluster whose DNA sequence has been rewritten, but it encodes the same function. The design process occurs on the computer, and then the resulting DNA sequence is constructed by using DNA synthesis (19). The first step of the process is to remove all noncoding DNA and regulatory genes. Next, each essential gene is recoded by selecting codons that produce a DNA sequence that is as distant as possible from the WT sequence. The intent is to introduce mutations throughout the gene to eliminate internal regulation (including that which is undiscovered), such as operators, promoters, mRNA secondary structure, pause sites, methylation sites, and codon regulation. Recoded sequences are further scanned by computational methods to identify putative functional sequences, which are then removed. The recoded genes are organized into artificial operons, and the expression levels are controlled by synthetic ribosome binding sites (RBSs), and spacer sequences physically separate the genes. The end result is a refactored gene cluster whose native regulation has been removed and has been organized into a set of discrete, well characterized genetic parts.

When the native regulation has been removed, synthetic regulation can be added back to control the dynamics and conditions under which the cluster is expressed. Constructing such regulation has been a major thrust of synthetic biology and involves the design of genetic sensors and circuits and the understanding of how to connect them to form programs (20). In our design, we genetically separate the sensing/circuitry from the refactored pathway by carrying them on different low-copy plasmids (Fig. 1). The plasmid containing the sensors and circuits is referred to as the controller, and the output of the circuits led to the expression of an engineered T7 RNA polymerase (T7* RNAP). The refactored cluster is under the control of T7 promoters. One advantage of this organization is that T7 RNA polymerase is orthogonal to native transcription and the T7 promoters are tightly off in the absence of the controller. In addition, changing the regulation is simplified to swapping the controller for one that contains different sensors and circuits, so long as the dynamic range of T7* RNAP is fixed.

As a demonstration, we have applied this process to refactor the gene cluster encoding nitrogen fixation in *Klebsiella oxytoca* (21).
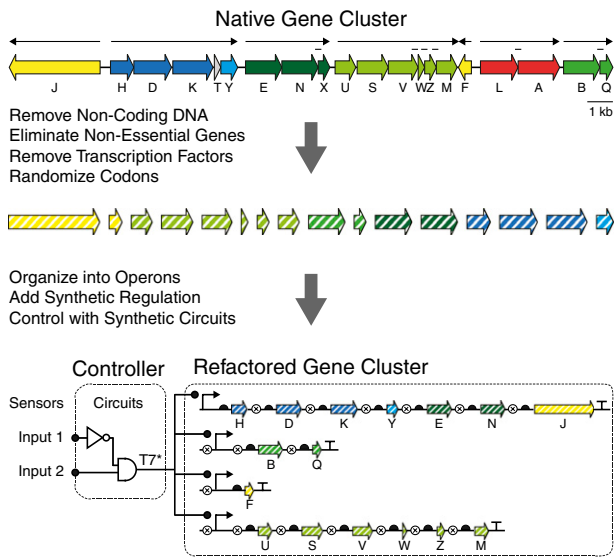
Fig. 1. The process of refactoring a gene cluster. *Top*: WT *K. oxytoca* nitrogen fixation gene cluster. The genes are colored by function: blue (nitrogenase), green (cofactor biosynthesis, shading corresponds to operons), yellow (e⁻ transport), and gray (unknown). The thin arrows show the length and orientation of the seven operons, and a horizontal bar indicates overlapping genes. The recoded genes are shown as dashed lines. The symbols used to define the refactored cluster and controller are defined in Figs. 4 and 5, respectively.

Nitrogen fixation is the conversion of atmospheric N_2 to ammonia (NH_3) so that it can enter metabolism (22). Industrial nitrogen fixation through the Haber–Bosch process is used to produce fertilizer. Many microorganisms fix nitrogen, and the necessary genes typically occur together in a gene cluster, including the nitrogenase subunits, the metallocluster biosynthetic enzymes and chaperones, e⁻ transport, and regulators (Fig. 2*A*) (23, 24). The gene cluster from *K. oxytoca* has been a model system for studying nitrogen fixation and consists of 20 genes encoded in 23.5 kb of DNA (Fig. 1, *Top*) (25). The biosynthesis of nitrogenase is tightly regulated by a two-layer transcriptional cascade in response to fixed nitrogen, oxygen, and temperature (26). The complete cluster has been transferred to *E. coli*, thus demonstrating that it has all the genes necessary for nitrogen fixation (27). The encoding of this function is complex, many of the genes overlap, the operons are oriented in opposite directions, and there are many putative hidden regulatory elements, including internal promoters and hairpins (25). The purpose of refactoring is to reorganize the cluster, simplify its regulation, and assign a concrete function to each genetic part.

Results

Tolerance of Native Gene Cluster to Changes in Expression. Before refactoring a cluster, a robustness analysis is performed to determine the tolerances of a gene or set of genes to changes in expression level (Fig. 2*B*). This informs the grouping of genes into operons and the selection of synthetic parts to obtain desired expression levels. In the WT background, genes are knocked out and complemented under inducible control. The tolerance is obtained by measuring nitrogenase activity as a function of the activity of the inducible promoter.

Nitrogenase function is notably sensitive to expression changes, and each tolerance has a clear optimum (Fig. 2*B*). The chaperone NifY is required to achieve full activity and broadens the tolerance to changes in expression level. NifT did not have an effect on activity, as observed previously (28), and it is frequently absent from homologous clusters (29). The genes controlling electron
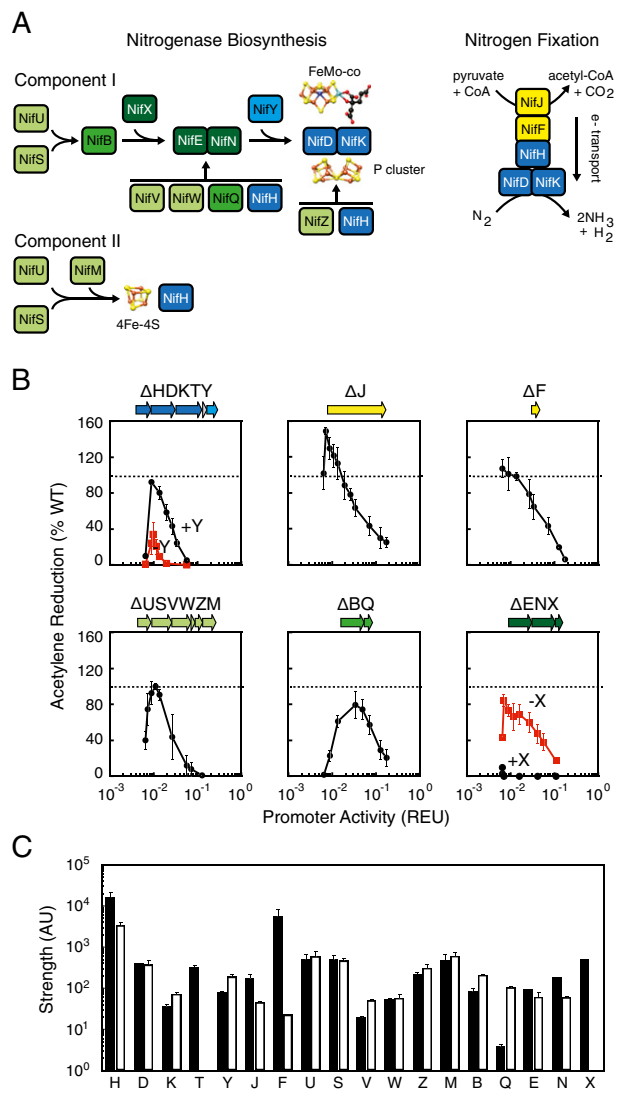

Fig. 2. The robustness of the nitrogen fixation pathway to changes in the expression of component proteins. (*A*) The pathway for nitrogenase maturation is shown and proteins are colored by function (Fig. 1). The metal clusters are synthesized by the biosynthetic pathway (23, 24). Nitrogen fixation catalyzed by the matured nitrogenase is shown with its in vivo electron transport chain. (*B*) The tolerance of nitrogenase activity to changes in the expression of component proteins are shown. Activity is measured via an acetylene reduction assay and the percentage compared with WT *K. oxytoca* is presented. WT operons are expressed from a P_{tac} promoter on a low-copy plasmid (*SI Materials and Methods*). The promoter activity is calculated as the output of the P_{tac} promoter at a given concentration of IPTG and compared with a constitutive promoter. The effect of not including NifY (-Y) and NifX (-X) are shown in red. (*C*) The comparison of the strength of WT (black) and synthetic (white) RBSs is shown. The RBSs were measured through an in-frame transcriptional fusion (−60 to +90) with mRFP. The strength is measured as the geometric average from a distribution of cells measured by flow cytometry (*SI Materials and Methods*). The synthetic RBSs of *nifF* and *nifQ* are not intended to match the WT measurement. Error bars represent the SD of at least three experiments performed on different days.

transport (*nifJ* and *nifF*) need to be expressed at low levels, and activity decreases rapidly as expression increases. The optima for genes participating in the metal cluster biosynthetic pathways vary. The *nifUSVWZM* operon, which encodes proteins for early Fe-S cluster formation and proteins for component maturation, needs to be expressed at low levels, whereas *nifBQ*, encoding proteins for FeMo-co core synthesis and molybdenum integration, need to be

expressed 10-fold higher. NifEN is tolerant to varied expression levels. However, activity is lost with the inclusion of *nifX*, which has been characterized as a negative regulator (30). The native cluster also includes the regulatory proteins NifL and NifA, which integrate environmental signals (26). The genes *nifT*, *nifX*, and *nifLA* are not included in the refactored cluster.

Complete Refactored Gene Cluster. The nitrogenase activities of the refactored operons were measured as a function of the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible P_{tac} promoter (Fig. 3A). Each operon has a different optimum. To combine the operons, the P_{tac} promoters were replaced with T7 promoters that have a strength close to the measured optimum (Fig. 3B and *SI Materials and Methods*). The nitrogenase genes (*nifHDK*) are highly expressed in *Klebsiella* under fixing conditions (as much as 10% of cell protein) (31), so the strongest promoter was used to control this operon [$P_{T7.WT}$, 0.38 relative expression units (REUs)] (32). A long operon was built to include the *nifEN* and *nifJ* genes, whereby the lower expression required for *nifJ* was achieved through transcriptional attenuation. The *nifF* gene was encoded separately under the control of a medium strength promoter ($P_{T7.3}$, 0.045 REU). Finally, the *nifUSVWZM* and *nifBQ* operons were controlled by weak promoters ($P_{T7.2}$, 0.019 REU). Each of the individual refactored operons under the control of a T7 promoter was able to recover the activity observed from the P_{tac} promoter and corresponding optimal IPTG concentration (Fig. 3C).

Transitioning the control to T7* RNAP and T7 promoters facilitates the assembly of the complete cluster from refactored operons. We first assembled half-clusters by using Gibson Assembly (33) and verified their function in strains with the corresponding genes deleted. The first half-cluster consisted of the *nifHDKYENJ* operon. The second half-cluster was assembled from the *nifBQ*, *nifF*, and *nifUSVWZM* operons. The half-clusters were able to recover 18% ± 0.7% and 26% ± 8.4% of WT activity, respectively. The full synthetic cluster was assembled from both half-clusters (Fig. 4), and its activity measured in a strain in which the full cluster is deleted. The synthetic gene cluster recovers nitrogenase activity at 7.4% ± 2.4% of the WT (Fig. 5A). Strains carrying the synthetic gene cluster used ambient N_2 as a nitrogen source, growing 3.5-fold slower than the WT strain (*SI Materials and Methods*) and incorporating ^{15}N-labeled nitrogen into 24% ± 1.4% of their cellular nitrogen content, as measured by isotope ratio mass spectrometry (Fig. 5B).

The complete refactored cluster consists of 89 genetic parts, including a controller, and the function of each part is defined and characterized. Therefore, the genetics of the refactored system are complete and defined in the schematic illustration in Fig. 4. However, the process of simplification and modularization reduces activity (18). This is an expected outcome of refactoring a highly evolved system.

Swapping Controllers to Change Regulation. The separation of the controller and the refactored cluster simplifies changing the regulation of the system. This can be achieved by transforming a different controller plasmid, as long as the dynamic range of the T7* RNAP expression is preserved. To demonstrate this, we constructed two additional controllers (Fig. 5A). Controller 2 changes the chemical that induces the system by placing the expression of T7* RNAP under the control of the aTc-inducible P_{tet} promoter. When induced, controller 2 produces nitrogenase activity identical to controller 1 (7.2% ± 1.7%). The controller can also serve as a platform to encode genetic circuits to control regulatory dynamics or to integrate multiple sensors. To this end, controller 3 contains two inducible systems (IPTG and aTc) and an ANDN gate (34, 35). In the presence of IPTG and the absence of aTc, nitrogen fixation is 6.6% ± 1.7% of WT activity. These controllers represent the simplicity by which the regulation of the refactored cluster can be changed.

In addition to making it possible to add new regulation, the process of refactoring eliminates the native regulation of the cluster. This is demonstrated through the decoupling of nitrogenase activity from the environmental signals that normally regulate its activity. For example, ammonia is a negative regulator that limits overproduction of fixed nitrogen (26). In the presence of
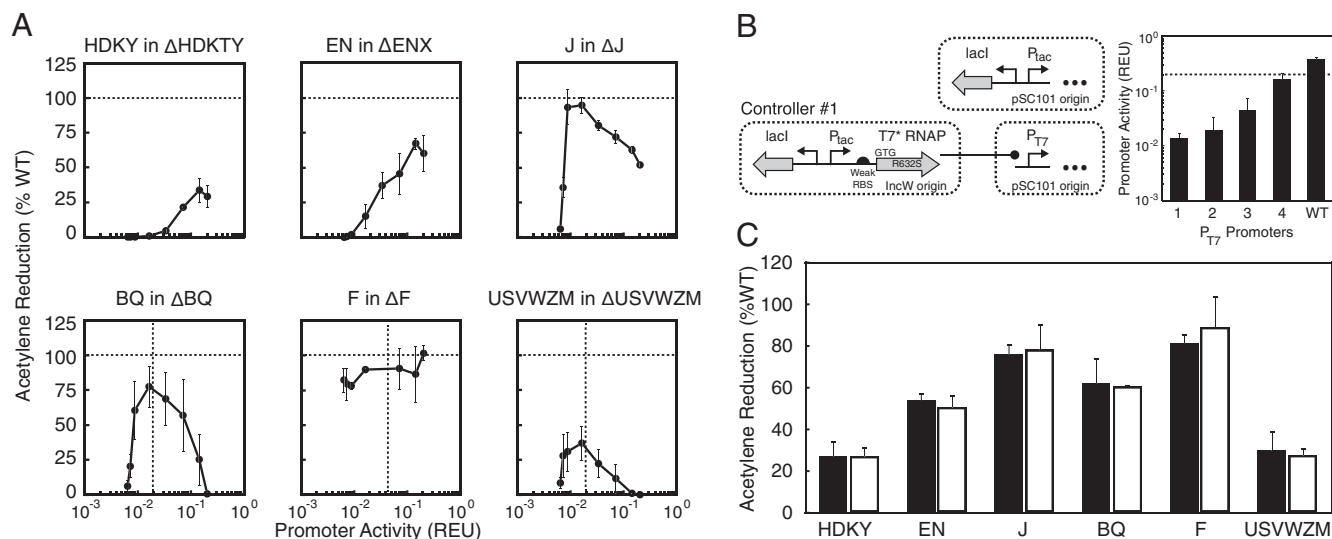
Fig. 3. Converting to T7* RNAP control. (*A*) Nitrogenase activity is shown as a function of promoter strength for each refactored operon in respective *K. oxytoca* KO strains (Δ*nifHDKTY*, Δ*nifENX*, Δ*nifJ*, Δ*nifBQ*, Δ*nifF*, and Δ*nifUSVWZM*). Vertical dashed lines indicate strength of the mutant T7 promoter that controls each operon in the complete refactored gene cluster. (*B*) A controller plasmid decouples operon expression from the inducible promoter. A T7 RNAP variant (T7* RNAP) was designed to reduce toxicity. A set of four mutated T7 promoters were used to control the expression of each operon (part numbers and sequences for mutants 1–4 are listed in *SI Materials and Methods*). P_{tac} activity under 1 mM IPTG induction is indicated by a dashed horizontal line. (*C*) Nitrogenase activity is compared for each refactored operon under the control of the P_{tac} promoter at the optimal IPTG concentration (black) and the controller with 1 mM IPTG and expression controlled by different T7 promoters (white). The T7 promoters used are $P_{T7.WT}$ for operons HDKY, EN, and J; $P_{T7.2}$ for operons BQ and USVWZM; and $P_{T7.3}$ for F. Error bars represent the SD of at least three experiments performed on different days.

Refactored Gene Cluster

Plasmid SBa_000534

P_T7.WT														
0.38	3333	42% nifH		367	51% nifD		72	51% nifK		187	45% nifY		61	48% nifE	
SBa_000443	SBa_000469	SBa_000453	SBa_000470	SBa_000454	SBa_000471	SBa_000455	SBa_000472	SBa_000456	SBa_000473	SBa_000457
1	24	52	934	984	1018	2470	2520	2554	4117	4157	4183	4846	4886	4918	6292

						P_T7.2								P_T7.3
59	47% nifN	45	48% nifJ	1.7		0.019		207	45% nifB		103	54% nifQ	2.8		0.044
SBa_000474	SBa_000475	SBa_000452	SBa_000458	SBa_000445	SBa_000459	SBa_000483	SBa_000460	SBa_000475	SBa_000451	SBa_000461	SBa_000446
6332	6363	7749	7776	11280	11320	11350	11373	11413	11452	12859	12899	12920	13424	13453	13471

			P_T7.2								
22	51% nifF	2.8		0.019		587	54% nifU		461	44% nifS		50	48% nifV		57
SBa_000482	SBa_000451	SBa_000462	SBa_000445	SBa_000463	SBa_000476	SBa_000464	SBa_000477	SBa_000465	SBa_000478	SBa_000466	SBa_000479
13494	13528	14059	14091	14109	14132	14172	14209	15034	15074	15113	16316	16356	16390	17533	17573

43% nifW		307	48% nifZ		601	45% nifM	2.2				
SBa_000467	SBa_000480	SBa_000468	SBa_000481	SBa_000450	SBa_000487	SBa_000488	SBa_000489	SBa_000490	SBa_000491	SBa_000492
17608	17866	17906	17942	18389	18429	18463	19264	19311	19446	19494	21539	22315	22407	22509

Controller

Plasmid SBa_000559

	P_BAD	P_tet		cI		P_cons		araC		tetR			
SBa_000488	SBa_000558	SBa_000493	SBa_000494	cI	SBa_000496	SBa_000497	SBa_000498	araC	SBa_000500	tetR	SBa_000491	SBa_000502	SBa_000503
1	43	377	439	456	1239	1374	1417	1435	2321	2340	2967	3276	4084	5208

Plasmid SBa_000560

		lacI	P_cons	P_tac/cI				T7* RNAP			
SBa_000491	SBa_000492	lacI	SBa_000505	SBa_000506	SBa_000508	SBa_000507	SBa_000509	T7* RNAP	SBa_000487	SBa_000488	SBa_000510	SBa_000511
1	93	196	1212	1910	1989	2029	2063	2189	4838	4973	5038	8685	9859

Legend:
- Promoter
- Spacer
- Ribosome Binding Site
- Degradation Tag
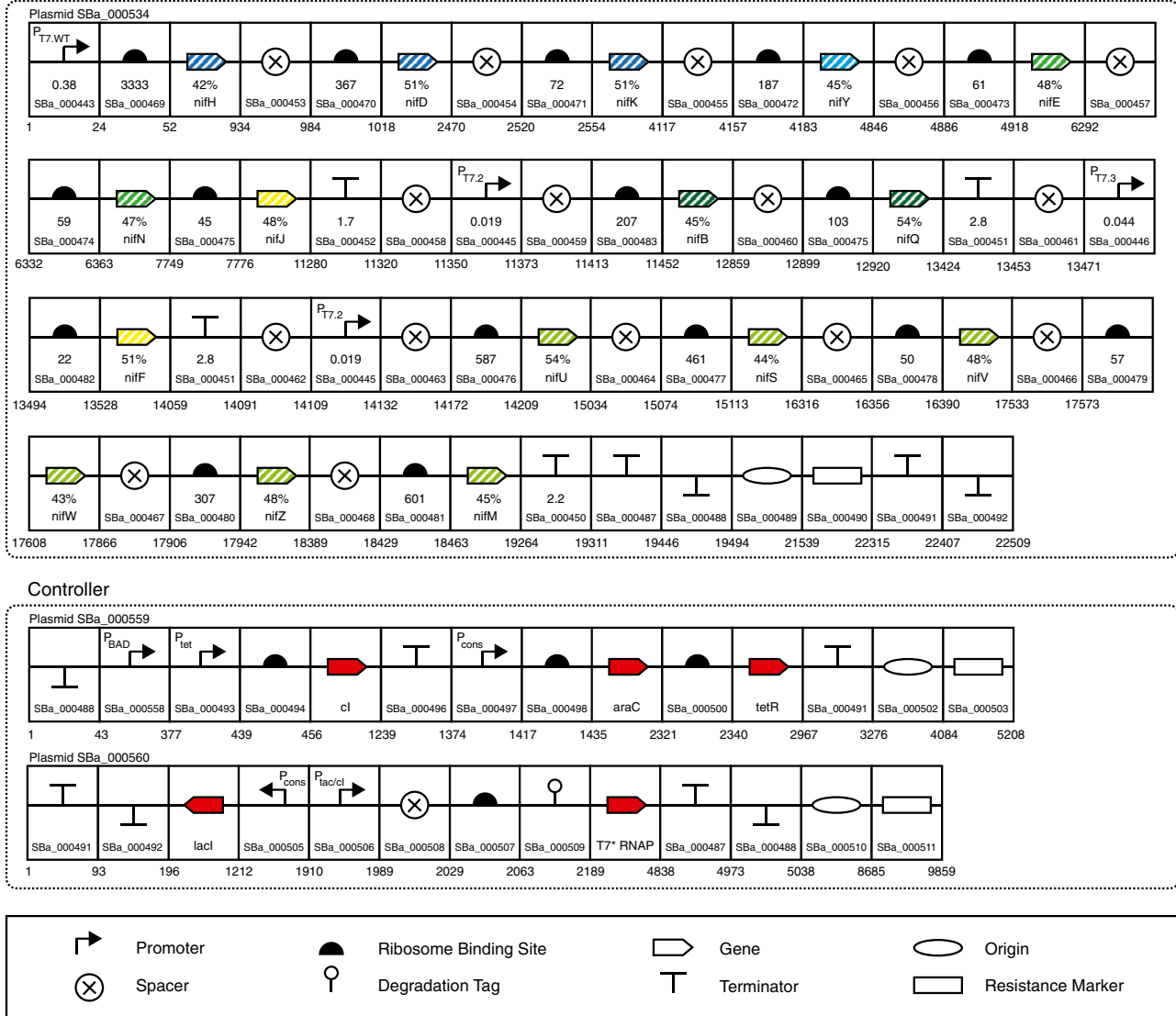- Gene
- Terminator
- Origin
- Resistance Marker

Fig. 4. Comprehensive schematic illustration for the complete refactored gene cluster and controller. Each of the 89 parts is represented according to the Synthetic Biology Open Language visual standard (www.sbolstandard.org), and the SynBERC Registry part number (registry.synberc.org) and part activity are shown. The full sequences of each plasmid have been deposited in GenBank (SBa_000534, JQ903614; SBa_000559, JQ903615; SBa_000560, JQ903616). The T7 promoter strengths are measured with monomeric red fluorescent protein and reported in REUs (*Materials and Methods*). Terminator strengths are measured in a reporter plasmid and reported as the fold reduction in monomeric red fluorescent protein (RFP) expression compared with a reporter without a terminator. The RBS strength is reported in as arbitrary units of expression from the induced P_{tac} promoter (1 mM IPTG) and a fusion gene between the first 90 nt of the gene and RFP. The nucleotide numbers for the plasmids containing the refactored cluster and controller are shown. The codon identity of each recoded gene compared with WT is shown as a percentage.

17.5 mM ammonia, no nitrogenase activity is observed for the WT cluster (Fig. 5C). In contrast, the refactored gene cluster maintains activity in the presence of ammonia ($1.1\% \pm 0.5\%$). Interestingly, this sevenfold reduction of activity is not caused by residual regulation present in the system. Rather, it occurs because the addition of ammonia to the media reduces the output of the controller by 4.5-fold (Fig. 5C). In theory, this could be fixed by increasing the expression level of T7* RNAP, but it speaks to the need to create genetic circuits that are robust to environmental context.

Discussion

The objective of refactoring is to facilitate the forward engineering of multigene systems encoded by complex genetics. Native gene clusters are the product of evolutionary processes; thus, they exhibit high redundancy, efficiency of information coding, and layers of regulation that rely on different biochemical mechanisms (36–38). These characteristics inhibit the quantitative alteration of function by part substitution because the effect can become embedded in a web of interactions. Here, modularizing the cluster, physically separating and insulating the parts, and simplifying its regulation have guided the selection and analysis of part substitutions. The information gleaned from screening the permutations in a refactored system can be cleanly fed back into the design cycle.

The refactored cluster can also serve as a platform for addressing questions in basic biology. First, it allows for the impact of regulatory interactions to be quantified in isolation. For example, in the natural system, one feedback loop could be embedded in many other regulatory loops. Systematically removing such regu-
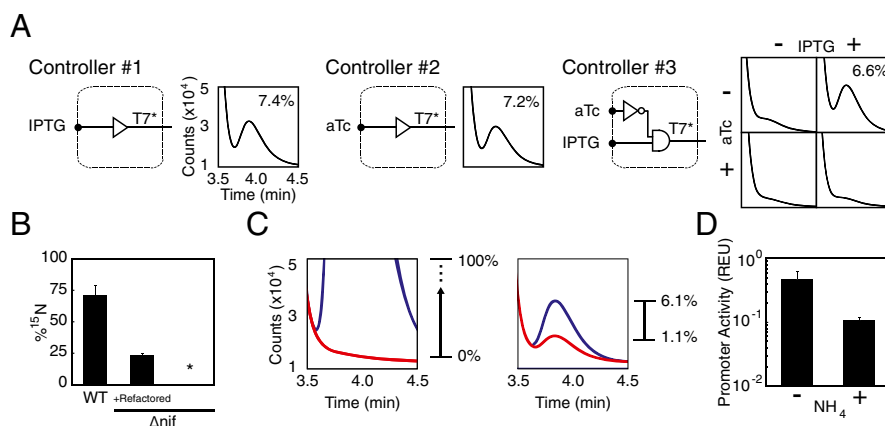
Fig. 5. Regulation of the complete refactored gene cluster. (*A*) Nitrogenase activity for the three controllers are shown: IPTG-inducible, aTc-inducible, and IPTG ANDN aTc logic. The gas chromatography trace is shown for each, as well as the calculated percent of WT activity (7.4% ± 2.4%, 7.2% ± 1.7%, and 6.6% ± 1.7% respectively). SD is calculated by using data from at least two experiments performed on different days. (*B*) Incorporation of ^{15}N into cell biomass is shown. Nitrogen fixation from N_2 gas by the refactored gene cluster was traced by using $^{15}N_2$ and measured by using isotope ratio mass spectrometry. Data are represented as the fraction of cellular nitrogen that is ^{15}N. The SD represents two experiments performed on different days. (*C*) The effect of ammonia on regulation of nitrogenase expression is shown. Acetylene reduction traces shown with (red) and without (blue) addition of 17.5 mM ammonium acetate for WT cells (*Left*) and cells bearing synthetic *nif* system (*Right*). The synthetic system was induced by controller 1 using 1 mM IPTG and exhibited nitrogenase activity of 1.1% ± 0.5% and 6.1% ± 0.4% with and without ammonium acetate, respectively. (*D*) T7* RNAP expression of controller 1 corresponding to *C* is shown. Strains carrying controller 1 and an RFP reporter plasmid were characterized under 1 mM IPTG induction with or without addition of ammonium acetate.

lation provides a clean reference system (potentially less active than WT) from which improvements can be quantified as a result of adding back regulation. It also serves as a basis for comparison of radically different regulatory programs or organizational principles, for example, to determine the importance of temporal control of gene expression (4, 39) or the need for genes to be encoded with a particular operon structure (40, 41). Second, the process of reconstruction and debugging is a discovery mechanism that is likely to reveal novel genetics and regulatory modes.

Refactoring may enable the access of functions encoded in gene clusters that are identified within sequenced genomes. With advances in DNA synthesis technology, it is possible to construct complete gene clusters and specify every nucleotide in the design. This capability eliminates the reliance on the natural physical DNA for construction and enables the simultaneous specification of every part in the system. The systematic replacement of gene regulation will be required if the cluster is silent (unexpressed in laboratory conditions) or if it needs to be transferred into a heterologous host (42). This is particularly important of the source of the DNA is unknown, for example, from a metagenomic sample.

Two relevant challenges were encountered when refactoring the nitrogenase gene cluster. First, not all of the necessary genetics will be known or there will be errors in the sequence. To address this, debugging methods will have to be developed that do not require a deep, specific understanding of the pathway. This will be aided by high-throughput part assembly techniques where many designs can be evaluated simultaneously. Second, there is a need for context-independent parts to control expression and computational methods to scan genetic designs for interfering functions. Together, these approaches will enable the rapid mining of multi-gene cellular functions from sequence databases for industrial, agricultural, and pharmaceutical applications.

Materials and Methods

Strains and Media. *E. coli* strain S17-1 was used for construction and propagation of all plasmids used in *K. oxytoca* KO mutant construction. *K. oxytoca* strain M5a1 (gift from Paul Ludden, University of California, Berkeley, CA) and mutants derived from M5al were used for nitrogen fixation experiments. Luria–Bertani/Lennox medium was used for strain propagation. All assays were carried out in minimal medium containing (per liter) 25 g Na_2HPO_3, 3g KH_2PO_4, 0.25g $MgSO_4 \cdot 7H_2O$, 1g $NaC1$, 0.1g $CaCl_2 \cdot 2H_2O$, 2.9 mg $FeCl_3$, 0.25 mg

$Na_2MoO_4 \cdot 2H_2O$, and 20 g sucrose. Growth medium is defined as minimal medium supplemented with 6 mL (per liter) of 22% (wt/vol) NH_4Ac. Derepression medium is defined as minimal medium supplemented with 1.5 mL (per liter) of 10% (wt/vol) serine. The antibiotics used were 34.4 μg·mL^{-1} chloramphenicol, 100 μg·mL^{-1} spectinomycin, 50 μg·mL^{-1} kanamycin, and/or 100 μg·mL^{-1} ampicillin.

Codon Randomization. Initial gene sequences were proposed by DNA2.0 to maximize the Hamming distance from the native sequence while seeking an optimal balance between *K. oxytoca* codon use and *E. coli* codon preferences experimentally determined by the company (43). Rare codons (<5% occurrence in *K. oxytoca*) were avoided, and mRNA structure in the translation initiation region was suppressed. Known sequence motifs, including restriction sites, transposon recognition sites, Shine–Dalgarno sequences, and transcriptional terminators, were removed by the DNA2.0 algorithm.

Elimination of Undesired Regulation. Each synthetic operon was scanned before DNA synthesis to identify and remove undesired regulation. Multiple types of regulation were identified using publicly available software. The RBS Calculator was used (Reverse Engineering mode; 16S RNA:ACCTCCTTA) to identify RBSs throughout the proposed DNA sequence of the operon (44). The Prokaryotic Promoter Prediction server was used to identify putative σ70 promoter sites (e-value cutoff of 5, sigma.hmm database) (45). The PromScan algorithm was used to identify putative σ54 promoter sites using default options (46). TransTermHP software was used with default parameters to identify terminator sequences in both the forward and reverse directions (47). RBSs greater than 50 AU and all identified promoters and terminators were considered significant.

Nitrogenase Activity Assay. In vivo nitrogenase activity is determined by acetylene reduction as previously described (48). For *K. oxytoca* whole-cell nitrogenase activity assay, cells harboring the appropriate plasmids were incubated in 5 mL of growth media (supplemented with antibiotics, 30 °C, 250 rpm, New Brunswick Scientific, Innova 44 incubator shaker) in 50-mL conical tubes for 14 h. The cultures were diluted into 2 mL derepression media (supplemented with antibiotics and inducers) to a final OD$_{600}$ of 0.5 in 14-mL bottles, and bottles were sealed with rubber stoppers (Z564702; Sigma). Headspace in the bottles was repeatedly evacuated and flushed with N_2 past a copper catalyst trap by using a vacuum manifold. After incubating the cultures for 5.5 h at 30 °C, 250 rpm in a New Brunswick Scientific, Innova 3100 water bath shaker, headspace was replaced by 1 atm Ar. Acetylene was generated from CaC_2 by using a Burris bottle, and 1 mL was injected into each bottle to start the reaction. Cultures were incubated for 1 h at 30 °C, 250 rpm in a New Brunswick Scientific, Innova 3100 water bath shaker, before the assay was stopped by injection of 300 μL of 4 M NaOH solution into

each bottle. To quantify ethylene production, 50 μL of culture headspace was withdrawn through the rubber stopper with a gas tight syringe and manually injected into a HP 5890 gas chromatograph. Nitrogenase activity is reported as a percentage of WT activity. Briefly, ethylene production by strains was quantified by integrating area under the peak by using Chem-Station software and dividing ethylene production of experimental strains by the ethylene production of a WT control included in each assay.

N_2-Dependent Growth and $^{15}N_2$ Incorporation Assay. Nitrogen fixation by synthetic *nif* cluster in *K. oxytoca* is further demonstrated by N_2-dependent growth and $^{15}N_2$ incorporation. Cells are diluted as described in the acety-

lene reduction assay. The headspace of the bottles is replaced by normal N_2 gas or by stable isotope nitrogen, $^{15}N_2$ (^{15}N atom 99.9%; catalog no. IN 5501; Icon Isotopes). After incubating the cultures for 36 h at 30 °C, 250 rpm in a New Brunswick Scientific, Innova 3100 water bath shaker, N_2-dependent growth of the cells is determined by measuring optical density at 600 nm (i.e., OD_{600}). To do the $^{15}N_2$ incorporation assay, the ^{15}N-enriched cells with corresponding control cultures under normal nitrogen gas are collected by centrifugation, and the cell pellets are dried in a laboratory oven at 100 °C for 12 h. The dried pellets are analyzed for $^{15}N/^{14}N$ ratio at the Center for Stable Isotope Biogeochemistry at the University of California, Berkeley, by using the Finnigan MAT Delta plus Isotope Ratio Mass Spectrometer.

1. Fischbach M, Voigt CA (2010) Prokaryotic gene clusters: A rich toolbox for synthetic biology. *Biotechnol J* 5:1277–1296.
2. Ishihama A (2010) Prokaryotic genome regulation: Multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol Rev* 34:628–645.
3. Mandal M, Breaker RR (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* 5:451–463.
4. Temme K, et al. (2008) Induction and relaxation dynamics of the regulatory network controlling the type III secretion system encoded within Salmonella pathogenicity island 1. *J Mol Biol* 377:47–61.
5. Georg J, Hess WR (2011) cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev* 75:286–300.
6. Güell M, Yus E, Lluch-Senar M, Serrano L (2011) Bacterial transcriptomics: What is beyond the RNA horiz-ome? *Nat Rev Microbiol* 9:658–669.
7. Johnson ZI, Chisholm SW (2004) Properties of overlapping genes are conserved across microbial genomes. *Genome Res* 14:2268–2272.
8. Zazopoulos E, et al. (2003) A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat Biotechnol* 21:187–190.
9. Medema MH, Breitling R, Bovenberg R, Takano E (2011) Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat Rev Microbiol* 9:131–137.
10. Gottelt M, Kol S, Gomez-Escribano JP, Bibb M, Takano E (2010) Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in Streptomyces coelicolor A3(2). *Microbiology* 156:2343–2353.
11. Lombó F, Braña AF, Méndez C, Salas JA (1999) The mithramycin gene cluster of Streptomyces argillaceus contains a positive regulatory gene and two repeated DNA sequences that are located at both ends of the cluster. *J Bacteriol* 181:642–647.
12. Medema MH, Breitling R, Takano E (2011) Synthetic biology in Streptomyces bacteria. *Methods Enzymol* 497:485–502.
13. Pickens LB, Tang Y, Chooi Y-H (2011) Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.* 2:1–26.
14. Smanski MJ, Peterson RM, Rajski SR, Shen B (2009) Engineered Streptomyces platensis strains that overproduce antibiotics platensimycin and platencin. *Antimicrob Agents Chemother* 53:1299–1304.
15. Biggins JB, Liu X, Feng Z, Brady SF (2011) Metabolites from the induced expression of cryptic single operons found in the genome of Burkolderia pseudomallei. *JACS* 133:1638–1641.
16. Watanabe K, et al. (2006) Total biosynthesis of antitumor nonribosomal peptides in Escherichia coli. *Nat Chem Biol* 2:423–428.
17. Fowler M, Beck K (1999) *Refactoring: Improving the Design of Existing Code* (Addison-Wesley, Reading, MA).
18. Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. *Mol Syst Biol* 1:2005.0018.
19. Czar MJ, Anderson JC, Bader JS, Peccoud J (2009) Gene synthesis demystified. *Trends Biotechnol* 27:63–72.
20. Purnick PE, Weiss R (2009) The second wave of synthetic biology: From modules to systems. *Nat Rev Mol Cell Biol* 10:410–422.
21. Stacey GS, Burris RH, Evans HJ (1992) *Biological Nitrogen Fixation* (Chapman and Hall, New York).
22. Burris RH (1991) Nitrogenases. *J Biol Chem* 266:9339–9342.
23. Hu Y, Fay AW, Lee CC, Yoshizawa J, Ribbe MW (2008) Assembly of nitrogenase MoFe protein. *Biochemistry* 47:3973–3981.
24. Rubio LM, Ludden PW (2005) Maturation of nitrogenase: A biochemical puzzle. *J Bacteriol* 187:405–414.
25. Arnold W, Rump A, Klipp W, Priefer UB, Pühler A (1988) Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae. *J Mol Biol* 203:715–738.
26. Dixon R, Kahn D (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* 2:621–631.
27. Dixon RA, Postgate JR (1972) Genetic transfer of nitrogen fixation from Klebsiella pneumoniae to Escherichia coli. *Nature* 237:102–103.
28. Simon HM, Homer MJ, Roberts GP (1996) Perturbation of nifT expression in Klebsiella pneumoniae has limited effect on nitrogen fixation. *J Bacteriol* 178:2975–2977.
29. Fani R, Gallo R, Liò P (2000) Molecular evolution of nitrogen fixation: The evolutionary history of the nifD, nifK, nifE, and nifN genes. *J Mol Evol* 51:1–11.
30. Gosink MM, Franklin NM, Roberts GP (1990) The product of the Klebsiella pneumoniae nifX gene is a negative regulator of the nitrogen fixation (nif) regulon. *J Bacteriol* 172:1441–1447.
31. Orme-Johnson WH (1985) Molecular basis of biological nitrogen fixation. *Annu Rev Biophys Biophys Chem* 14:419–459.
32. Kelly JR, et al. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 3:4.
33. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–345.
34. Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci USA* 99:16587–16591.
35. Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 469:212–215.
36. Alon U (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, Boca Raton, FL).
37. Kitano H (2002) Systems biology: A brief overview. *Science* 295:1662–1664.
38. Palsson B (2006) *Systems Biology: Properties of Reconstructed Networks* (Cambridge Univ Press, New York).
39. Kalir S, et al. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 292:2080–2083.
40. Zaslaver A, Mayo A, Ronen M, Alon U (2006) Optimal gene partition into operons correlates with gene functional order. *Phys Biol* 3:183–189.
41. Kovács K, Hurst LD, Papp B (2009) Stochasticity in protein levels drives colinearity of gene order in metabolic operons of Escherichia coli. *PLoS Biol* 7:e1000115.
42. Wenzel SC, Müller R (2005) Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr Opin Biotechnol* 16:594–606.
43. Welch M, et al. (2009) Design parameters to control synthetic gene expression in Escherichia coli. *PLoS ONE* 4:e7002.
44. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27:946–950.
45. Zomer AL (2011) PPP: Perform Promoter Prediction. Available at http://bioinformatics.biol.rug.nl/websoftware/ppp/ppp_start.php. Accessed June 1, 2010.
46. Studholme D (2011) PromScan Available at http://molbiol-tools.ca/promscan/. Accessed June 1, 2010.
47. Kingsford CL, Ayanbule K, Salzberg SL (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8:R22.
48. Stewart WD, Fitzgerald GP, Burris RH (1967) In situ studies on nitrogen fixation with the acetylene reduction technique. *Science* 158:536.

Supporting Information

Temme et al. 10.1073/pnas.1120788109

SI Materials and Methods

***Klebsiella oxytoca* KO Strains.** All *K. oxytoca* mutants are constructed from M5al by allele exchange by using suicide plasmid pDS132 [gift from Dominique Schneider at Universite Joseph Fourier (Grenoble, France)] carrying the corresponding *nif* gene deletion (1). We made a slight modification to a previously published protocol (2). Here, a kanamycin resistance cassette was cloned into the suicide plasmid upstream of the left homologous exchange fragment. These operon deletions in *nif* gene cluster span the promoter and the complete amino acid coding sequences except when specifically designated (Fig. S1). All mutants were verified by DNA sequencing of the PCR product of the corresponding gene region to confirm physical DNA deletion and by whole-cell acetylene reduction assay to confirm the lack of nitrogenase activity (Table S1).

Promoter Characterization. *Relative expression units.* In this manuscript, the output of promoters is reported as relative expression units (REU). This is simply a linear factor that is multiplied by the arbitrary units measured by the flow cytometer. The objective of normalizing to REUs is to standardize measurements between laboratories and projects. The linear factor is 1.66×10^{-5}, and the division by this number back-converts to the raw arbitrary units. This number was calculated to be a proxy to the relative promoter units (RPUs) reported by Kelly et al. (3). Our original standardized measurements were made before the publication of the study of Kelly et al. (3) and involved a different reference promoter, monomeric red fluorescent protein (mRFP), ribosome binding sites (RBSs), and plasmid backbone. Because of these differences, one cannot calculate RPUs as defined by Kelly et al. Instead, a series of plasmids was made (Fig. S2A) to estimate the relative expression of reporter protein from experimental constructs compared the standard construct in the work of Kelly et al. (3). Conversion factors between constructs were measured and multiplied to obtain the linear factor described here. We renamed the unit REUs because it is intended to be a simple normalization of fluorescent units (akin to a fluorescent bead) and not a direct measurement of the activity of a promoter (e.g., polymerase flux).

Promoter characterization assay. Cells were grown as in the acetylene reduction assay with two modifications. The initial flush of headspace with N_2 was not performed, and the assay was halted after the 5.5-h incubation. To halt the assay, 10 μL of cells were transferred from each bottle to a 96-well plate containing PBS solution supplemented with 2 mg·mL^{-1} kanamycin. Fluorescence data were collected by using a BD Biosciences LSRII flow cytometer. Data were gated by forward and side scatter, and each data set consisted of at least 10,000 cells. FlowJo software was used to calculate the geometric means of the fluorescence distributions. The autofluorescence value of *K. oxytoca* cells harboring no plasmid was subtracted from these values to give the values reported in this study. The strengths of T7 promoter mutants were characterized by swapping them in place of the P$_{tac}$ promoter in plasmid N149 (SBa_000516), cotransforming with controller 1 (plasmid N249), and measuring fluorescence via flow cytometry under 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) induction.

Replacement of P$_{tac}$ promoters by T7 promoters. To replace the P$_{tac}$ promoter by a T7 promoter in each synthetic operon, we followed a simple process. First, we identified the IPTG concentration corresponding to the maximal functional activity of each synthetic operon. Second, we translated this IPTG concentration into REUs based on characterization of the P$_{tac}$ promoter (Fig. S2B, *Left*). Third, we selected the T7 mutant promoter with the closest strength in REUs. For the synthetic *nifF* operon, we observed broad, robust fixation under the P$_{tac}$ promoter. We found that T7 mut 3 produced inducible functional activity with a maximum at 1 mM IPTG induction of the T7 RNAP. For the synthetic *nifJ* operon, our method suggests that we use a weak T7 mutant promoter. However, we found that a WT T7 promoter produced inducible activity with a maximum at 1 mM IPTG. We attribute this deviation to a change in RBS strength caused by contextual differences between P$_{tac}$ and the T7 promoter.

Debugging Synthetic Operons. Some of the initial designs for refactored operons showed little or no activity. When this occurs, it is challenging to identify the problem because so many genetic changes have been made simultaneously to the extent that there is almost no DNA identity with the WT sequence. To rapidly identify the problem, a debugging method was developed that can be generalized when refactoring different functions (Fig. S3A). Chimeric operons are created by replacing a WT region of DNA with its synthetic counterpart. The function of each chimera in this library is assessed to identify which region of synthetic DNA caused a loss of activity. New chimeras are then constructed with increasingly fine resolution changes between synthetic and WT DNA. This approach "zooms in" on the problematic region of DNA, which can then be fixed. The most common problem is caused by errors in the reference DNA sequence (GenBank, X13303.1) (4). Refactored genes were designed using only the amino acid sequence information from the database; thus, they were sensitive to sequencing errors leading to missense mutations that reduced or eliminated activity. Indeed, 18 such mutations were identified and confirmed by carefully resequencing the WT cluster (Table S2). Fifteen of the 18 mutations occurred in refactored operons that required debugging and were corrected (Fig. S3B). This demonstrates the challenge of reconstituting biological functions by using only database information and DNA synthesis (5).

Modifying synthetic RBS strength was also important to debugging. The function of the synthetic *nifUSVWZM* operon was significantly improved by changing RBSs to match a 1:1 ratio of NifU:NifS. The initial selection of RBSs led to an observed 10:1 ratio in their respective RBS strengths. After debugging, *nifU* and *nifS* RBS strength was better balanced (1.25:1), and this improved activity. For one RBS, the measurement method proved to be inaccurate. We found the measured strength of the WT *nifQ* RBS was extremely low (Fig. 2C), and the synthetic *nifBQ* operon showed low activity when the synthetic *nifQ* RBS was matched to the measured strength. In contrast, the robustness analysis showed a requirement for high expression level of the *nifBQ* operon (Fig. 2B). Thus, a strong synthetic RBS near the strength of the *nifB* RBS was used and significantly improved *nifBQ* operon activity. In one case, our initial recoded *nifH* gene did not express well by using WT or synthetic regulation (Fig. S5). We designed a new synthetic gene, requiring that it diverge in DNA sequence from the native and first synthetic DNA sequences, and found that the new synthetic gene expressed well and recovered activity.

Growth by Nitrogen Fixation. Cells capable of nitrogen fixation should exhibit measurable growth on media that lacks nitrogen by using atmospheric N_2 as a source of nitrogen. Conversely,

cells incapable of nitrogen fixation should not grow on nitrogen-free media.

In parallel to the $^{15}N_2$ incorporation assay, we monitored strain growth under nitrogen-limited media conditions and 100% $^{15}N_2$ atmosphere (*Materials and Methods*, N$_2$-Dependent Growth and $^{15}N_2$ Incorporation Assay). Cells were grown on derepression media as used in the nitrogenase activity assay. Depression media is not strictly nitrogen-free, containing 1.43 mM serine to promote ribosomal RNA production and hasten nitrogenase biosynthesis (6).

Strains containing controller 1 and the refactored gene cluster grew nearly 30% as much as WT strains. In contrast, minimal growth was observed in Δ*nif* strains, consistent with the limited nitrogen available from serine and cell lysis products (7).

Western Blot Assay for Synthetic *nifH* Expression. The first synthetic *nifHDK* did not exhibit nitrogenase activity under induction ranging from 0 to 1 mM IPTG, and the *nifH* gene (synthetic *nifH*$_{v1}$) was identified as a problematic part by using the debugging protocol shown in Fig. S3. However, there was no mutation found. Western blots were further used to confirm low synthetic *nifH* expression.

A Western blot for NifH protein in Fig. S5 (*Left*) showed that WT *nifH* expressed well with synthetic *nifD* or *nifK* (constructs N10, N12, N14), whereas synthetic *nifH*$_{v1}$ was not expressed regardless of the context of *nifDK* (constructs N1 and N19). A second synthetic *nifH* (synthetic *nifH*$_{v2}$) was used to replace synthetic *nifH*$_{v1}$. The Western blot in Fig. S5 (*Right*) showed the synthetic *nifH*$_{v2}$ (construct N38) expressed well.

Samples for Western blots were prepared by boiling collected *K. oxytoca* cells in SDS/PAGE loading buffer and run on 12% SDS-polyacrylamide gels (Lonza). Proteins on the gels were transferred to PVDF membranes (cat no. 162–0177; BioRad) by using Trans-Blot SD Semi-Dry Transfer Cell (cat no. 170–3940; BioRad). Blocking the membrane and antibody binding were performed by using a SNAP i.d. Protein Detection System (cat no. WBAVDBA; Millipore). The membranes were blocked by Tris-buffered saline solution/Tween 20 with 1% BSA. The anti-NifH and anti-NifDK antibodies (provided by the Paul Ludden laboratory, University of California, Berkeley, CA) were used as the primary antibodies. The anti-NifH antibody was a universal anti-NifH made against a mixture of purified NifH proteins from *Azotobacter vinelandii*, *Clostridium pasteurianum*, *Rhodospirillum rubrum*, and *K. oxytoca*. The anti-NifDK antibody was made against purified NifDK protein from *A. vinelandii*. The anti-NifH and anti-NifDK antibodies were used at 1:500 and 1:2,000 respectively. The secondary antibody (goat anti-rabbit IgG-HRP; catalog no. A0545; Sigma) was used at 1:10,000. Development was done using an enhanced chemiluminescent substrate for HRP (catalog no. 32209; Pierce) and captured on film (catalog no. 178–8207; Kodak).

Construction of Plasmids and Parts. *Plasmid construction.* Plasmids were designed in silico. Synthetic parts (promoters, RBSs, terminators, and spacers) were combined with the initial synthetic gene sequences proposed by DNA2.0 in A Plasmid Editor (http://biologylabs.utah.edu/jorgensen/wayned/ape/) and GeneDesigner (8) to create synthetic operons. Synthetic operons were computationally scanned to eliminate unintended regulation (*Materials and Methods*, "Elimination of Undesired Regulation"), and parts containing such regulation were replaced. This iterative process continued until the synthetic operons included only designed regulation.

Physical DNA was constructed using standard manipulation techniques. Assembly methods followed published protocols and included BioBrick (9), MEGAWHOP (10), Phusion Site-Directed Mutagenesis, or Gibson Assembly methods (11). We found that Gibson Assembly was the most efficient DNA assembly method, except when making small (i.e., <10 bp) changes

in plasmids smaller than 10 kb in size. We noted assembly failures were infrequent, more common in assemblies greater than 15 kbp, and linked to the presence of homology within approximately 500 bp of part termini. In these cases, we observed annealing of unexpected parts to create nonintended junctions.

Plasmid pIncW (pSa, SpR) was generated from pEXT21 (pSa, SpR) by deletion of *osa*, *nuc1*, the Tn21 integrase gene, and ORF18 (12).

Plasmid pSB4C5 (pSC101, CmR) was obtained from the Registry of Standard Biological Parts and serves as the base vector for wild-type complementation, RBS characterization, and synthetic operons (9).

Plasmid N58 (pSC101, CmR) was generated by inserting the P$_{tac}$ cassette (SynBERC Registry, SBa_000561) between the BioBrick prefix and BioBrick suffix of pSB4C5.

Plasmid N292 (SBa_000566) was generated by inserting a terminator characterization cassette between the BioBrick prefix and BrioBrick suffix of pSB4C5. The cassette consists of the PT7 promoter, RBS (SBa_000498), GFP, the WT T7 terminator, RBS D103 (SBa_000563) from Salis et al. (13), and mRFP (SBa_000484).

Plasmid N149 (SBa_000516) was constructed by inserting the P$_{tac}$ promoter cassette (SBa_000563), RBS D103 (SBa_000563) from Salis et al. (13), and mRFP (SBa_000484) between the BioBrick prefix and BioBrick suffix of pSB4C5.

Plasmid N505 (SBa_000517) was constructed by inserting the P$_{tet}$ promoter cassette (SBa_000562), RBS D103 (SBa_000563), and mRFP (SBa_000484) between the BioBrick prefix and BioBrick suffix of pSB4C5.

Plasmid N110 (SBa_000564) was constructed by inserting a constitutive promoter (SBa_000565), a strong RBS (SBa_000475), and mRFP (SBa_000484) between the BioBrick prefix and BioBrick suffix of pSB4C5.

Plasmid N573 (SBa_000559) was constructed by inserting the AmpR resistance marker in pNOR1020 (14).

WT complementation plasmids. It has been shown that the multicopy expression of some nitrogen fixation genes can eliminate nitrogenase maturation and function (i.e., multicopy inhibition) (15, 16). An additional uncertainty is that the replacement of the native promoter with an inducible promoter could disrupt their function. To examine these effects, we constructed plasmids to complement the activities of the KO strains (Fig. S1) and tested their activity under inducible control. These plasmids are also the basis for the experiments to quantify the robustness to changes in expression (Fig. 2).

Complementation plasmids were constructed by inserting the DNA encoding each WT operon between the P$_{tac}$ promoter and BioBrick suffix of plasmid N58 (pSC101, CmR). One exception was plasmid Nif18, which was constructed by cloning the *nifHDKTY* operon into the multicloning site of pEXT21 (12). WT operon sequences were defined by published transcription initiation sites (17).

RBS characterization vectors. WT RBS characterization vectors were constructed by inserting the region from −60 bp to +90 bp for each native gene and mRFP (SBa_000484) between the P$_{tac}$ cassette (SBa_000561) and the BioBrick suffix of plasmid N58 (pSC101, CmR). The native gene sequence from +1 bp to +90 bp formed an in-frame fusion with mRFP. In cases in which the gene transcript does not extend to −60 bp, a shorter cassette was cloned into N58. RBS strength was characterized by using the Promoter Characterization Assay.

Synthetic RBSs of sufficient length to capture the full ribosome footprint (~35 bp) were generated with the RBS Calculator (13). The strength of each was measured using a synthetic RBS characterization vector. These vectors were constructed similar to the WT RBS characterization vectors by using −60 bp to +90 bp of the designed synthetic gene. This region includes part of a buffer sequence, the synthetic RBS, and the region from +1 bp

to +90 bp of the synthetic gene. If the synthetic and WT RBSs differed by more than threefold in expression, new RBS sequences were generated and screened.

Synthetic operon vectors. Synthetic operons were cloned into the pSB4C5 (pSC101, CmR) backbone between the BioBrick prefix and BioBrick suffix.

Synthetic part generation. T7* RNA polymerase. The T7 RNA polymerase was modified to be nontoxic to *Klebsiella* and *Escherichia coli* at high expression levels. The RNAP was expressed from a low-copy origin (pSa) under control of a weak RBS (SBa_000507, TATCCAAACCAGTAGCTCAATTGGAGTCGTCTAT) and N-terminal degradation tag (SBa_000509, TTGTTTATCAAG-CCTGCGGATCTCCGCGAAATTGTGACTTTTCCGCTATT-TAGCGATCTTGTTCAGTGTGGCTTTCCTTCACCGGCA-GCAGATTACGTTGAACAGCGCATCGATCTGGGTGGC). The start codon was changed from ATG to GTG, and the active site contained a mutation (R632S).

T7 promoters. T7 promoters were generated from a random library. The T7 promoter seed sequence was TAATACGACT-CACTANNNNNAGA. Table S3 provides sequences of individual promoters.

T7 terminators. T7 terminators were generated from a random library and inserted into the terminator characterization vector N292 (SBa_000566). The T7 terminator seed sequence was TANNNAACCSSWWWSSNSSSSTCWWWCGSSSSSSWWSSG-TTT. Terminator plasmids were cotransformed with plasmid N249 and characterized under 1 mM IPTG induction of T7* RNAP. RFP expression was measured for each terminator, and data are reported as the fold reduction in measured fluorescence compared with a derivative of N292 carrying no terminator. Table S3 provides the sequences of individual terminators.

RBSs. The RBS Calculator was used to generate an RBS that matched the measured strength of the WT RBS. In three cases, synthetic RBSs were selected from existing parts (SBa_000475 for *nifJ* and *nifQ*, and SBa_000469 for *nifH*). In cases in which the strength of the initial synthetic RBS differed from the WT RBS by more than threefold (*nifV*, *nifZ*, and *nifM*), a library of synthetic RBS was constructed by replacing the 15 bp upstream of the start codon with NNNAGGAGGNNNNNN. We screened

mutants in each library to identify synthetic RBSs within threefold of the WT RBS strength. RBS strength is reported in arbitrary fluorescence units measured by using the fluorescence characterization assay.

Spacer sequences. Spacer sequences were generated using the Random DNA Generator (http://www.faculty.ucr.edu/~mmaduro/random.htm) using a random GC content of 50%.

ANDN logic. We constructed a genetic circuit encoding the logic A ANDN B and used this circuit to control T7* RNAP in controller 3. In this circuit, the A ANDN B logic corresponds to the presence or absence of the inducers, IPTG and aTc, such that the cell computes IPTG ANDN aTc. The circuit was constructed by modifying the P_{tac} promoter in controller 1 (SBa_000520) to include the cI repressor binding sites OR1 and OR2 to produce plasmid N639 (SBa_000560). Additionally, plasmid pNOR1020 encodes the repressor cI under control of the P_{tet} promoter (14). We modified pNOR1020 by changing the resistance marker to confer ampicillin resistance to produce N573 (SBa_000559). When N639 and N573 are cotransformed, they produce the logic circuit IPTG ANDN aTc.

P_{tac} (SBa_000512) sequence was tattctgaaatgagctgttgacaatta-atcatcggctcgtataatgtgtggaattgtgagcggataacaatt. P_{tac} plus OR1 and OR2 (SBa_000506) sequence was tattaacaccgtgcgtgttgacagctata-cctctggcggttataatgctagcggaattgtgagcggataacaatt.

Plasmid Maps. Fig. S6 shows maps for key plasmids. Synthetic Biology Open Language graphical notation is used to describe genetic parts.

Native Gene Cluster Resequencing. The *nif* gene cluster in *K. oxytoca* Ma5L was resequenced from PCR fragments. The resequenced DNA sequence was compared with the reference sequence from GenBank (X13303.1) (4). Sequence differences are listed in Table S2. The nucleotide locations are numbered relative to X13303.1. Amino acid mutations to correct errors in the X13303.1 record are shown (Impact).

Part Sequences. DNA sequences of synthetic parts are provided in Table S3.

1. Philippe N, Alcaraz JP, Coursange E, Geiselmann J, Schneider D (2004) Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. *Plasmid* 51:246–255.
2. Zhao D, Curatti L, Rubio LM (2007) Evidence for nifU and nifS participation in the biosynthesis of the iron-molybdenum cofactor of nitrogenase. *J Biol Chem* 282: 37016–37025.
3. Kelly JR, et al. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 3:4.
4. Arnold W, Rump A, Klipp W, Priefer UB, Pühler A (1988) Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae. *J Mol Biol* 203:715–738.
5. Bayer TS, et al. (2009) Synthesis of methyl halides from biomass using engineered microbes. *J Am Chem Soc* 131:6508–6515.
6. Jacob GS, Schaefer J, Garbow JR, Stejskal EO (1987) Solid-state NMR studies of Klebsiella pneumoniae grown under nitrogen-fixing conditions. *J Biol Chem* 262: 254–259.
7. Mason CA, Hamer G (1987) Cryptic growth in Klebsiella-pneumoniae. *Appl Microbiol Biotechnol* 25:577–584.
8. Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S (2006) Gene Designer: A synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* 7:285.

9. Shetty RP, Endy D, Knight TF, Jr. (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2:5.
10. Miyazaki K (2003) Creating random mutagenesis libraries by megaprimer PCR of whole plasmid (MEGAWHOP). *Methods Mol Biol* 231:23–28.
11. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–345.
12. Dykxhoorn DM, St Pierre R, Linn T (1996) A set of compatible tac promoter expression vectors. *Gene* 177:133–136.
13. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27:946–950.
14. Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 469:212–215.
15. Buck M, Cannon W (1987) Frameshifts close to the Klebsiella pneumoniae nifH promoter prevent multicopy inhibition by hybrid nifH plasmids. *Mol Gen Genet* 207: 492–498.
16. Riedel GE, Brown SE, Ausubel FM (1983) Nitrogen fixation by Klebsiella pneumoniae is inhibited by certain multicopy hybrid nif plasmids. *J Bacteriol* 153:45–56.
17. Beynon J, Cannon M, Buchanan-Wollaston V, Cannon F (1983) The nif promoters of Klebsiella pneumoniae have a characteristic primary structure. *Cell* 34:665–671.
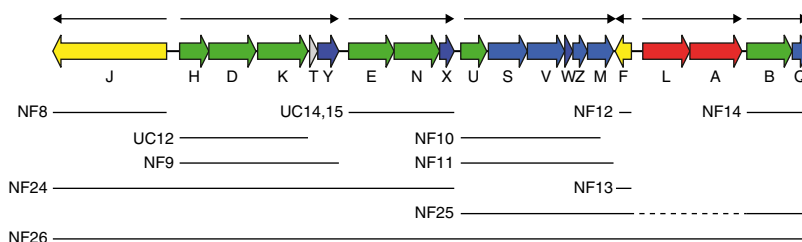
Fig. S1. *nif* operon deletions used in this study. The solid lines show the region of deleted *nif* operons. The dashed line in NF25 shows the retained *nifLA* operon.
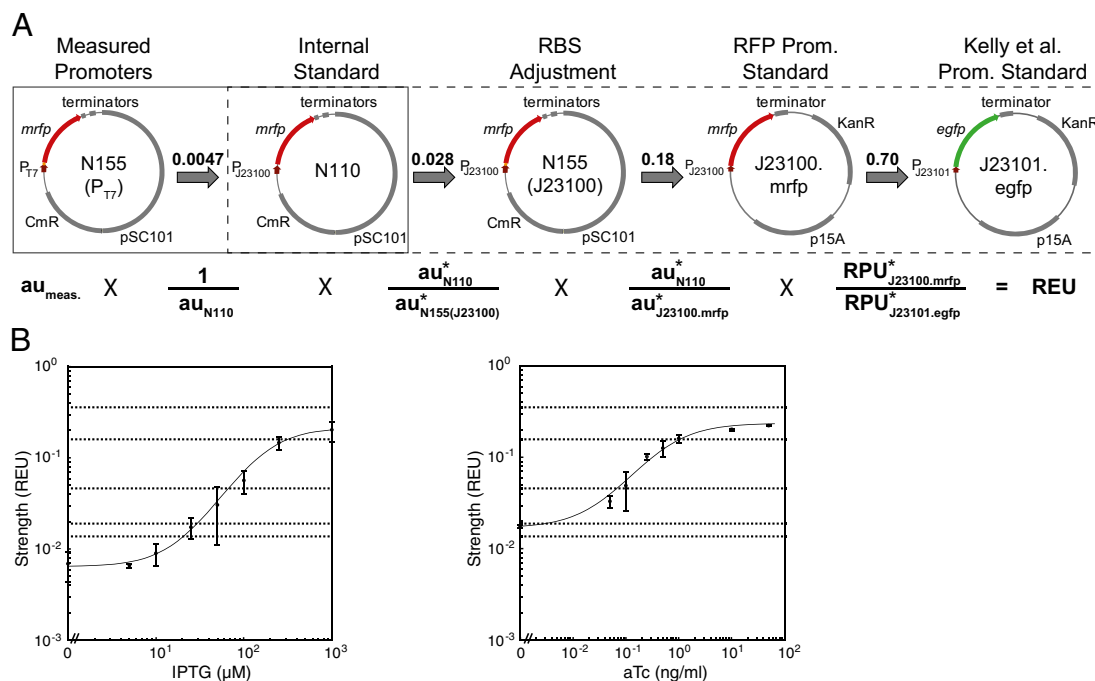

Fig. S2. Promoter characterization using relative expression units. (*A*) Conversion of arbitrary units into REUs. Promoters were characterized using mRFP1 fluorescent reporter protein in N155 (measured promoters). Data were first normalized by the fluorescence of N110 (internal standard) and then scaled by the fluorescence of N155 (J23100) to account for RBS differences between N155 and N110 (RBS adjustment). To compare our measurements to expression levels of the Kelly et al. standards (3), we further multiplied by the ratio of N110 fluorescence to the fluorescence of a Kelly et al. standard plasmid expressing mRFP1 (RFP promoter standard). A final conversion factor is applied to compare all measurements to the Kelly et al. J23101-EGFP promoter standard (3) based on a strong linear correlation of promoter strength (in RPUs) between constructs expressing mRFP and EGFP. Solid and dashed boxes were drawn to indicate which plasmids were measured at different facilities. Asterisked and nonasterisked units were measured in different facilities and correspond to the conversion factors directly above. (*B*) Promoter characterization for P_{tac} promoter (*Left*) and P_{tet} promoter (*Right*). The promoter strengths of P_{tac} promoter and P_{tet} promoter were measured under varied concentrations of inducers (IPTG or aTc). The strengths of T7 promoters (WT and mutants; Fig. 3*B*) are shown as horizontal dotted lines.
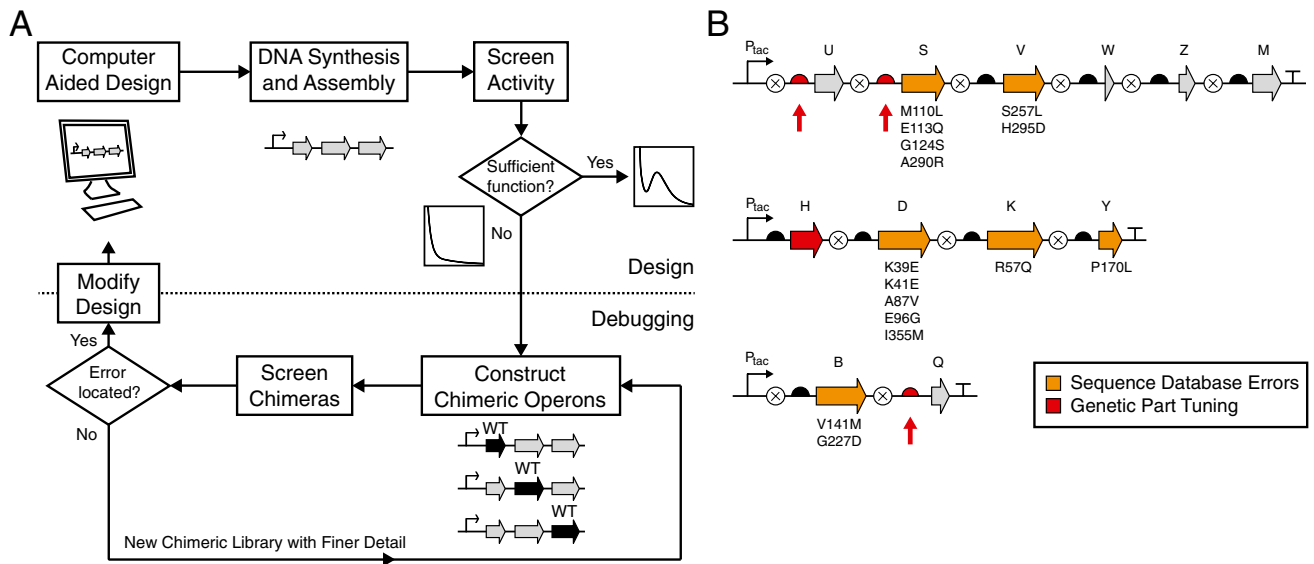
Fig. S3. Debugging refactored operons. (*A*) The process is shown for the identification of problem sequences within a refactored operon. After design and synthesis, the problematic DNA is crossed with WT to create a chimeric library, which is screened. This is done iteratively to reduce the size of the problematic region until the specific errors are identified. (*B*) The debugging process led to the correction of RBS strengths (red arrows), the recoded sequence of *nifH*, and numerous nucleotide errors found in the sequenced cluster in the database. Amino acid mutations to correct errors in the synthetic sequence are shown.
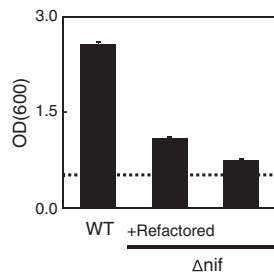

Fig. S4. Cell growth supported by nitrogen fixation. The dotted line indicates initial seeding density of OD_{600} 0.5. WT *Klebsiella* grew to an OD_{600} of 2.57 ± 0.07 after 36 h of incubation in depression conditions. Eliminating the full *nif* cluster severely inhibited cell growth (*Δnif*, OD_{600} 0.76 ± 0.02). Complementing the KO strain with the refactored cluster and controller 1 under 1 mM IPTG induction yielded growth of OD_{600} 1.10 ± 0.03.
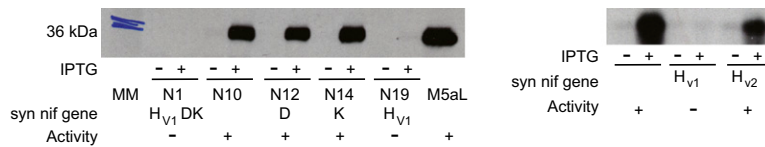

Fig. S5. Expression of synthetic *nifH* variants. Western blot assay to detect the expression of synthetic $nifH_{v1}$ (*Left*) and synthetic $nifH_{v2}$ (*Right*). All constructs bore P_{tac}-*nifHDK* with the synthetic gene indicated. Cultures were induced with 50 μM IPTG.
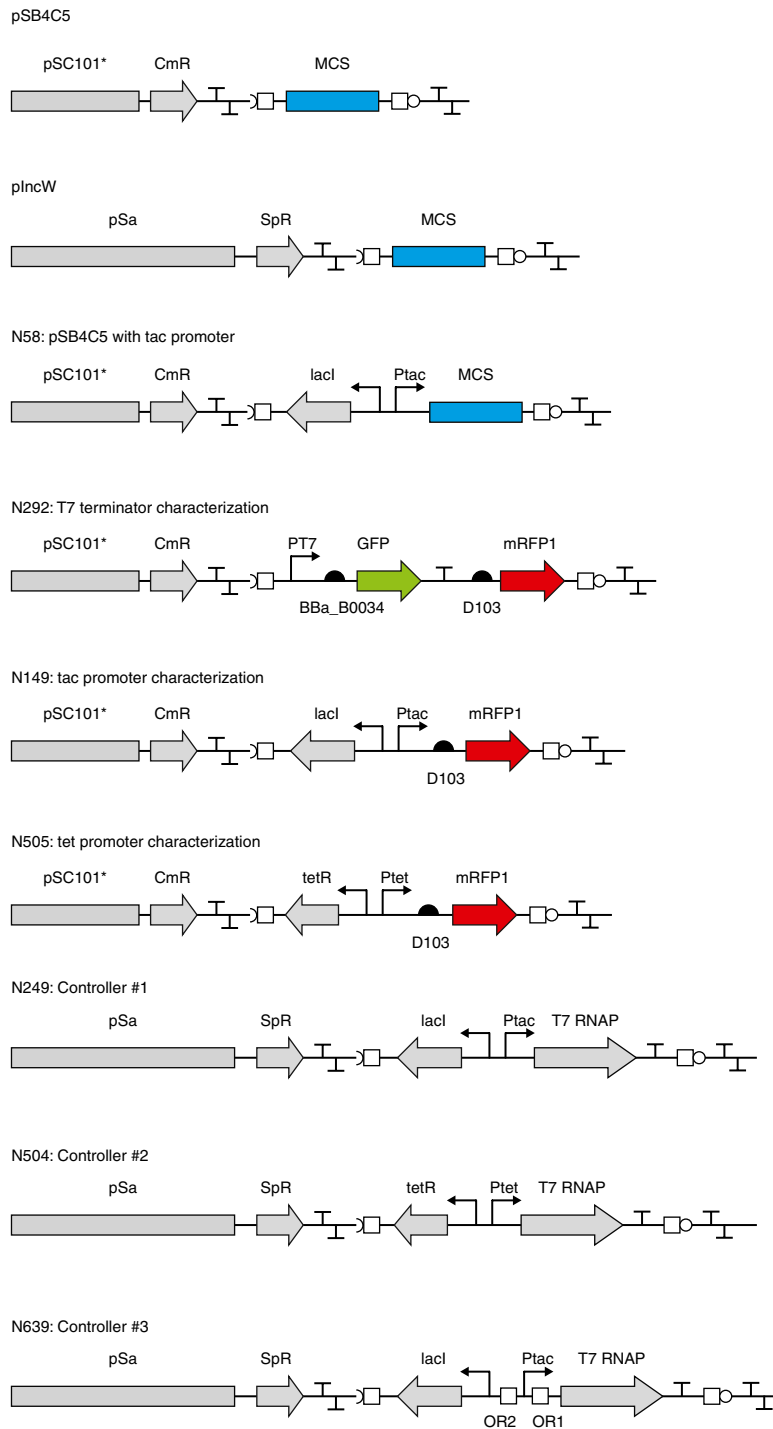
Fig. S6. Maps for key plasmids. Synthetic Biology Open Language graphical notation is used to describe genetic parts: the BioBrick prefix and suffix are open squares, and terminators are in the shape of a *T*.

Other Supporting Information Files