Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1120788109">http://dx.doi.org/10.1073/pnas.1120788109</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Jun 23 16:26:08 EDT 2017</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/89066">http://hdl.handle.net/1721.1/89066</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*

Karsten Temme,1, Dehua Zhao,1, and Christopher A. Voigt2

1Joint Graduate Group in Bioengineering, University of California, Berkeley/University of California, San Francisco, CA 94158; and 2Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited* by Christopher T. Walsh, Harvard Medical School, Boston, MA, and approved March 23, 2012 (received for review December 16, 2011)

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 promoters (15). The entire echinomycin biocatalyst has been shown that a promoter can be inserted upstream to induce expression (15). The entire echinomycin biocatalyst has been shown that a promoter can be inserted upstream to induce expression (15). 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 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).
Nitrogen fixation is the conversion of atmospheric N2 to ammonia (NH3) so that it can enter metabolism (22). Industrial nitrogen fixation, as observed previously (28), and it is frequently absent in the natural environment. NifT did not have an effect on nitrogenase activity of the inducible promoter. Our results indicate that the refactored cluster and controller are determined in Figs. 4 and S, respectively.

**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. 2B). 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 into operons and the selection of synthetic parts to obtain desired expression level is measured as the geometric average from a distribution of cells measured by flow cytometry (SI Materials and Methods).

**Component II**

- **ΔHDKTY**
  - **ΔJ**
  - **ΔF**

**Component III**

- **ΔUSVWZM**
- **ΔBQ**
- **ΔENX**

**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 nif/USVWZM operon, which encodes proteins for early Fe-S cluster formation and proteins for component maturation, needs to be expressed at low levels, whereas nif/BQ, encoding proteins for FeMo-co core synthesis and molybdenum integration, need to be expressed at high levels.**
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 nifβ genes, whereby the lower expression required for nifβ 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 nifBOQ 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 nifHDKYEN 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 N2 as a nitrogen source, growing 3.5-fold slower than the WT strain ([SI Materials and Methods](#)) and incorporating 15N-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 8% ± 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
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% ± 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 a refactored system can be cleanly fed back into the design cycle. The information gleaned from screening the permutations in a refactored system can be cleanly fed back into the design cycle. The information gleaned from screening the permutations in a refactored system can be cleanly fed back into the design cycle.
Refactoring 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 MSA1 (gift from Paul Ludden, University of California, Berkeley, CA) and mutants derived from MSA1 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 Na2HPO4, 3g KH2PO4, 0.25g MgSO4·7H2O, 1g NaCl, 0.1g CaCl2·2H2O, 2.9 mg FeCl3, 0.25 mg Na2MoO4·2H2O, and 20 g sucrose. Growth medium is defined as minimal medium supplemented with 6 mL (per liter) of 22% (wt/vol) NH4Ac. 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; 165 RNA-ACTCCTTA) 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 PromSite can 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 OD600 of 0.5 in 14-mL bottles, and bottles were sealed with rubber stoppers (ZSK64702; Sigma). Headspace in the bottles was repeatedly evacuated and flushed with N2 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 CaC2 by using a Burrell 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 the bottles.
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 ChemStation software and dividing ethylene production of experimental strains by the ethylene production of a WT control included in each assay.

**N2-Dependent Growth and 15N2 Incorporation Assay.** Nitrogen fixation by synthetic nif cluster in K. oxytoca is further demonstrated by N2-dependent growth and 15N2 incorporation. Cells are diluted as described in the acetylene reduction assay. The headspace of the bottles is replaced by normal N2 gas or by stable isotope nitrogen, 15N2 (15N 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, N2-dependent growth of the cells is determined by measuring optical density at 600 nm (i.e., OD600). To do the 15N2 incorporation assay, the 15N-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 15N/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.


Temme et al.
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 $1.66 \times 10^{-7}$, and the division by this number back-converts to the raw arbitrary units. This number was calculated to be a proxy to the expression of the corresponding 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 REU 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 $\mu$L 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 $P_{\text{tac}}$ promoter in plasmid N149 (SBA_000516), cotransferring with the T7 RNAP and measuring fluorescence via flow cytometry under 1 mM isopropyl-$\beta$-D-thiogalactopyranoside (IPTG) induction.

Replacement of $P_{\text{tac}}$ promoters by *T7* promoters. To replace the $P_{\text{tac}}$ 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 $P_{\text{tac}}$ 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 $P_{\text{tac}}$ promoter. However, we found that a WT T7 promoter produced inducible functional activity with a maximum at 1 mM IPTG induction of the T7 RNAP. For the synthetic *nifF* 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 $P_{\text{tac}}$ 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 *nifU*/*WZM* 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 from the natural 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. Depressia media is not strictly nitrogen-free, containing 1.45 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 Δnif 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$_{1,1}$) 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-y), whereas synthetic nifH$_{1,1}$ was not expressed regardless of the context of nifD (constructs N1 and N19). A second synthetic nifH (synthetic nifH$_{2,2}$) was used to replace synthetic nifH$_{1,1}$. The Western blot in Fig. S5 (Right) showed that the synthetic nifH$_{2,2}$ (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 (Lanza). 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, RBBS, terminators, and spacers) were combined with the initial synthetic gene sequences proposed by DNA2.0 in A Plasmid Editor (http://biologylab.sjsu.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 nusA, nuc1, the Tn21 integrase gene, and ORF18 (12). Plasmid pSB4C5 (pSCL101, 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 (pSCL101, CmR) was generated by inserting the P$_{lac}$ 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 BioBrick 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$_{lac}$ 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$_{lac}$ 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 multi-copy 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$_{lac}$ promoter and BioBrick suffix of plasmid N58 (pSCL101, CmR). One exception was plasmid Nif18, which was created 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$_{lac}$ cassette (SBa_000561) and the BioBrick suffix of plasmid N58 (pSCL101, 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.** 

*77 T* RNA polymerase. The T7 RNA polymerase was modified to be nontoxic to *Klebsiella* and *Escherichia coli* at high expression levels. The RNASP was expressed from a low-copy origin (pSa) under control of a weak RBS (Sba_000507, TATCCAAACCAGTAGCTCAATGGAGTCTCCTAT) and N-terminal degradation tag (Sba_000509, TGTGGTTATCAAGCTTGCGGATCTCCGGAAATTTGTAATTTCCGCATTATTAGCGATTCGGTTGCCTTCTTCACCGGCAGCGATTTAGTGAAACACCGCATCGATCTGGTGCC). The start codon was changed from ATG to GTG, and the active site contained a mutation (R632S).

**77 T** promoters. T7 promoters were generated from a random library. The T7 promoter seed sequence was TAAACGACTCACTANNNNNAGA. **Table S3** provides sequences of individual promoters.

**77 T** 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 TANNNACCCSSWSSSSTCTWCWGSSSSSSWSGGTTT. Terminator plasmids were co-transformed 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 RBSs was created 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 AND B and used this circuit to control T7* RNAP in controller 3. In this circuit, the A AND B logic corresponds to the presence or absence of the inducers, IPTG and aTc, such that the cell computes IPTG AND 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_tac 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 AND aTc.

**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). **Table S2** lists 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**.

---

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. 3B) 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

- [Table S1 (DOCX)](https://www.pnas.org/cgi/content/short/1120788109/6?6)
- [Table S2 (DOCX)](https://www.pnas.org/cgi/content/short/1120788109/6?6)
- [Table S3 (DOCX)](https://www.pnas.org/cgi/content/short/1120788109/6?6)