Supplementary Information for

Caffeine Junkie: an unprecedented GST-dependent oxygenase required for caffeine degradation by *P. putida* CBB5

Ryan M. Summers[1], Jennifer L. Seffernick[2], Erik M. Quandt[3], Chi Li Yu[1], Jeffrey E. Barrick[3], Mani V. Subramanian[1*]

[1]Center for Biocatalysis and Bioprocessing and Department of Chemical and Biochemical Engineering, The University of Iowa, Coralville, IA 52241, USA.

[2] Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, St. Paul, MN 55108, USA.

[3]Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, TX 78712, USA.

*Correspondence should be addressed to Mani Subramanian (mani-subramanian@uiowa.edu)

SUPPLEMENTARY METHODS

Determination of the full *ndmE* and *orf9* sequences. All PCR primers in this study (Table S3) were purchased from Integrated DNA Technologies (Coralville, IA). The theoretical peptide sequence of the known 5′ end of *ndmE*, previously reported as incomplete *orf8* (Summers *et al*., 2012), comprised 123 amino acids. The remaining 3′ end of the *ndmE* gene and sequence 3′- to *ndmE* were obtained by sequencing two DNA fragments generated by PCR amplification from an EcoRI genomic DNA library (Summers *et al*., 2012). A BlastP search of the 123 N-terminal amino acids of NdmE revealed two highly homologous glutathione-*S*-transferase proteins: Gst9 from *Janthinobacterium* sp. Marseille (GenBank accession no. YP_001355366, E-value of 2×10^{-58}) and G1E_06908 from *Pseudomonas* sp. TJI-51 (ZP_08139011, E-value of 2×10^{-53}). An alignment of the two sequences revealed matching peptide sequences near the C-terminus (PGFTVMSGIFP, residues 198 to 208 of both proteins) with nearly identical DNA sequences. A degenerate reverse primer, gst-degR2, was designed from the DNA sequence encoding peptides 202 through 208 of both proteins. Primer gst-degR2 was used with primer gst-F1 and a thermal profile of 10 s at 98°C, 30 s at 55°C, and 15 s at 72°C for 35 cycles to generate an approximately 400 bp DNA fragment. This PCR product was gel-purified and directly sequenced. The 5′ end of the PCR product completely matched the 3′ end of the known *ndmE* sequence, while the 3′ end brought the total known *ndmE* sequence to 620 bp. The stop codon was still missing from the incomplete *orf8*.

A nested PCR approach was used to amplify the missing 3′ end of *ndmE* and an additional 1.3 kb of DNA 3′ to *ndmE*. Two specific forward primers, gst-F3 and gst-F4, and two specific reverse primers, Marcy and Marcy2, were designed from the incomplete *ndmE*

ORF and the vector backbone of pUC19-Kan (Summers *et al.*, 2012). Using primers gst-F3 and Marcy2, a primary PCR reaction was run using the *EcoRI* genomic DNA library as a template with *Taq* DNA polymerase and a thermal profile of 30 s at 95°C, 30 s at 60°C, and 3 min at 72°C for 35 cycles. Then, 1 μL of the primary PCR reaction was used as a template in a second round of PCR with primers gst-F4 and Marcy and *Taq* DNA polymerase. A thermal profile of 30 s at 95°C, 30 s at 60°C, and 1.75 min at 72°C for 35 cycles was used in the second round of PCR, resulting in amplification of an approximately 1.6 kb PCR product. Control PCR reactions using either gst-F4 or Marcy primers alone did not yield this 1.6 kb PCR product. DNA sequencing of the gel-purified 1.6-kb PCR product revealed the missing 3′ end of *ndmE* and an additional ORF 3′ to *ndmE*, designated as *orf9* (Fig 1A).

Deletion of *ndmC*, *ndmD*, and *ndmE* genes. CBB5 strains with a single gene deletion were created by the method of Link *et al*. (7) to create a truncated gene comprising the first and last 21 bp (14 amino acids total) of the gene of interest (GoI). The sequence starting between 600-800 bp upstream of the GoI and including the first 21 bp of the GoI was PCR-amplified using primers NdmX-KO-NF and NdmX-KO-NR (reaction N), where NdmX is either NdmC, NdmD, or NdmE. Similarly, the sequence downstream of the gene of interest, starting with the last 21 bp of the GoI and ending between 600-800 bp downstream of the GoI was PCR-amplified using primers NdmX-KO-CF and NdmX-KO-CR (reaction C). A thermal profile of 10 s at 98°C, 30 s at 58°C, and 30 s at 72°C for 35 cycles was used for both reaction N and C. Upon completion, reactions N and C were mixed 1:1 (v/v) and used as a template for a secondary PCR reaction with a thermal profile of 10 s at 98°C, 30 s at 58°C, and 60 s at 72°C for 35 cycles. This secondary PCR reaction resulted in a DNA insert with desired restriction sites at each end and the truncated GoI in the middle. The

PCR-amplified insert was digested with EcoRI and either BamHI (for the *ndmC* and *ndmD* knock outs) or XbaI (for the *ndmE* knock out) and ligated to the pEX18Gm plasmid previously cut with the same restriction enzymes, resulting in plasmids pCKO, pDKO, and pEKO.

Electrocompetent CBB5 cells were prepared by first growing overnight at 30°C and 225 rpm in a modified M9 medium supplemented with 0.4% soytone and 0.25% caffeine (M9CS medium) (Yu *et al*., 2009). When the optical density at 600 nm (OD_{600}) reached 1.0-1.3, cells were incubated on ice for 10 min, followed by centrifugation at 4,000□*g* for 10 min at 4°C. Cells were then subsequently washed in 50, 25, and 12.5 mL 1 mM HEPES (pH = 7.5), and ultimately suspended in 200 µL 1 mM HEPES (PH = 7.5). Between 1-2 µg plasmid pCKO, pDKO, or pEKO were added to 40 µL freshly prepared electrocompetent CBB5 and transformed in a 2 mm cuvette using 2.5 kV, 25 µF, and 200 Ω. Transformed cells were rescued 1 h at 30°C in 1 mL SOC, plated on agar plates containing M9 salts, 4 $g{\cdot}L^{-1}$ soytone, 3 $g{\cdot}L^{-1}$ glucose, and 2.75 $g{\cdot}L^{-1}$ ammonium chloride (M9SGN medium) with 5 $\mu g{\cdot}mL^{-1}$ gentamycin (Gm), and incubated 2 days at 30°C. The pEX18Gm plasmid cannot replicate in CBB5. Therefore, any resultant transformants must have integrated the plasmid into the chromosome *via* crossover with the sequence flanking the GoI.

One of each transformant was grown overnight in 5 mL M9SGN containing 5 $\mu g{\cdot}mL^{-1}$ Gm at 30°C and 225 rpm. About 5 µL of overnight culture was then inoculated into 5 mL M9SGN without antibiotic and cultured at 30°C and 225 rpm for 20-24 h to allow for removal of the plasmid from the chromosome. These cell cultures without antibiotic were plated on LB agar containing 5% sucrose and incubated overnight at 30°C. Any resultant cells growing on the plates with 5% sucrose must have excised the plasmid, as presence of

the *sacB* gene would have toxic effects in the presence of sucrose. Because removal of the plasmid could result in either gene deletion or reversion to the native genotype, colonies were screened by PCR using primers flanking the GoI. Each gene deletion was confirmed by sequencing of PCR-amplified DNA.

Cloning of *ndm* genes. Forward primer ndmC-F-NdeI and reverse primer ndmC-R-XhoI were used for PCR amplification of *ndmC* from CBB5 genomic DNA using a thermal profile of 10 s at 98°C, 30 s at 58°C, and 30 s at 72°C for 30 cycles. The 0.8-kb PCR product was digested with NdeI and XhoI and ligated to the plasmid pET32a(+) previously digested with NdeI and XhoI, producing plasmid p32CHis. Primers ndmE-F-NdeI and ndmE-R-EagI were used to amplify *ndmE* from CBB5 genomic DNA with a thermal profile of 10 s at 98°C, 30 s at 58°C, and 30 s at 72°C for 35 cycles. The PCR product was digested with NdeI and EagI and ligated into the pET28a(+) plasmid previously cut with NdeI and EagI, producing plasmid p28HisE. The *ndmCDE* genes were PCR-amplified from CBB5 genomic DNA using primers ndmC-F-NdeI and ndmE-R-EagI with a thermal profile of 10 s at 98°C, 30 s at 58°C, and 105 s at 72°C for 35 cycles. The approximately 3.3-kb PCR product was digested with NdeI and EagI and ligated to the pET28a(+) plasmid previously digested with NdeI and EagI, resulting in plasmid p28HisCDE.

The *ndmA* and *ndmB* genes were cloned into the pACYCDuet-1 plasmid for co-expression with His_6-NdmCDE. *ndmA* was PCR amplified from CBB5 genomic DNA using primers NdmA-F-NcoI and NdmA-R-BamHI with a thermal profile of 10 s at 98°C, 30 s at 58°C, and 75 s at 72°C for 35 cycles. The approximately 1.1-kb PCR product was digested with NcoI and BamHI and ligated to pACYCDuet-1 previously cut with NcoI and BamHI, resulting in plasmid dA0. The NcoI and NdeI restriction sites were removed from *ndmB*

using the procedure described in the QuikChangeII site-directed mutagenesis kit (Stratagene, Santa Clara, CA). PCR primers Bsdm-NcoI-F and Bsdm-NcoI-R were used with plasmid pET32-ndmB-His (Summers *et al*., 2012) as template in the first round of site-directed mutagenesis to produce plasmid p32BHis-sdm1. This plasmid was then used as template for a second round of site-directed mutagenesis with primers Bsdm-NdeI-F and Bsdm-NdeI-R to produce plasmid p32BHis-sdm2. *ndmB* was PCR amplified from p32BHis-sdm2 using primers NdmB-F-NdeI and NdmB-R-KpnI. The resulting 1.1-kb PCR product was digested with NdeI and KpnI and ligated to plasmid dA0 previously cut with NdeI and KpnI, resulting in plasmid dAB.

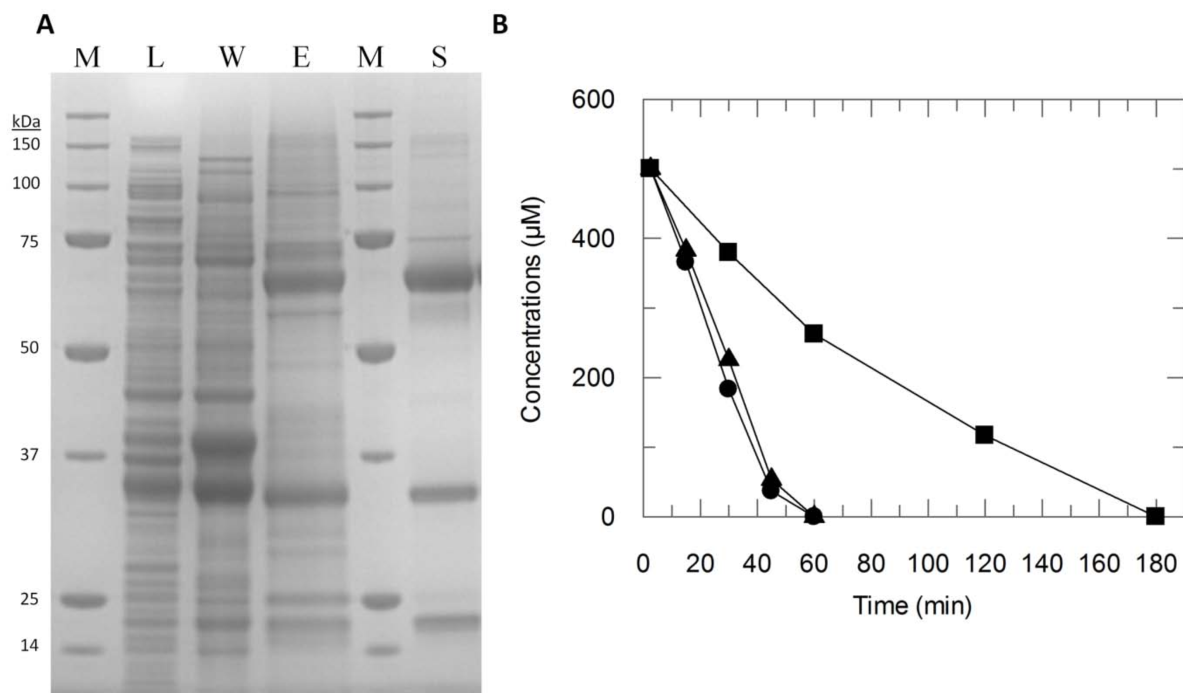

Fig. S1. NdmAB do not complex with His$_6$-NdmCDE. (A) SDS-PAGE of Ni-eluted fractions. Lane M, MW ladder with size of peptides shown on the left; L, sample loaded on the column; W, unbound protein washed from the column; E, bound protein eluted from the column; S, Sephacryl S200-purified His$_6$-NdmCDE loaded as a standard. Fraction E contained NdmC activity, but was not active toward caffeine or 7-methylxanthine. Fraction W alone was not active toward caffeine, theobromine, or 7-methylxanthine. The size of His$_6$-NdmC (lanes E and S) is slightly smaller than that of NdmB (lane W). (B) Cell extracts of *E. coli* expressing His$_6$-NdmCDEAB (lane L) consumed 500 µM caffeine (■) in 180 min and 500 µM theobromine (▲) or 7-methylxanthine (●) in less than 60 min. No *N*-demethylation activity was detected in unbound protein (lane W). Bound and eluted protein (lane E) converted 148.3 µM 7-methylxanthine to xanthine, but was not active toward caffeine or theobromine.

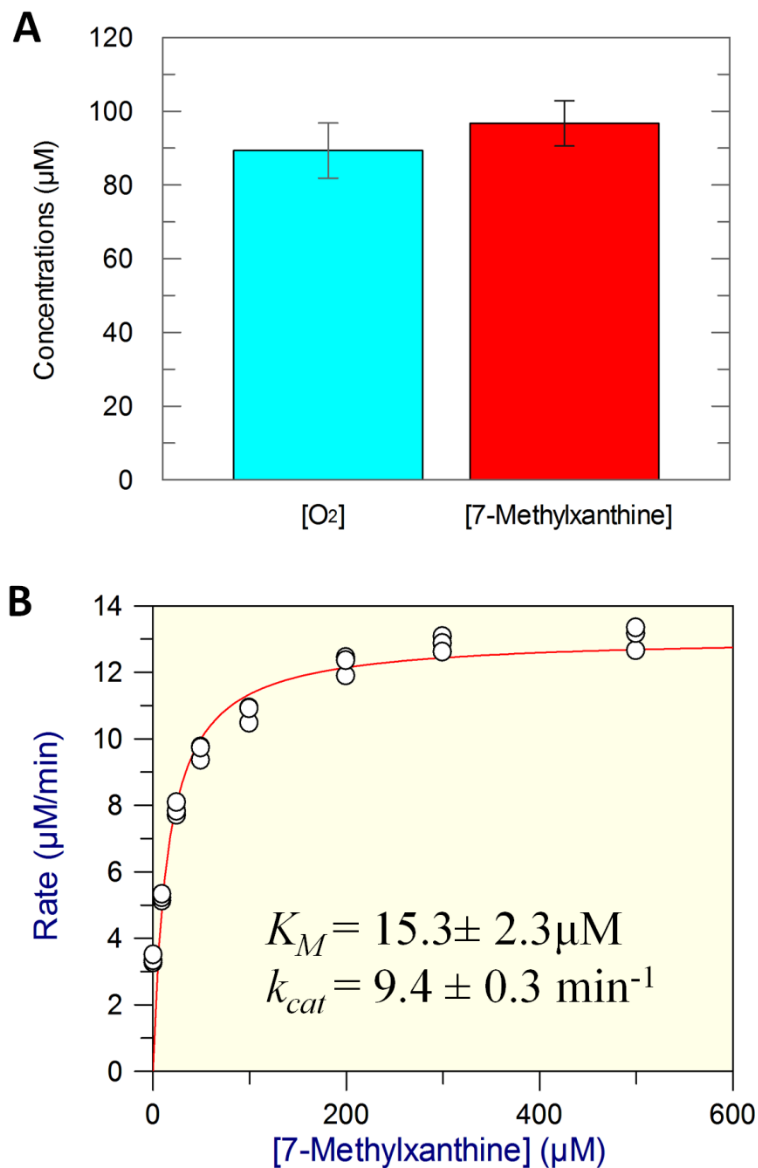

Fig. S2. (A) His_6-NdmCDE consumed 89.3 ± 7.5 µM oxygen for N_7-demethylation of 96.7 ± 6.2 µM 7-methylxanthine. (B) Michaelis Menten plot for kinetic determination of His_6-NdmCDE toward 7-methylxanthine. The apparent K_m and k_{cat} values of His_6-NdmCDE for 7-methylxanthine in 50 mM KP_i (pH 7.5) and 30°C were 15.3 ± 2.3 µM and 9.4 ± 0.3 min^{-1}.

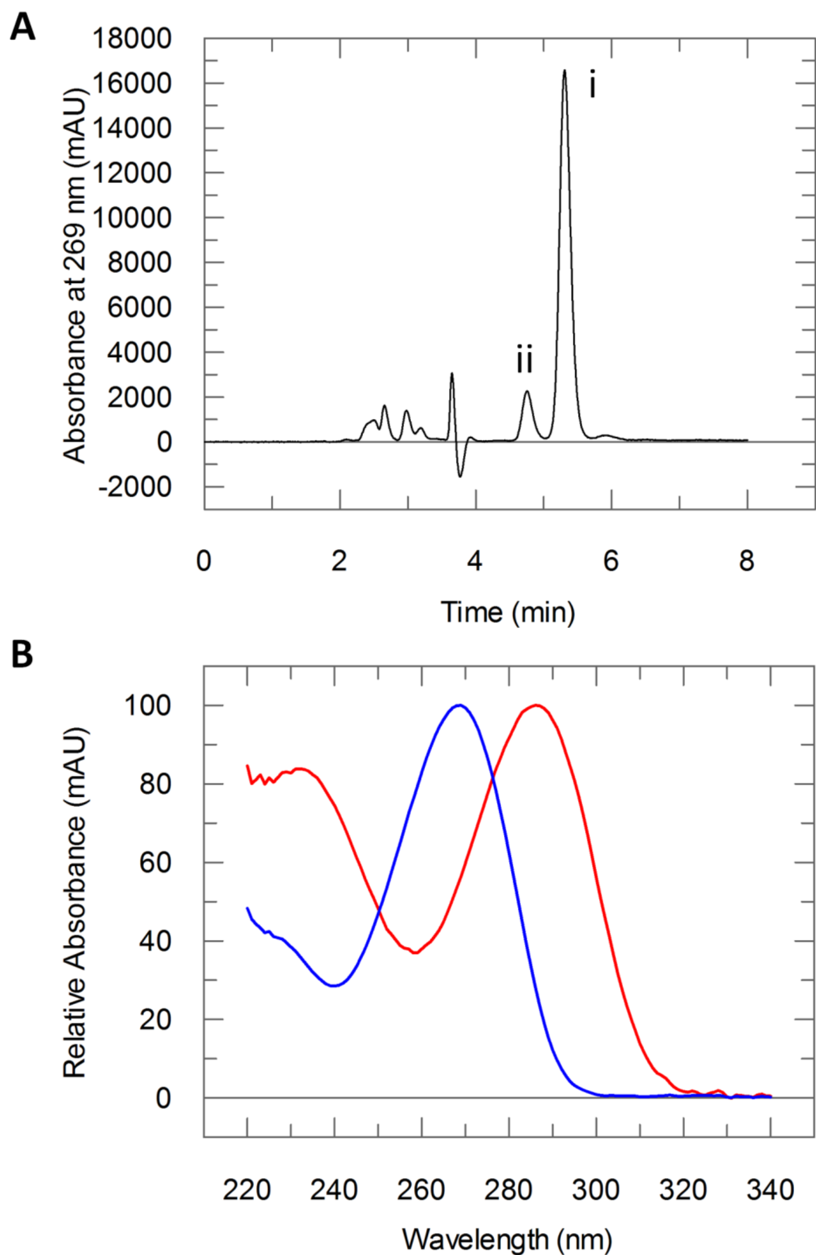

Fig. S3. Characterization of 7-methyluric acid produced from 7-methylxanthine by single gene knock out strains of *P. putida* CBB5. (A) HPLC chromatogram of the CBB5Δ*ndmD* reaction at 24 h analyzed at 269 nm. Peak i, 7-methylxanthine; peak ii, 7-methyluric acid. (B) UV spectra of the 7-methylxanthine (blue) and 7-methyluric acid (red) peaks from the chromatogram in part A (peaks i and ii, respectively).

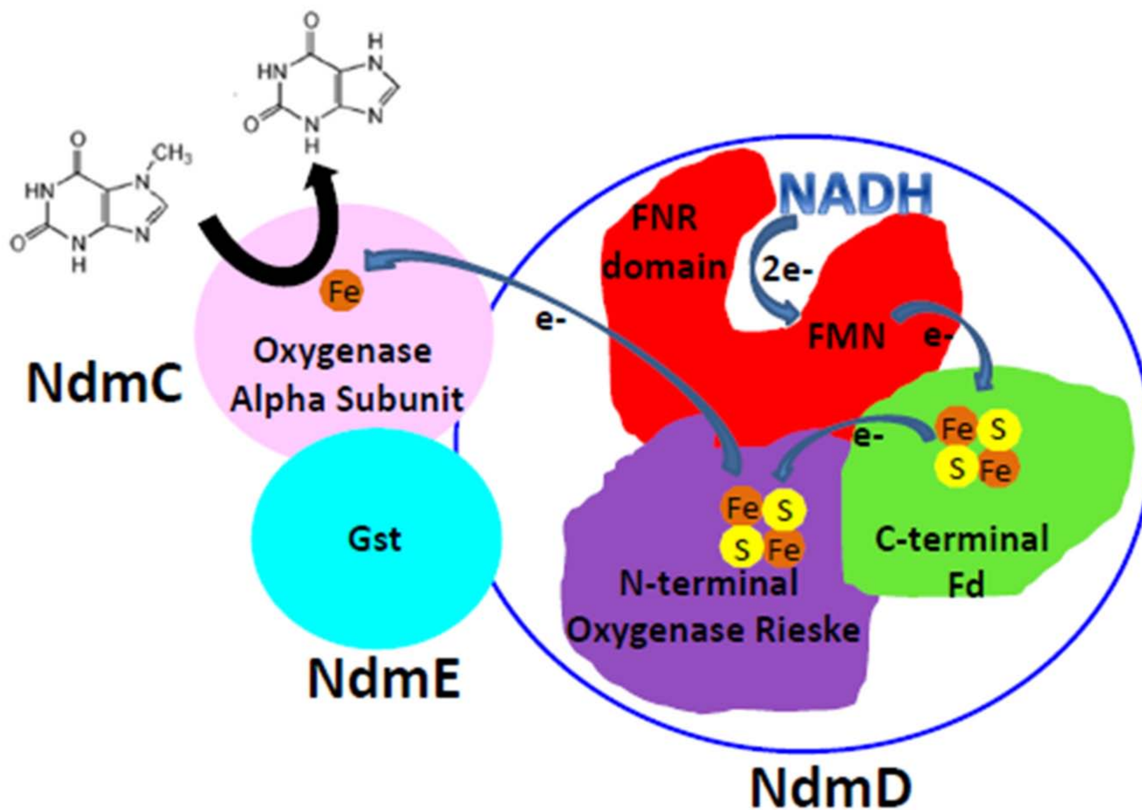

Fig. S4. A schematic representation of predicted electron flow through the 7-methylxanthine N_7-demethylation components in the caffeine degradation pathway of *P. putida* CBB5 based on that observed in other RO systems (*39*). The circles represent individual proteins. Two electrons are transferred from NADH to the FMN-containing FNR domain in NdmD (red). One electron at a time is transferred to the C-terminal ferredoxin domain of NdmD (green), which shuttles the electron to the NdmD N-terminal oxygenase Rieske domain (blue). The electrons finally transfer to the iron in the oxygenase alpha subunit (NdmC, pink), which can then N_7-demethylate 7-methylxanthine to produce xanthine. NdmE (cyan) is required for function and may be involved in complex assembly.

Table S1. Apparent solubility of NdmCDE expressed in different combinations in *E. coli* BL21(DE3).

Protein(s) Expressed	% Soluble[a] NdmC	% Soluble[a] NdmD	% Soluble[a] NdmE
NdmC	0	N. A.	N. A.
NdmD	N. A.	<10%	N. A.
NdmE	N. A.	N. A.	0
NdmCD	0	<10%	N. A.
NdmCE	0	N. A.	0
NdmDE	N. A.	<10%	0
NdmCDE	>50%	>50%	>50%

[a]% Soluble protein was determined by comparing the soluble and insoluble protein bands observed by SDS-PAGE.

Table S2: Protein accession and GI numbers for NdmCDE gene clusters.

Organism[a]	Plasmid encoded	Accession and GI Numbers			
		1st NdmC	2nd NdmC	NdmD	NdmE
Pseudomonas putida CBB5	No	JQ061129 379334193	-	AFD03119 379334196	KC778191
Janthinobacterium sp. Marseille	No	YP_001355368 152981996	-	YP_001355367 152981877	YP_001355366 152981788
Klebsiella pneumoniae subsp. pneumoniae WGLW2	No	ZP_18487397 425084301	-	ZP_18487398 425084302	ZP_18487399 425084303
Pseudomonas sp. TJI-51[b]	draft	ZP_08139009 325272654	-	ZP_08139010 325272655	ZP_08139011 325272656
Mesorhizobium alhagi CCNWXJ12-2	No	ZP_09296067 359793307	ZP_09296068 359793308	ZP_09296069 359793309	ZP_09296070 359793310
Mesorhizobium australicum WSM2073	No	YP_007302813 433772346	YP_007302814 433772347	YP_007302815 433772348	YP_007302816 433772349
Mesorhizobium ciceri biovar biserrulae WSM1271	No	YP_004141258 319781782	YP_004141259 319781783	YP_004141260 31978178	YP_004141261 319781785
Oceanicola sp. S124	No	ZP_09518460 372282424	-	ZP_09518461 372282425	ZP_09518462 372282426
Octadecabacter arcticus 238	No	ZP_05063840 254450403	-	ZP_05068329 254454892	ZP_05062978 254449541
Sinorhizobium medicae WSM419	plasmid pSM	YP_001314498 150377903	-	YP_001314497 150377902	YP_001314496 150377901
1st -*Sinorhizobium meliloti* 1021	pSymA	NP_436402 16263609	NP_436401 16263608	16263607 NP_436400	NP_436399 16263606
2nd - *Sinorhizobium meliloti* 1021	pSymA	NP_436064 16263271	NP_436063 16263270	NP_436062 16263269	NP_436061 16263268
Sinorhizobium meliloti AK83	chromosome 3	AEG57528 334099519	AEG57527 334099518	AEG57526 334099517	AEG57525 334099516
Sinorhizobium meliloti BL225C	pSINMEB01	YP_005718215 384532611	YP_005718214 384532610	YP_005718213 384532609	YP_005718212 384532608
Sinorhizobium meliloti	Draft-	ZP_12973398	ZP_12973399	ZP_12973400	ZP_12973401

CCNWSX0020	contig00009 Not determined	418399852	418399853	418399854	418399855
Sinorhizobium meliloti GR4	pRmeGR4c	YP_007192877 433616082	YP_007192878 433616083	YP_007192879 433616084	YP_007192880 433616085
Sinorhizobium meliloti Rm41	pSYMA	YP_006814644 407691060	YP_006814643 407691059	YP_006814642 407691058	YP_006814641 407691057
Sinorhizobium meliloti SM11	pSmeSM11c	AEH81905 336035974	AEH81906 336035975	AEH81907 336035976	AEH81908 336035977

[a]These are the only 18 organisms in the Genbank with an *ndmCDE* gene cluster as of March 11, 2013.
[b]The *ndmD* homolog in *P. putida* TJI-51does not contain an annotated N-terminal Rieske. The annotated start site is potentially wrong, as an upstream start site has the Rieske domain included. Proteins from this organism are not included in the phylogenetic trees.

Table S3. PCR Primers used in this study. Restriction sites are underlined, and mutations are italicized.

Primer	DNA sequence
gst-degR2	5'-TGGRAAAATCCCCGACATBAC-3'
gst-F1	5'-TAAGAGCTGCTACTGGTACC-3'
gst-F3	5'-CGTGCCCGAGATTTGTTGAG-3'
gst-F4	5'-CGCAGACATCGCCTGTTTCC-3'
Marcy	5'-CAGGAAACAGCTATGACC-3'
Marcy2	5'-GTTAGCTCACTCATTAGGCACC-3'
NdmC-KO-NF	5′-CCCTTTGGATCCAAATCGCGGTCGCCCAATTGCGAGG-3′
NdmC-KO-NR	5′-TTAGTCCCGCAGAGCACCATAAATTACTTGGTCAGTAGACAT-3′
NdmC-KO-CF	5′-ATGTCTACTGACCAAGTAATTTATGGTGCTCTGCGGGACTAA-3′
NdmC-KO-CR	5′-CCCCCCGAATTCAAGTTCGTATCCTGGACGTCCCGGA-3′
NdmD-KO-NF	5′-CCCTTTGGATCCCAGTGAGATGTTCGATATTGCAGAG-3′
NdmD-KO-NR	5′-GCCTCACAGATCGAGAACGATTTTGTTGACGTCAAGTTTGTTCACGGC-3′
NdmD-KO-CF	5′-GCCGTGAACAAACTTGACGTCAACAAAATCGTTCTCGATCTGTGAGGCC-3′
NdmD-KO-CR	5′-CCCCCCGAATTCATAGAAAAACAAGTCATGTAATCGCGCGG-3′
NdmE-KO-NF	5′-TTGACGGAATTCCGAGGAACGGTTTTCCAC-3′
NdmE-KO-NR	5′-CTATGCACTCACCCGTCGCGCATAGTCATAGAGTGTAATCACGG-3′
NdmE-KO-CF	5′-GTGATTACACTCTATGACTATGCGCGACGGGTGAGTGCATAG-3′
NdmE-KO-CR	5′-CCACTGTCTAGAAATAAGCGAGCACGCATC-3′
ndmC-F-NdeI	5'-GCGGCCCATATGTCTACTGACCAAGTAA-3'
ndmC-R-XhoI	5'-TTTAATCTCGAGGTCCCGCAGAGCACCATA-3'
ndmE-F-NdeI	5'-GGCGCGCATATGATTACACTCTATGACTATG-3'
ndmE-R-EagI	5'-ATTTTACGGCCGTTATGCACTCACCCGTCGCG-3'
NdmA-F-NcoI	5′-GCAAGGTCCATGGAGCAGGCGATCATCAATGATGA-3′
NdmA-R-BamHI	5′-CCTCCGGGATCCTTATATGTAGCTCCTATCGCTT-3′
NdmB-F-NdeI	5′-GCAAGGTCATATGAAAGAACAGCTCAAGCCGCTGC-3′
NdmB-R-KpnI	5′-CCGCCGGGTACCTTACTGTTCTTCTTCAATAAC-3′
Bsdm-NcoI-F	5′-CCAGTGGGCACGGCCC*A*ATGGGCGTCACCTTGC-3′
Bsdm-NcoI-R	5′-GCAAGGTGACGCCCAT*T*GGGCCGTGCCCACTGG-3′
Bsdm-NdeI-F	5′-GGATAATCAACCTGATCA*C*ATGCACCTGGCTTTCC-3′
Bsdm-NdeI-R	5′-GGAAAGCCAGGTGCATGT*G*ATCAGGTTGATTATCC-3′