PNAS
www.pnas.org

Supplementary Information for

A polyprotein strategy for stoichiometric assembly of nitrogen fixation components for synthetic biology

Jianguo Yang, Xiaqing Xie, Nan Xiang, Zhe-Xian Tian, Ray Dixon, and Yi-Ping Wang

Ray Dixon & Yi-Ping Wang
Email: ray.dixon@jic.ac.uk (R.D.) & wangyp@pku.edu.cn (Y.-P.W.)

This PDF file includes:

SI Methods
Figs. S1 to S9
Tables S1 to S3
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S2

www.pnas.org/cgi/doi/10.1073/pnas.1804992115

SI Methods

Protein extraction and tryptic digestion. Bacterial cells were collected immediately after the acetylene reduction assay. The samples were resuspended in approximately 4 times volume of Lysis buffer (4% SDS, 100 mM Hepes pH=7.6 containing protease inhibitor cocktail and PMSF) and then boiled for 20 min. After centrifugation at 25 000 g for 30 min at 4℃, the supernatants were collected and stored at -80℃. The total protein concentration was measured using the BCA Kit. Equal amounts of samples were reduced, alkylated and digested with proteomics grade trypsin (Sigma, #T6567) at a trypsin-to-protein ratio of 1:50 (w/w) at 37℃ overnight to achieve complete digestion (1). Resulting peptides were vacuum dried and stored at -20℃.

Protein identification by high-resolution Orbitrap LC MS/MS analysis. Peptides were reconstituted in 0.1% formic acid (FA) and subjected to LC- MS/MS analysis. Online LC-MS/MS analysis was carried out on Q Exactive mass spectrometer (Thermo) coupled with a nano ACQUITY ultra performance liquid chromatography system (Waters). Peptides were separated on a BEH130 C18 analytical column (1.7 μm particles, 100 μm id × 150 mm length) at 300 nL/min, and subsequently eluted as follows (solvent A, 0.1% formic acid; solvent B, acetonitrile/ 0.1% formic acid): 0~2 min, isocratic with 5% B; 2~95 min, linear gradient to 30% B; 95~98 min, linear gradient to 80% B; 98~108 min, isocratic with 80% B; 108~110 min, isocratic to 5% B and for 120 min. Data-dependent MS/MS acquisition was performed following a full MS survey scan by Orbitrap at a resolution of 60,000 over the m/z range of 350-1800, and MS/MS measurements of the top 20 most intense precursor ions. The target values of automatic gain controls (AGC) were set to 200,000 for Orbitrap MS and 10,000 for ion-trap MS/MS detection. Dynamic exclusion was enabled for 60 seconds.

The resulting raw data from the LC-MS/MS analyses were separately converted to MGF files using the P Extract software (Thermo Fisher Scientific), and then searched against the in-house database using MASCOT (Matrix Science). The search parameter for tryptic digestion was restricted to a maximum of one missed cleavage of the protein. Mass tolerances were set up to 10 ppm for MS ions, and 0.05 Da for MS/MS fragment ions. Cysteine carbamidomethylation was set as a fixed modification. Oxidation of Met and acetylation of protein N-termini were considered as variable modifications.

Development of SRM-based quantification method. One representative peptide was selected for each targeted protein from the identified peptides list in Data-dependent MS/MS result. The selected peptides needed to be unique to the targeted proteins and show a high mass spectrometry signal response to maximize the sensitivity of the assay. The synthesized peptides (stable isotope-labeled and un-labeled) were ordered from BANKPEPTIDE LTD (Hefei, China).

Synthesized peptides were subjected to LC-SRM, performed on a hybrid triple quadrupole/ion trap mass spectrometer (6500 QTRAP, AB Sciex) interfaced with a nanoLC 1D plus high performance liquid chromatography system (Eksigent Technologies) at Keecloud Biotech (http://www.keecloud.com). The peptides were loaded onto a reverse-phase C18 trap column (Chrom XP, 350 μm id×0.5 mm length, 3 μm, 120A), and then separated through a C18 analytical column (75 μm id×150 mm, 3 μm, 120A). The LC mobile phase consisted of 0.1% FA in water (solvent A) and 0.1% FA in acetonitrile (solvent B). Peptides were eluted at the flow rate of 300 nL/min with a linear LC gradient from 5% to 40% solvent B over 45 min, followed by an increase to 80% solvent B over the next 5 min, before switching to 5% solvent B for a duration of 60 min. The SRM scans of the targeted peptides were selectively monitored with optimized transitions predicted by the Skyline software (1). SRM data were acquired in the positive ion mode with an ion spray voltage of 2,500 V and the nitrogen curtain gas of 25 psi. The pause between mass ranges was 5 ms, and the dwell time was set to 10 ms. The ion resolution for both the first quadrupole and the third

quadrupole was set to “unit” [0.7 Da, Full Width at half maximum (FWHM)]. Results were analyzed using Skyline. Three SRM transitions with optimized collision energy were selected for each peptide.

SRM-based quantification of targeted proteins by mass spectrometry. The absolute quantities of targeted proteins in samples were determined by liquid chromatography-isotope dilution mass spectrometry (2). Stable isotope-labeled ($^{13}C6^{15}N2$ lysine and $^{13}C6^{15}N4$ arginine) peptides were mixed with different amounts of standard peptides. The reference peptide mixture was then subjected to LC-SRM analysis. A linear response curve was plotted for each targeted peptide. Each sample was added to an equal amount of stable isotope-labeled peptides. The LC-SRM of samples was performed on a 6500 QTRAP (AB Sciex) with the established SRM method. Results were analyzed using Skyline (3). The concentration of target protein in each sample was calculated using the linear response curve.

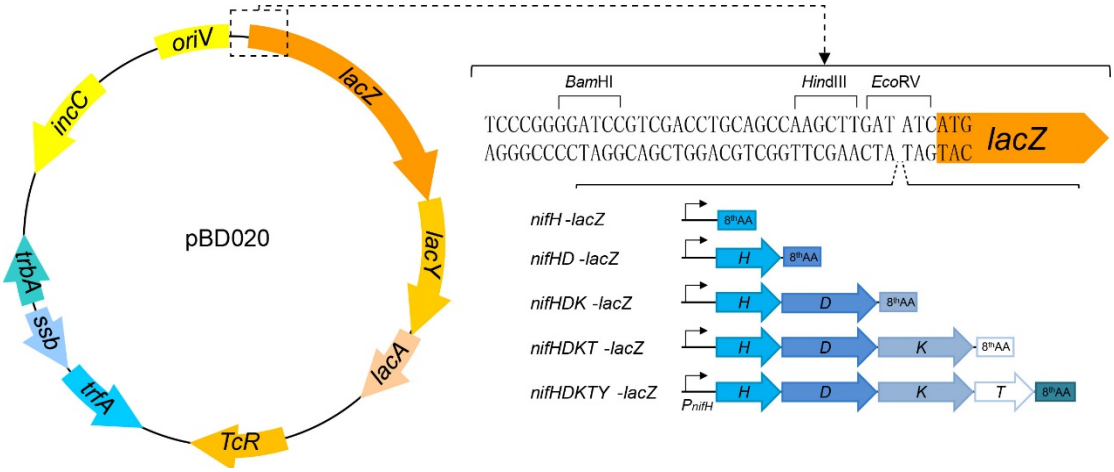

Fig. S1. Plasmid map of the vector pBD020 used for determining the expression level of each *nif* gene. The promoter-less *lacZYA* operon was used as a reporter, using β-galactosidase activity as a measure of gene expression. An *Eco*RV restriction site was introduced into the vector to enable blunt-end ligation of PCR products into *Eco*RV digested vector to generate in frame fusions with the *lacZYA* operon.

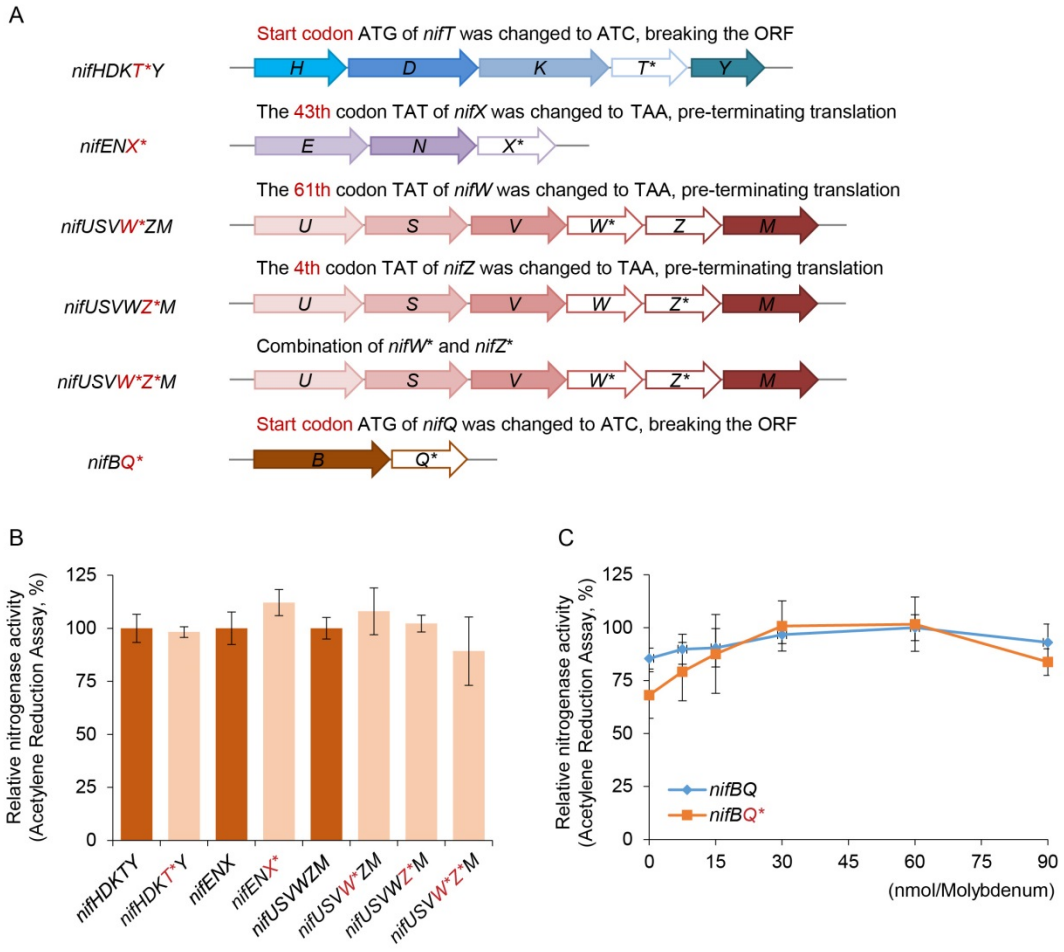

Fig. S2. Analysis of the requirements for accessory *nif* genes in the reconstituted operon-based Biobrick system in *E. coli* (this 7 operon *nif* system is shown in Fig. 1B). (A) Schematic diagram showing the construction of *nifT*, *nifX*, *nifW*, *nifZ* and *nifQ* deficient mutants. Primers used are provided in the *Dataset S1*. Mutations were introduced into Biobrick operon plasmids and tested for complementation of *nif* operon deletions. (See *Dataset S1* for details of the plasmids used in this figure). (B) Influence of mutations in *nifT*, *nifX*, *nifW*, *nifZ* and (C) Complementation of *nifQ* function with different concentrations of molybdenum as determined by the acetylene reduction assay. Nitrogenase activities obtained from complementation with the original biobrick *nif* operon plasmids were normalized to 100% in each case (specific activity nmol C_2H_4/min/mg total protein; *nifHDKTY*, 33.1±2.2; *nifENX*, 29.2±2.2; *nifUSVWZM*, 31.7±0.8). The *nifQ* deficient mutant (yellow squares) could be rescued by adding > 30 nmol molybdenum to the medium. Results with the native *nifBQ* operon are indicated by the blue diamonds. Error bars indicate the SD observed from at least two biological replicates.

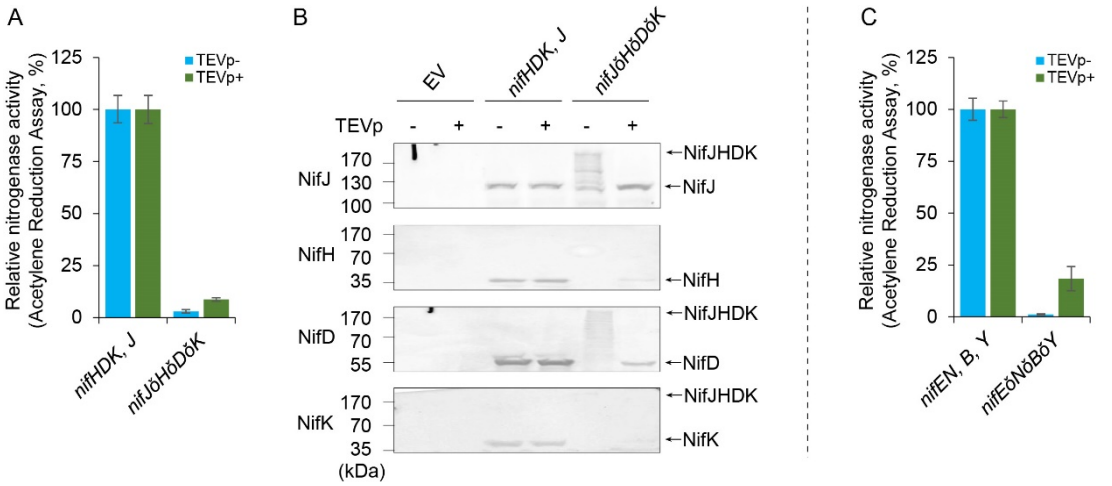

Fig. S3. Complementation of deletions in structural and biosynthetic *nif* genes by reassembled giant genes in *E. coli*. (A) Acetylene reduction exhibited by the giant gene *nifJöHöDöK* in the *ΔnifHDKTY* & *ΔnifJ* background. The level of activity shown by the complementation plasmid pBD367 carrying the native *nifHDKY* & *nifJ* genes (designated here as *nifHDK, J*) is represented as 100% (specific activity 28.5±1.8 (TEVp-), 27.3±1.9 (TEVp+) nmol C_2H_4/min/mg total protein respectively). (B) Western blotting of samples taken from the acetylene reduction assays shown in (A). Antibodies used are listed in *Materials and Methods* and full gels of the Western blots are available in Fig. S9. EV represents samples from the strain carrying the empty vector (pBDS1024K). Note that lower levels of NifH, NifD and NifK are expressed from the *nifJöHöDöK* gene. (C) Acetylene reduction exhibited by the giant gene *nifEöNöBöY* in the *ΔnifENX* & *ΔnifBQ* & *ΔnifTY* background. The level of activity shown by the plasmid pBD321 carrying native *nifENX*, *nifBQ* genes and [P_{Tc}-*nifY*] cassette (designated here as *nifEN, B, Y*) is represented as 100% (29.9±1.2 (TEVp-), 29.6±1.6 (TEVp+) nmol C_2H_4/min/mg total protein respectively). In all cases TEVp- and TEVp+ indicate the absence or presence of TEV proteinase encoding sequences respectively. Error bars indicate the SD observed from at least two biological replicates. The plasmids used in this figure are listed in *Dataset S1*.

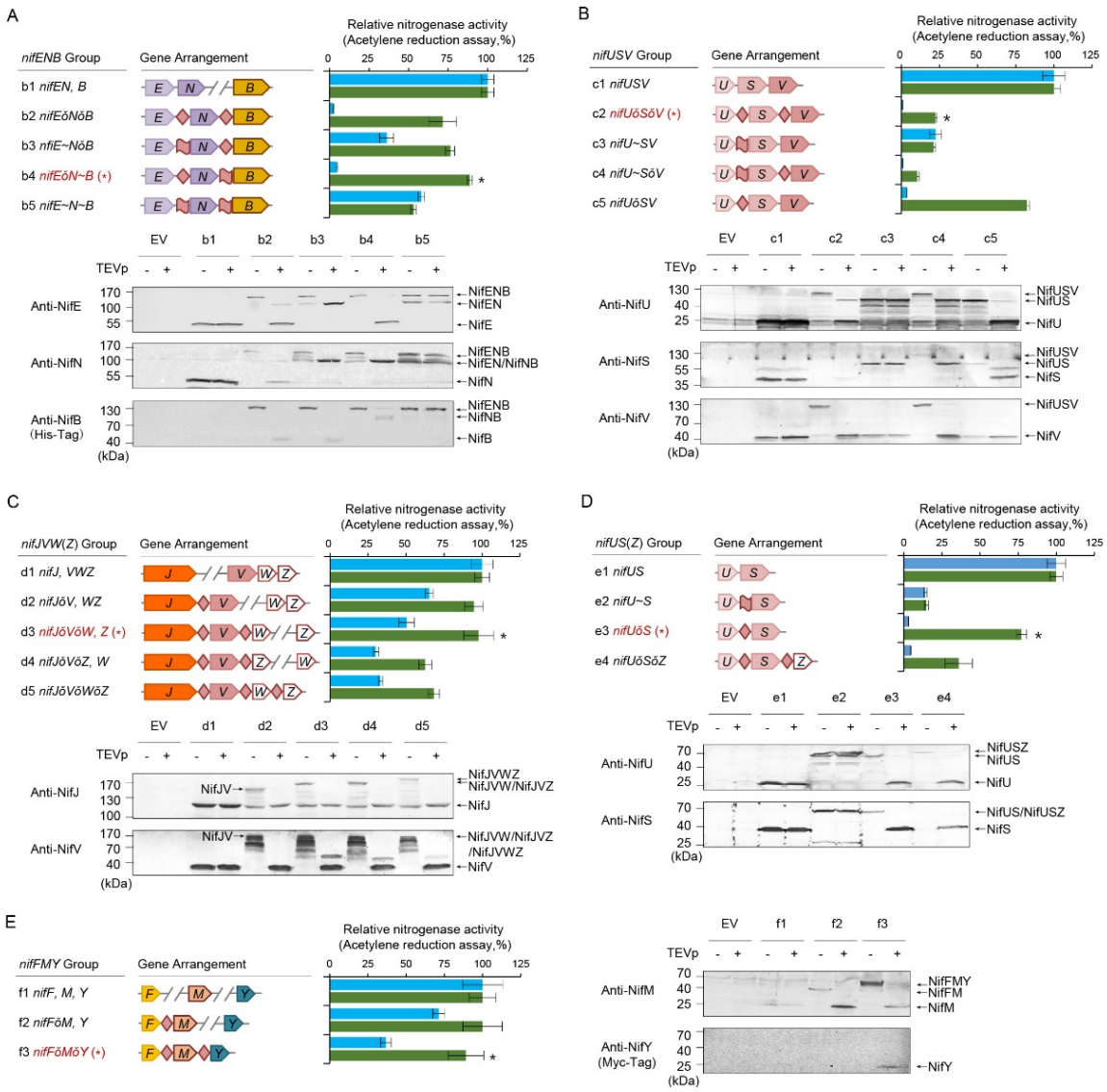

Fig. S4. Assessment of giant genes for complementation of nitrogenase activity and cleavage of their encoded polyproteins (A), *nifENB* group; (B), *nifUSV* group; (C), *nifJVWZ* group; (D), *nifUSZ* group and *nifFMY* group, (E). The description of this figure and experimental details are the same as in Fig. 2A. Acetylene reduction activities by the reconstituted operon-based system in *E. coli* were assigned as 100% (specific activity: nmol C_2H_4/min/mg total protein; *nifENX*, 32.6±1.3 (TEVp-), 30.1±1.3 (TEVp+); *nifUSV*, 28.3±1.1 (TEVp-), 26.9±1.9 (TEVp+); *nifJVWZ*, 31.2±1.5 (TEVp-), 30.0±2.1 (TEVp+); *nifUSZ*, 25.9±1.0 (TEVp-), 24.9±1.5 (TEVp+); *nifFMY*, 26.6±2.4 (TEVp-), 26.1±3.4 (TEVp+),). Error bars indicate the standard deviation (SD) observed from at least two biological replicates. Samples were immediately collected after the acetylene reduction assay for western blotting. For NifB and NifY, no specific antibodies were available, so a 10×His-tag coding sequence were added to the *nifB* gene of each construct from the *nifENB* group for detection of NifB using an anti-His antibody, and a Myc-tag was added to the *nifY* gene of each construct from the *nifFMY* group to detect NifY protein with anti-Myc antibody (Antibodies used are listed in *Materials and Methods*. Full gels of the Western blots are available in Fig. S9. The plasmids used in figure are listed in *Dataset S1*.). EV represents empty vector, used as a negative control.

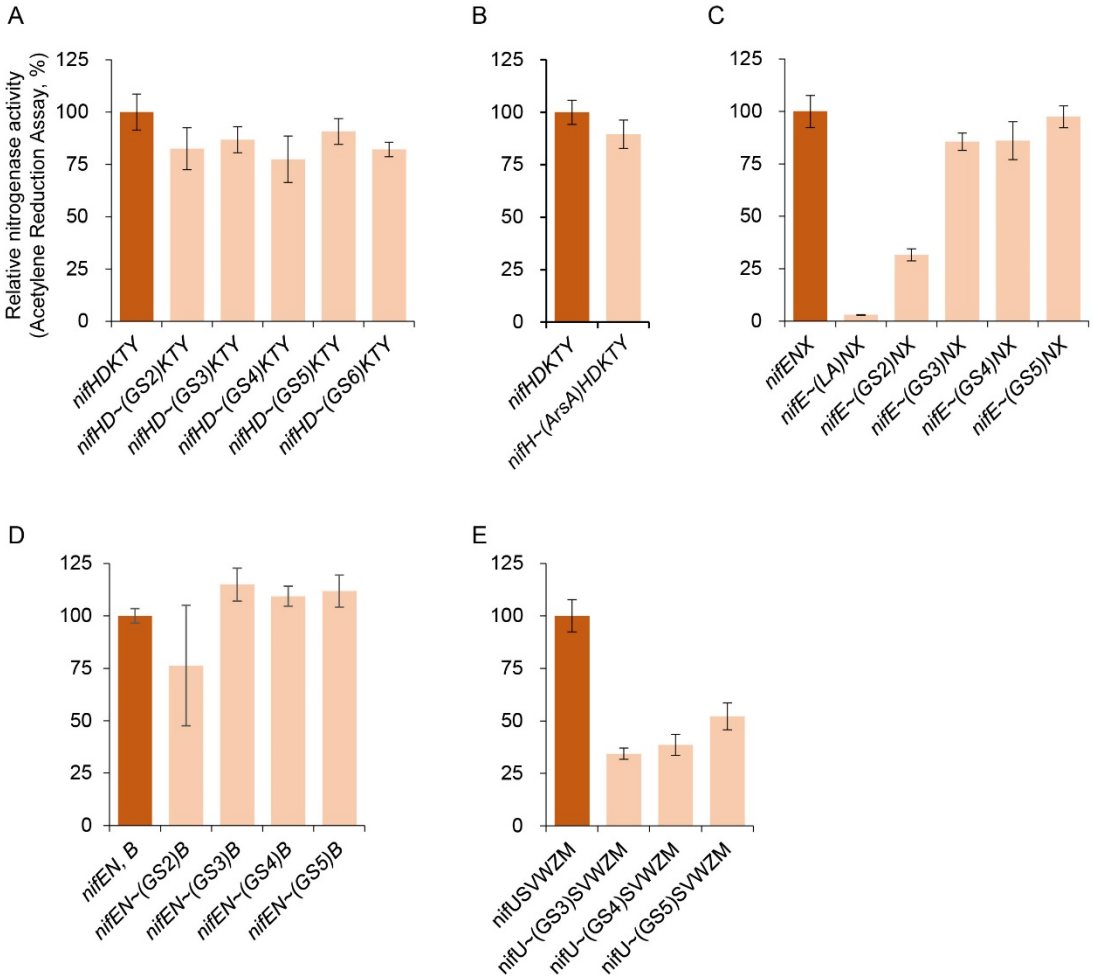

Fig. S5. Fusibility analysis of Nif proteins, determined by introducing *nif* gene fusions into the Biobrick operon-based system in *E. coli*. (A) NifDK, (B) fusion of two copies of NifH, (C) NifEN, (D) NifNB, and (E) NifUS. Linkers as indicated as follows: GS2 to GS6 represents the numbers of GGGGS linkers from 2 to 6 respectively; ArsA represents the specific linker from the ArsA protein of *E. coli*; LA in (C) represents the LA-linker from the naturally fused *nifEN* gene from *Anabaena variabilis* (4). Fusibility was determined by measuring acetylene reduction activities of the corresponding gene fusion derivatives of pKU7017 in *E. coli* strain JM109. (See *Dataset S1* for details of the plasmids used in this figure). In each case, activities exhibited by the original Biobrick construct pKU7017 are represented as 100% (29.6±2.6 nmol/min/mg total protein). Error bars indicate the SD observed from at least two biological replicates.

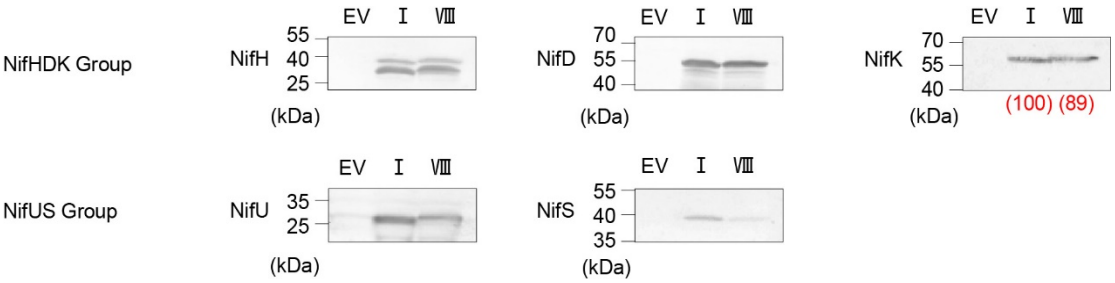

Fig. S6. Western Blot analysis of the protein levels of the polyprotein-based nitrogenase system (NifHDK and NifUS Groups). (Antibodies used are listed in *Materials and Methods*). EV represents empty vector, used as a negative control; I represents the native operon-based *nif* system (Construct Ⅰ in Fig. 3A); VIII represents the polyprotein-based *nif* system (Construct VIII in Fig. 3A). Image J software was used for quantification of NifK protein and relative expression levels are shown in red font as a percentage (in parentheses).

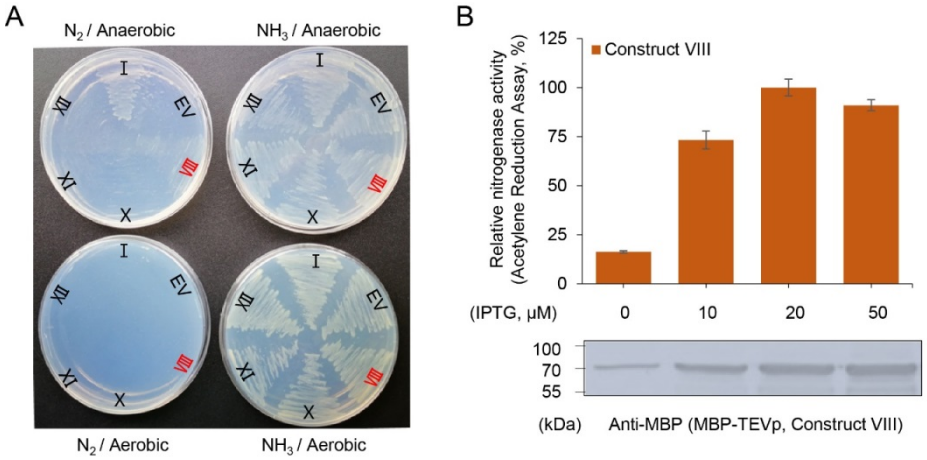

Fig. S7. (A) Diazotrophic growth promoted by polyprotein-based nitrogenase systems in the absence of IPTG. Roman numerals represent the corresponding assemblies in Fig. 3A. EV represents empty vector, used as a negative control. (B) Nitrogenase activity assay of the polyprotein-based system under different concentration of IPTG. In each case, activity exhibited by the construct VIII in the presence of 20 μM IPTG is represented as 100%. (C) Western Blot assay of MBP-TEVp induced with different concentrations of IPTG. Some MBP-TEVp protein was observed in the absence of IPTG due to leaky expression from the P_{tac} promoter. Antibodies used are listed in *Materials and Methods* and full gels of the Western blots are available in Fig. S9.

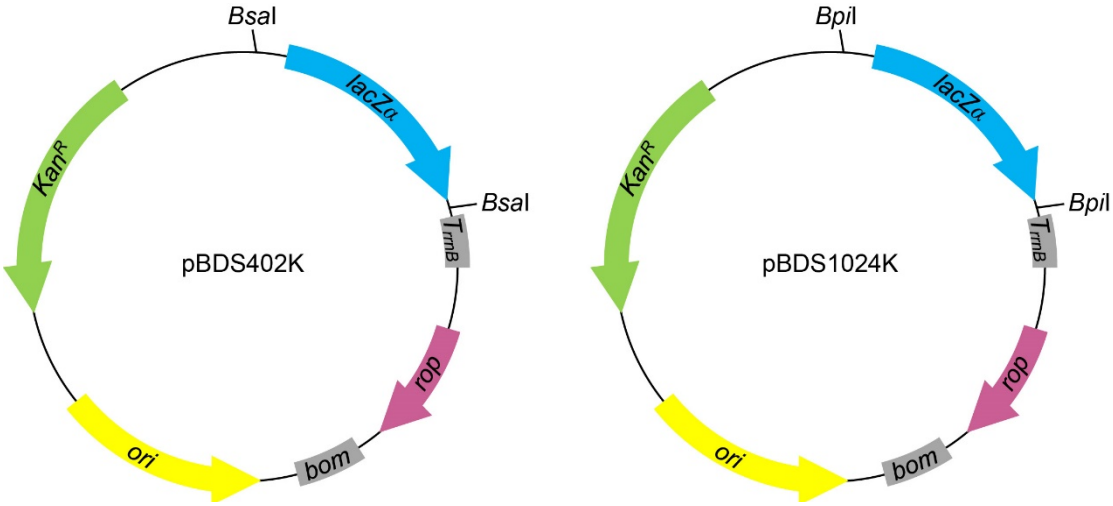

Fig. S8. Plasmid maps of the vectors pBDS402K and pBDS1024K used for assembling giant genes in this study. The *lacZα* coding sequence is flanked with either *Bsa*I or *Bpi*I restriction sites to facilitate Golden Gate assembly.

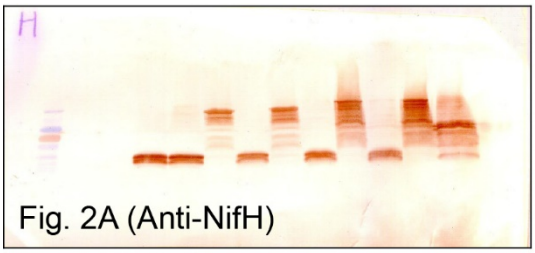

Fig. 2A (Anti-NifH)

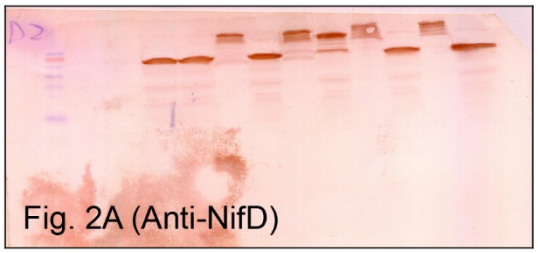

Fig. 2A (Anti-NifD)

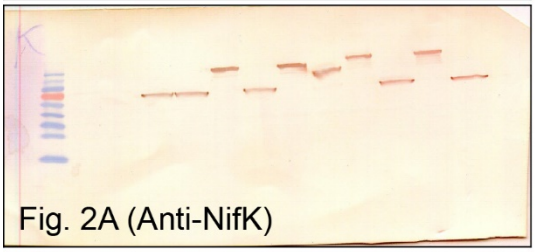

Fig. 2A (Anti-NifK)

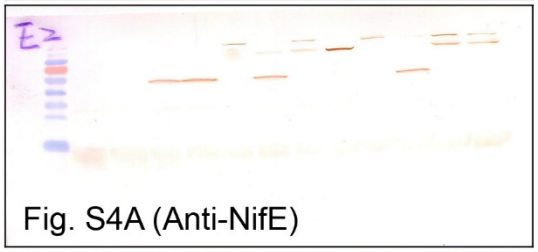

Fig. S4A (Anti-NifE)

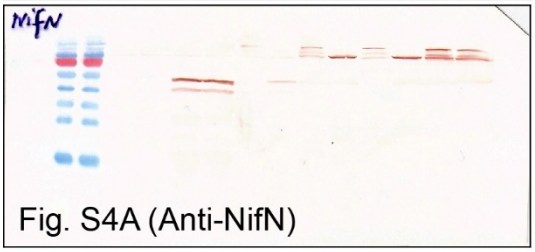

Fig. S4A (Anti-NifN)

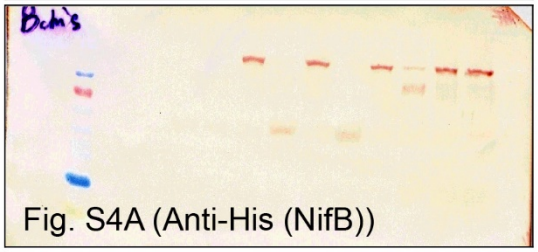

Fig. S4A (Anti-His (NifB))

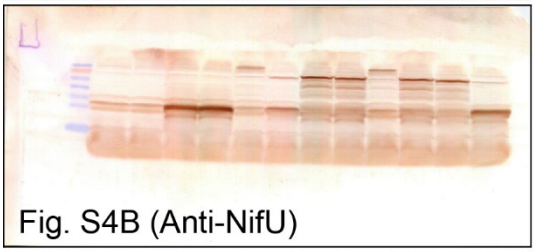

Fig. S4B (Anti-NifU)

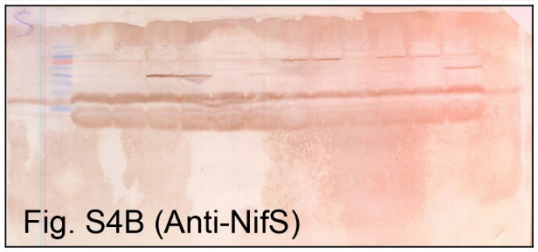

Fig. S4B (Anti-NifS)

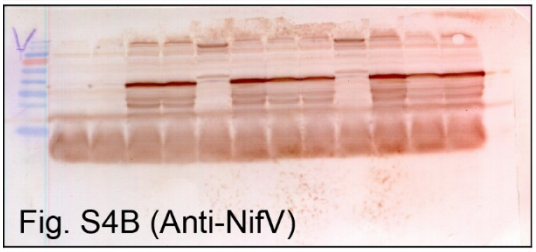

Fig. S4B (Anti-NifV)

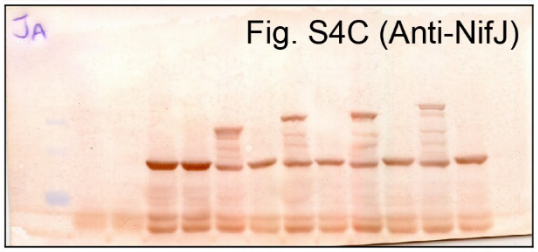

Fig. S4C (Anti-NifJ)

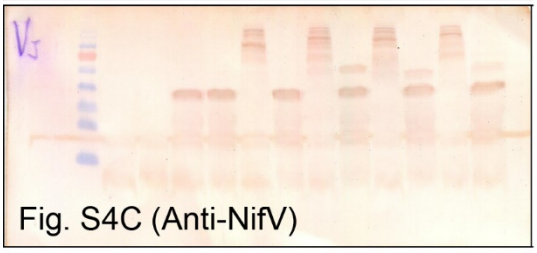

Fig. S4C (Anti-NifV)

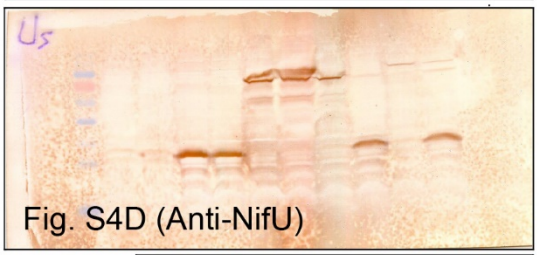

Fig. S4D (Anti-NifU)

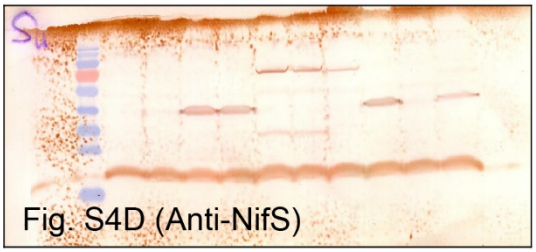

Fig. S4D (Anti-NifS)

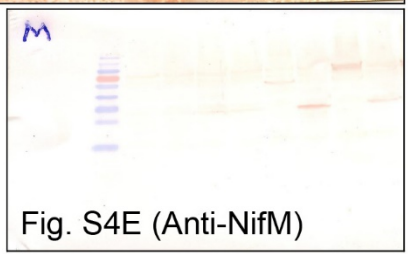

Fig. S4E (Anti-NifM)

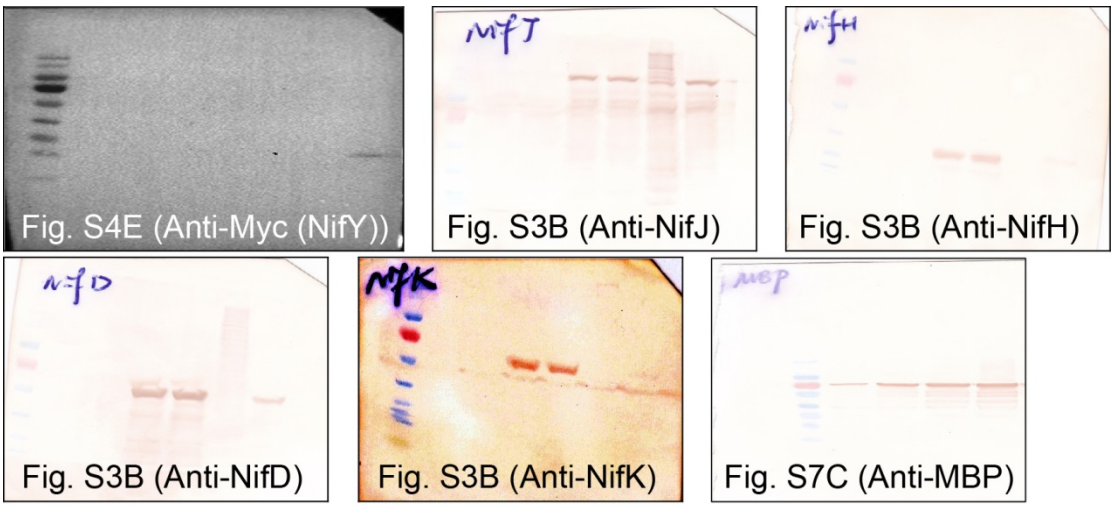

Fig. S9. Full gels of the Western blots shown in Fig. 2, Fig.S3, Fig.S4, and Fig.S7.

Table S1. Relative expression level of each *nif* gene.

***nif* gene**	**Relative expression level[a]**
H	100 ± 11
D	55 ± 10
K	45 ± 8
T	8 ± 0 [b]
Y	17 ± 2
E	23 ± 2
N	27 ± 5
X	19 ± 2
B	16 ± 4
Q	1 ± 0 [b]
U	8 ± 2
S	16 ± 4
V	9 ± 2
W	2 ± 1
Z	6 ± 2
M	2 ± 1
J	38 ± 7
F	5 ± 0[b]

[a.] The eighth codon of each *nif* gene was fused in frame to the *lacZYA* reporter (see Supplementary Fig 1) and the resultant plasmids were co-transformed with plasmid pKU7017 into *E. coli* strain JM109 to measure β-galactosidase activity under diazotrophic conditions. The expression level of the *nifH* gene is represented as 100%.
[b.] SD values < 0.5.

Table S2. Tailing-tolerance of each *nif* gene product

***nif* gene**	**Tailing-tolerance[a]**
H	97 ± 6
D	89 ± 9
K	1 ± 0[b]
T	ND[c]
Y	104 ± 7
E	85 ± 5
N	90 ± 9
X	106 ± 5
B	71 ± 8
Q	80 ± 3
U	85 ± 2
S	97 ± 6
V	117 ± 7
W	103 ± 6
Z	116 ± 9
M	126 ± 7
J	87 ± 5
F	90 ± 11

[a.] Each *nif* gene carrying the coding sequence of the extended ENLYFQ-tail was used to replace the corresponding native gene in the operon-based biobrick system. Values represent the acetylene reduction activity exhibited by each gene replacement, normalized against the activity exhibited by the native genes in the biobrick system in *E. coli* strain JM109 (100%).
[b.] SD value < 0.5.
[c.] Not Determined.

Table S3. Tailing-tolerance of *anf* structural gene products

***anf* gene**	**Tailing-tolerance[a]**
H	47 ± 7
D	101 ± 9
G[b]	10 ± 0[c]
K	4 ± 0

[a.] Each *anf* gene carrying the coding sequence of the extended ENLYFQ-tail was used to replace the corresponding native gene in the operon-based biobrick system. Values represent the acetylene reduction activity exhibited by each gene replacement, normalized against the activity exhibited by the native genes in the biobrick system in *E. coli* strain JM109 (100%).
[b.] Assayed by ^{15}N assimilation as an *anfG* deletion has a negligible effect on acetylene reduction activity of the iron-only nitrogenase in *E. coli* (5).
[c.] SD < 0.5.

References

1. Wiśniewski JR, Zougman A, Nagaraj N, & Mann M (2009) Universal sample preparation method for proteome analysis. *Nature Meth* 6:359.
2. Kirkpatrick DS, Gerber SA, & Gygi SP (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* 35(3):265-273.
3. MacLean B*, et al.* (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26(7):966-968.
4. Suh M-H, Pulakat L, & Gavini N (2002) Functional expression of the FeMo-cofactor-specific biosynthetic genes nifEN as a NifE-N fusion protein synthesizing unit in *Azotobacter vinelandii*. *Biochem Bioph Res Co* 299(2):233-240.
5. Yang J, Xie X, Wang X, Dixon R, & Wang Y-P (2014) Reconstruction and minimal gene requirements for the alternative iron-only nitrogenase in *Escherichia coli*. *Proc Natl Acad Sci USA* 111(35):E3718-E3725.

Other supplementary materials for this manuscript include the following (separate files):

Dataset S1. Plasmids and Primers used in this study.

Dataset S2. Complete DNA sequences of the Construct VIII defined in Fig. 3.