Oak Ridge National Laboratory

Nitration enzyme toolkit for the biosynthesis of energetic materials

**SERDP Project Number
WP-2332**

PI: David E. Graham
Jim C. Spain
Ronald J. Parry
Robert L. Hettich

February 5, 2018

**Approved for public release.
Distribution is unlimited.**

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* 02-05-2018	2. REPORT TYPE Final Report	3. DATES COVERED *(From - To)* 01-09-2013 to 09-30-2017

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER WP-2332
Nitration enzyme toolkit for the biosynthesis of energetic materials	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Graham, David E; Spain, Jim C; Parry, Ronald J; Hettich, Robert L; Mahan, Kristina M; Klingeman, Dawn M; Giannone, Richard J; Gulvick, Christopher A; Fida, Tekle T	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Oak Ridge National Laboratory PO Box 2008, MS-6038 Oak Ridge, TN 37831	8. PERFORMING ORGANIZATION REPORT NUMBER ORNL/SPR-2017/498

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Strategic Environmental Research and Development Program 4800 Mark Center Dr, Suite 17D03 Alexandria, VA 22350-3605	10. SPONSOR/MONITOR'S ACRONYM(S) SERDP
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The most common secondary explosives and propellants contain nitro groups that are produced using chemical nitration reactions. Conventional manufacturing processes for these energetic materials often use hazardous and corrosive substances, such as nitric acid, and produce hazardous waste streams. Highly reactive nitration reactions can also create multiple isomers and by-products that degrade performance of the energetic products. To reduce the environmental impacts of these processes, new strategies are needed to produce energetic chemicals and precursors. This project identified new bionitration mechanisms used by microorganisms to produce nitro-containing natural products. We investigated biosynthetic pathways for 2-nitroimidazole (azomycin), *N*-nitroglycine, and nitrophenols. Five new enzymes were discovered to catalyze 2-aminoimidazole and 2-nitroimidazole biosynthesis in a revised pathway from *Streptomyces eurocidicus*. Stable isotope labeling experiments identified L-arginine and glycine as precursors for the nitramine *N*-nitroglycine in *Streptomyces noursei*. Comparative genomic and proteomic analyses identified a set of proteins that could be responsible for *N*-nitroglycine biosynthesis. Studies of *Salegentibacter* sp. demonstrated that nitrate reduction to nitrite was associated with bionitration of phenolic substrates, producing a diverse set of nitrophenols. This growing bionitration toolkit represents a diverse range of nitration mechanisms and products that can be adapted for the green chemistry production of nitro compounds and precursors.

15. SUBJECT TERMS
Nitration, Biosynthesis, Microorganisms, Energetic compounds, Nitroimidazole, Nitramine, Biosynthesis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON David Graham
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED	UU	56	19b. TELEPHONE NUMBER *(include area code)* (865)574-0559

SERDP Project WP-2332 Final Report

Nitration enzyme toolkit for the biosynthesis of energetic materials

Author(s)

David E. Graham
Jim C. Spain
Ronald J. Parry
Robert L. Hettich
Kristina M. Mahan
Dawn M. Klingeman
Richard J. Giannone
Christopher A. Gulvick
Tekle T. Fida

Date Published:
February 5, 2018

TABLE OF CONTENTS

List of Tables

List of Figures

List of Acronyms and Abbreviations

2-AI   2-Aminoimidazole
2-NI   2-Nitroimidazole
ANI   Average nucleotide identity
ANOVA   Analysis of Variance
ANTA   3-Amino-5-nitro-1,2,4-triazole
ATCC   American Type Culture Collection
bp   Base pair
BTN   1,2,4-Butanetriol trinitrate
CDS   Coding DNA sequence
DNAN   2,4-Dinitroanisole
DNI   2,4-Dinitroimidazole
DNU   Dinitrourea
DOE   Department of Energy
FDR   False Discovery Rate
GC   Guanine + cytosine base content
HC   Hierarchical Cluster
HMX   Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine
HPLC   High pressure liquid chromatography
HR-MS   High resolution mass spectrometry
LC-HRMS   Liquid chromatography-high resolution mass spectrometry
LC-MS   Liquid chromatography-mass spectrometry
MNA   *N*-Methyl-4-nitroaniline
MS/MS   Tandem mass spectrometry
MudPIT   Multidimensional Protein Identification Technology
NMR   Nuclear Magnetic Resonance
NNG   *N*-Nitroglycine
NQ   Nitroguanidine
NTO   3-Nitro-1,2,4-triazol-5-one
ORNL   Oak Ridge National Laboratory
PCA   Principal Component Analysis
PSM   Peptide Specific Matches
RAST   Rapid Annotation using Subsystems Technologies
RDX   Hexahydro-1,3,5-trinitro-1,3,5-triazine
RSD   Relative Standard Deviation
TATB   1,3,5-triamino-2,4,6-trinitrobenze
TLC   Thin layer chromatography
TNI   2,4,5-Trinitroimidazole
TNT   2,4,6-Trinitrotoluene

Keywords

Acknowledgements

ABSTRACT

Objective
The most common secondary explosives and propellants contain nitro groups that are produced using chemical nitration reactions. Conventional manufacturing processes for these energetic materials often use hazardous and corrosive substances, such as nitric acid, and produce hazardous waste streams. Highly reactive nitration reactions can also create multiple isomers and by-products that degrade performance of the energetic products. To reduce the environmental impacts of these processes, SERDP Statement of Need WPSON-13-04 requested new synthetic biology strategies to produce energetic chemicals and precursors. This project identified new bionitration mechanisms used by microorganisms to produce nitro-containing natural products, which could be applied to the future production of energetic materials.

Technical Approach
We investigated biosynthetic pathways for 2-nitroimidazole (azomycin), *N*-nitroglycine, and nitrophenols. Bacteria reported to the produce these compounds were obtained from culture collections, and we developed reliable growth conditions for the cells to produce these natural products in small bioreactors. A combination of genome sequencing, comparative genome analysis, proteomics, transcriptomics, and gene cluster analysis provided data to identify candidate genes that could encode bionitration enzymes. Heterologous gene expression in an *E. coli* host bacterium produced candidate proteins that were tested in bioassays or purified for in vitro biochemical assays of enzymatic activity. Reaction products were detected by chromatography and mass spectrometry.

Results
Five new enzymes were discovered to catalyze 2-aminoimidazole and 2-nitroimidazole (azomycin) biosynthesis in a revised pathway from *Streptomyces eurocidicus*. Stable isotope labeling experiments identified L-arginine and glycine as precursors for the nitramine *N*-nitroglycine in *Streptomyces noursei*. Comparative genomic and proteomic analyses identified a set of proteins that could be responsible for *N*-nitroglycine biosynthesis. Studies of *Salegentibacter* sp. demonstrated that nitrate reduction to nitrite was associated with bionitration of phenolic substrates, producing a diverse set of nitrophenols. Gene expression experiments and comparative genome analysis identified two clusters of genes associated with nitrophenol production.

Benefits
This growing bionitration toolkit represents a diverse range of nitration mechanisms and products that can be adapted for the green chemistry production of nitro compounds and precursors. This work demonstrates feasibility of applying bionitration enzymes to produce relevant nitrated precursors of energetic materials without the use of heavy metal catalysts or strong acids and without producing hazardous waste.

OBJECTIVE

More than 200 natural products containing nitro groups have been identified, establishing clearly that enzymes can catalyze both aromatic and aliphatic nitration reactions. They can even synthesize nitramines. Although intact cells produce the compounds, the enzymes that catalyze the reactions are mostly unknown. It is essential to advance the understanding of the biochemistry to provide the efficient, broad-specificity bionitration enzymes that will be required for the biocatalytic synthesis of nitro compounds on process scales.

The objective of this project was to identify and characterize new bionitration enzymes from bacteria that produce nitro compounds with structural similarity to energetic chemicals. This project benefits the Department of Defense (DoD) and manufacturers of energetic materials by providing a path to reduce the environmental impact of energetic chemical production. The proposed approach would identify and develop novel biocatalysts that form nitro groups in organic molecules. These enzymes will serve as modular parts for engineering biosynthetic pathways to produce energetic materials for use in explosives or propellants. This research specifically addresses the second objective in WPSON-13-04 for "broad-based research to develop the fundamentals of synthetic biology as related to energetic materials." The biological parts identified here constitute an enzyme toolkit for the future construction of green, biological pathways to produce nitro compounds through collaborations with materials chemists with expertise in process-scale synthesis of energetics. A broader repertoire of bionitration enzymes will enable new syntheses of fuel additives, pharmaceuticals, photochemicals, pesticides and dyes from renewable materials.

We proposed using systems biology techniques to identify new bionitration enzymes from: 1) *Streptomyces eurocidicus* that produce 2-nitroimidazole; 2) *Streptomyces noursei* that produces the rare nitramine compound *N*-nitroglycine; 3) *Salegentibacter* bacteria that produce an abundant variety of nitroaromatic, nitroaliphatic and nitroheterocyclic compounds; and 4) *Streptomyces* sp. UC 11065 that produces pyrrolomycins (Table 1). This research reveals a set of nitrating enzymes (and their specifications) as prototypes of new biocatalysts for protein engineering. These parts will constitute an enzyme toolkit for the future construction of green, biological pathways to produce nitro compounds through collaborations with energetic materials chemists with expertise in process-scale synthesis of energetic materials.

Table 1. Natural product nitro compound analogs to energetic compounds

Natural product compound(s)	Bacterial host	Energetic compound relevance
2-Nitroimidazole	*Streptomyces eurocidicus*	DNI, NTO propellants
N-Nitroglycine	*Streptomyces noursei*	Nitroguanidine (NQ) propellant, nitramines like HMX, RDX
Nitroaromatics	*Salegentibacter* spp.	DNAN, MNA secondary explosives
Pyrrolomycins	*Actinosporangium vitaminophilum*, etc.	DNI, DNAN

BACKGROUND

The nitrogen-based explosives and propellants currently in wide use and under development contain nitro groups in a variety of configurations including *N*-nitro, *O*-nitro, nitroarene and nitroalkane substituents. Some are readily synthesized using simple nitration chemistry (2, 3). However, these processes often require hazardous reagents, and they produce hazardous waste streams. Green chemistry approaches for the synthesis of energetic materials are currently being developed for sustainable manufacturing (4, 5). Some desirable energetic compounds are more difficult to synthesize because of constraints of organic reactions using simple reagents. For example, the classical nitration of toluene using nitric and sulfuric acids produces a mixture of TNT isomers and nitrated phenolic compounds: these impurities degrade TNT performance and decrease its stability (6). Biosynthetic processes usually produce single isomers. For example, *S. eurocidicus* cells produce a single isomer of 2-nitroimidazole, while the chemical nitration of imidazole compounds produces a mixture of 4-nitro and 5-nitroimidazole tautomers (7).

Over 200 natural nitro-substituted organic compounds have been reported to be produced by living organisms (8) and new compounds are reported regularly. They include analogs of all the above types of energetic materials. In a few instances, the biosynthesis mechanism is well established. For most, however, the biochemistry and molecular biology are unknown. We proposed to determine the mechanisms involved in synthesis of natural nitro compounds to harness the remarkable power of biotechnology for green chemistry synthesis of energetic materials with high nitrogen content.

Biocatalysis is playing a growing role in producing materials by green chemistry –often using renewable feedstocks (9, 10). Previous bioprospecting programs supported by the Office of Naval Research and SERDP developed biosynthetic pathways that produced organic precursors to several energetic compounds. Expression of genes from *Pseudomonas fluorescens* in *Escherichia coli* led to the formation of phloroglucinols, which are precursors of 1,3,5-triamino-2,4,6-trinitrobenze (TATB) and TNT (11, 12). 1,2,4-Butanetriol, a precursor of 1,2,4-butanetriol trinitrate (BTN) explosive plasticizer, was produced from sugars by four bacterial enzymes (13). Although both of these precursors required chemical nitration to produce energetic compounds (14), the successes demonstrated the value of synthetic biology approaches to make energetic materials.

Many natural products containing nitro groups have been identified (8, 15), and a complete cycle of nitro compound biosynthesis and biodegradation has naturally evolved. However, the various mechanisms of nitro compound degradation by bacteria are better understood than the biosynthetic pathways (16-19). Several modes of nitro group biosynthesis have been proposed (Figure 1). Currently the oxidation of amines to nitro compounds is best understood. A variety of dissimilar *N*-oxygenase enzymes catalyze the stepwise oxidation of amines to hydroxylamine, nitroso and nitro groups. A Rieske oxygenase catalyzes the oxidation of an aromatic amine to a nitro group in the biosynthesis of pyrrolnitrin (20). An unrelated enzyme, AurF, catalyzes the oxidation of *p*-aminobenzoate to *p*-nitrobenzoate for aureothin biosynthesis (21-25). Studies of the substrate specificity of AurF demonstrated activity on a variety of substituted *p*-aminobenzoate analogs (26, 27), and the crystal structure model of the homologous CmlI protein involved in chloramphenicol biosynthesis demonstrated the possible diversity of arylamine

substrates (28). Several flavoproteins catalyze the partial oxidation of amines to hydroxylamines (29, 30).
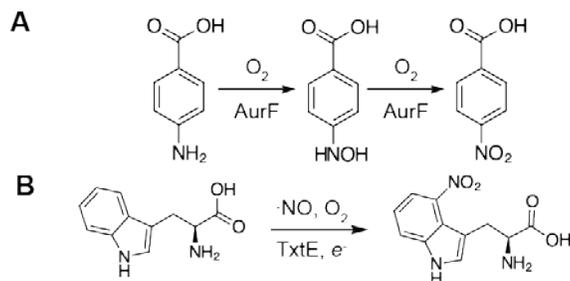
Figure 1. Biological and chemical nitration reactions

The N-oxygenase activity of AurF enzyme catalyzes the oxidation of p-aminobenzoate for the biosynthesis of aureothin (part B).The TxtE enzyme catalyzes the direct nitration of L-tryptophan for thaxtomin biosynthesis (part A).

The 4-nitrotryptophan moiety of thaxtomin phytotoxins from *Streptomyces* spp. is produced by a different mechanism, using nitric oxide produced by a cytochrome P450-type enzyme (Figure 1) (31-34). Proteins can also become nitrated when a nitrogen dioxide radical ($\cdot$NO$_2$) reacts spontaneously with aromatic amino acid side chains to form 3-nitrotyrosine and 6-nitrotryptophan (35, 36). Similar chemical reactions are responsible for lipid nitration (37). The $\cdot$NO$_2$ species is formed either by homolytic cleavage of peroxynitrite (ONOO$^-$) or by the oxidation of nitrite by a heme peroxidase enzyme (36). Other *Streptomyces* produce dioxapyrrolomycin by incorporating nitrate (1, 38) or reactive nitrogen species derived from the amino acids and the growth environment (39). Despite this recent progress in elucidating nitro group biosynthesis, known nitration enzymes cannot account for the remarkable diversity of nitro groups observed in natural products. The enzymes that produce nitroheterocycles, *N*-nitro and *O*-nitro substituents are a complete mystery; clearly, additional nitration systems remain to be discovered.

Nitroimidazole biosynthesis by *Streptomyces eurocidicus*

Several bacteria, including *Streptomyces eurocidicus*, produce 2-nitro-1*H*-imidazole (azomycin or 2-NI) (40). This compound has an oxygen balance of -78% (by weight), similar to many secondary explosives. It is a potential precursor for the synthesis of explosive, insensitive dinitro- or trinitroimidazole explosives and their energetic salts (41-44) (Figure 2). The nitroimidazoles are also analogs of thermally stable, high-energy triazole compounds including 3-nitro-1,2,4-triazol-5-one (NTO), a component of IMX-101 (45, 46).

Figure 2. 2-Nitroimidazole and analogous nitroimidazole energetic compounds

We recently identified a nitrohydrolase enzyme in a mycobacterium that converts 2-NI into an imidazolone (47). However, *Streptomyces* cells use a different pathway to make 2-NI by oxidizing 2-aminoimidazole (2-AI) (40, 48). Metabolic labeling studies using L-[U-^{14}C]arginine revealed it to be the imidazole precursor (49), supporting the biosynthetic pathway proposed in Figure 3. However, exogenous 4-hydroxyarginine was not incorporated in 2-AI biosynthesis, and guanidinobutyraldehyde was not identified as intermediate, raising questions about this proposed pathway. No enzyme involved in the proposed pathway had been isolated, and no genome sequence was available for any strain that produces 2-NI prior to this project. Therefore, a reverse genetic approach of genome sequencing and proteomic analysis was used to identify the requisite enzymes, focusing on the cyclase that produces 2-AI and the *N*-oxygenase or that oxidizes it to form 2-NI.
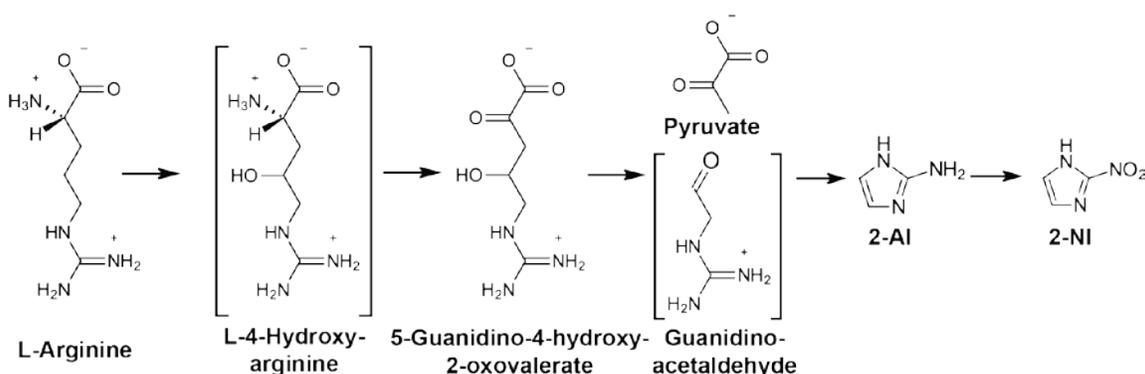
Figure 3. Proposed biosynthesis of 2-nitroimidazole from Nakane et al.

N-Nitroglycine (nitraminoacetic acid) biosynthesis by *Streptomyces noursei*

S. noursei produces *N*-nitroglycine (NNG), one of the few known nitramine natural products (50). Nitroglycine is an analog of the energetic materials nitroguanidine (NQ) and dinitrourea (DNU) (Figure 4). It has a relatively high oxygen balance (-26% by weight), and is predicted to decompose to $2CO + 2H_2O + N_2$, although the free acid was reported to be stable (50, 51). Nitramino ester compounds have been considered as high-energy plasticizers for nitropolymers (52). Successes of nitramine compounds like RDX, HMX and CL-20 suggest that nitroglycine and nitroguanidine analogs could be useful precursors for the synthesis of new insensitive energetic nitramines (53). Furthermore, nitramines decompose in sulfuric acid and could be used as alternative, green nitration reagents for the synthesis of other energetic materials (54). Although the biodegradation of nitramines has received considerable attention (55), these

reactions are unlikely to be reversible *in situ*. Therefore, elucidating the biosynthesis of NNG would be a significant contribution to the toolkit of high nitrogen materials biosynthesis.
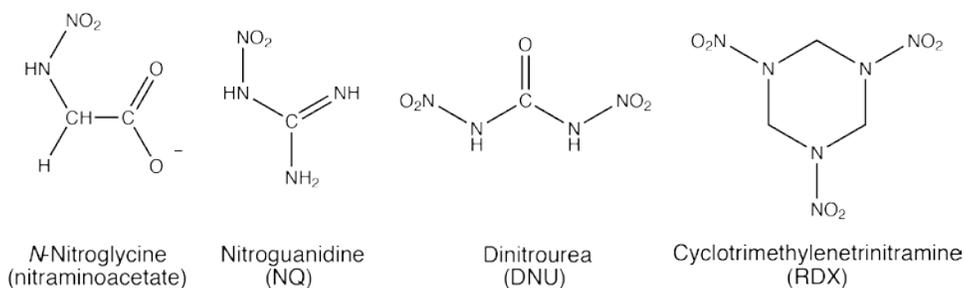
Figure 4. N-Nitroglycine and analogous nitramine energetics

Nitroaromatic biosynthesis by *Salegentibacter*

The arctic marine bacterium *Salegentibacter* sp. T436 produces almost 30 nitrophenol compounds (Figure 5) (56, 57). Some of these nitroaromatic compounds are thought to derive from tyrosine, although the nitration mechanism is unknown. Nitroaliphatic and nitroheterocyclic compounds are also produced by the strain so other mechanisms must also be involved. No genome sequence was available for any member of the *Salegentibacter* genus at the start of this project, and no genetic tools are currently available for this lineage. Therefore, a systems biology approach using high-throughput genome sequencing and transcriptomic analysis was used to identify new nitration enzymes from *Salegentibacter*. Identifying these broad specificity nitration enzymes would provide the basis for synthetic biology pathways to produce insensitive energetic compounds like 2,4-dinitroanisole (DNAN) or *N*-methyl-4-nitroaniline (MNA) (Figure 5) (58).

Nitration reactions in pyrrolomycin biosynthesis

The pyrrolomycins are produced by *Actinosporangium* and *Streptomyces* spp. bacteria (59, 60). The skeleton of dioxapyrrolomycin is derived from L-proline and three acetate molecules, while the methylenedioxy group originates from L-methionine (Figure 6) (61). Several of the pyrrolo- mycins, including pyrrolomycin A and dioxapyrrolomycin, contain a nitro group on C-4 of the

R=COOH
R=CH$_2$COOH
R=CH$_2$COOCH$_3$
R=CH$_2$CH$_2$COOH
R=CH$_2$CH(OH)COOCH$_3$
R=CH$_2$CH(Cl)COOCH$_3$
R=CH$_2$CH$_2$OH
R=CH$_2$CH$_2$NHCOCH$_3$
R=CH=CHNO$_2$

R=CH$_2$COOH
R=CH$_2$COOCH$_3$
R=CH$_2$CH$_2$COOH

R=CH$_2$CH(Cl)COOCH$_3$
R=CH$_2$CH$_2$OH
R=CH$_2$CH$_2$NHCOCH$_3$
R=CH=CHNO$_2$
R=CH$_2$CH$_2$Cl

2,4-Dinitroanisole
(DNAN)

N-Methyl-4-nitroaniline
(MNA)
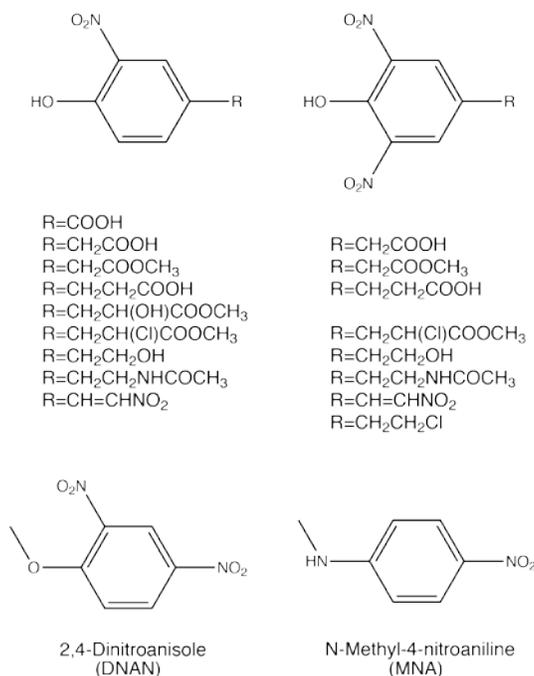
Figure 5. Nitrophenol compounds reported from Salegentibacter sp. T-436 and structurally related energetics

pyrrole ring. In order to gain insight into the nitration mechanism that introduces this functional group, we cloned and sequenced the pyrrolomycin biosynthetic gene clusters of *Streptomyces vitaminophilum* and the genetically transformable *Streptomyces* sp. UC 11065 (1). The identity of the cluster was confirmed by the single crossover disruption of two pyrrolomycin biosynthetic genes (1).
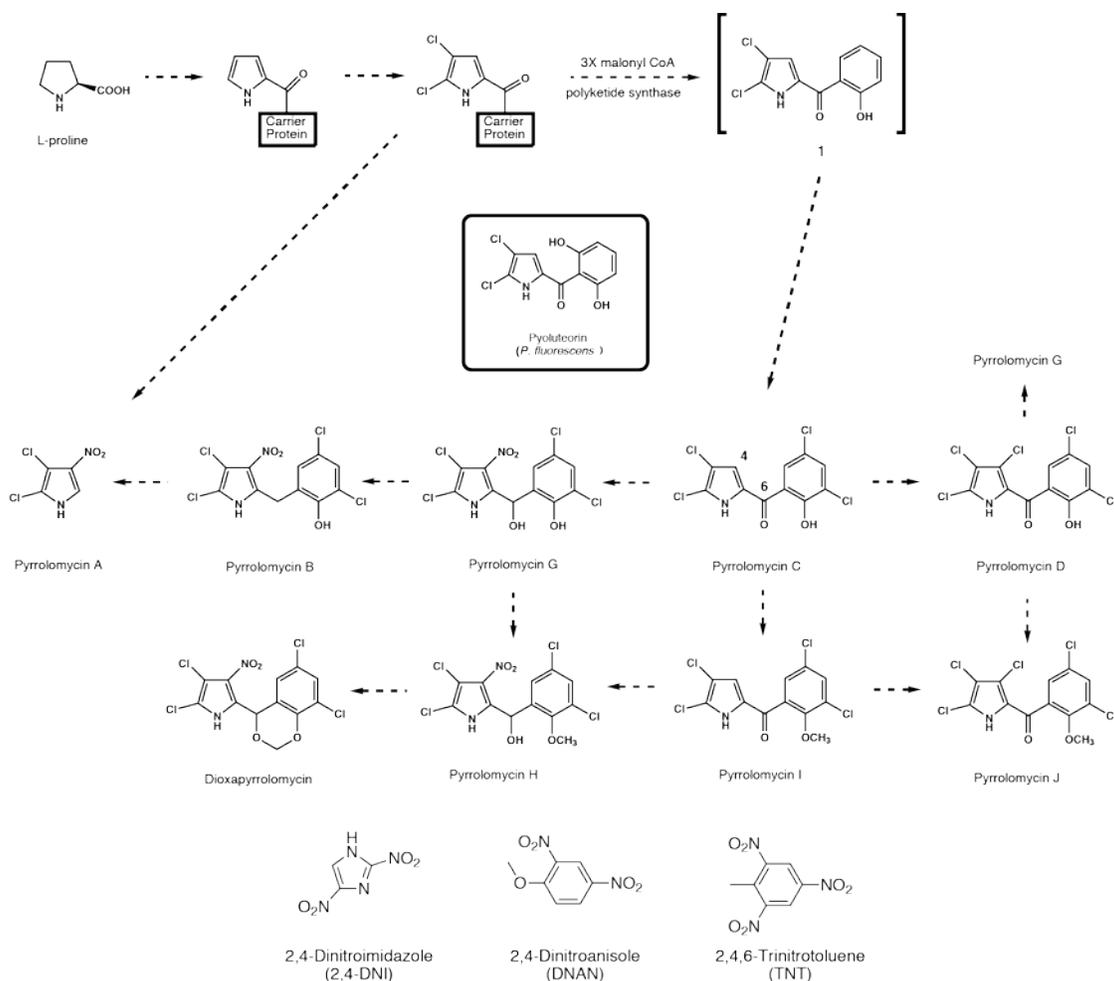
Figure 6. Hypothetical pathway for the biosynthesis of pyrrolomycins produced by S. vitaminophilum, Streptomyces fumanus and Streptomyces sp. UC11065

Multiple enzymes from Streptomyces spp. are proposed to convert L-proline into the halogenated precursor pyrrolomycin C (1). Subsequent reactions including a predicted electrophilic nitration reaction convert pyrrolomycin C into the nitro-pyrrole compounds pyrrolomycin A and dioxapyrrolomycin. Analogous aromatic nitration reactions could be used in the biosynthesis of insensitive energetic compounds.

Sequence analysis of the two pyrrolomycin biosynthetic gene clusters does not reveal the mechanism of the nitration reaction. Both clusters encode the two components of an assimilatory nitrate reductase, strongly suggesting that nitrite, formed by the reduction of nitrate, is required for the nitration reaction. However, sequence analysis fails to reveal how nitrite might be utilized to nitrate the pyrrole ring. The C-4 position of the precursor compound, pyrrolomycin C, should be the most reactive position in the molecule towards an electrophilic nitrating agent. Since the nitration occurs at this position, we hypothesize that the nitrating enzyme(s) may not play a role in determining the regiospecificity of the nitration reaction.

Consequently, if the enzyme has broad substrate specificity, then it might accept a variety of electron-rich aromatic compounds as substrates for nitration. This possibility was explored using by adding potential substrates to the fermentation broth. Potential substrates include pyrrole-2-carboxylic acid, phenol, resorcinol, and phloroglucinol (11). A successful biological nitration of

8

phloroglucinol would be especially interesting since 2,4,6-trinitrophloroglucinol can be converted chemically to the heat- resistant, insensitive explosive TATB in two steps (62). A broad specificity nitration enzyme would be advantageous for the synthesis of many other nitroaromatic compounds discussed here, including DNAN, NTO, TNI and TNT.

Project Organization and Workflow

This project was designed with four parallel tasks to identify pathways and biocatalytic proteins responsible for nitrocompound biosynthesis:

Task 1. Nitroimidazole biosynthesis from *S. eurocidicus*
Task 2. *N*-Nitroglycine biosynthesis from *S. noursei*
Task 3. Nitroaromatic biosynthesis from *Salegentibacter*
Task 4. Pyrrolomycin nitration from *A. vitaminophilum*

Each task would pursue a similar workflow (Figure 7), depending on strains, substrates and analytical methods needed to monitor the bionitration processes. Due to the inherent technical risk in discovery-based projects, we anticipated that one or more tasks would be terminated at a Go/No-Go decision point to focus resources on the most promising pathways.

Materials & Methods common to the four tasks are presented in the following section. For clarity, Results & Discussion are divided by task in this report. Finally, a conclusions section synthesizes results from this project and identifies opportunities for future development benefitting DoD programs.
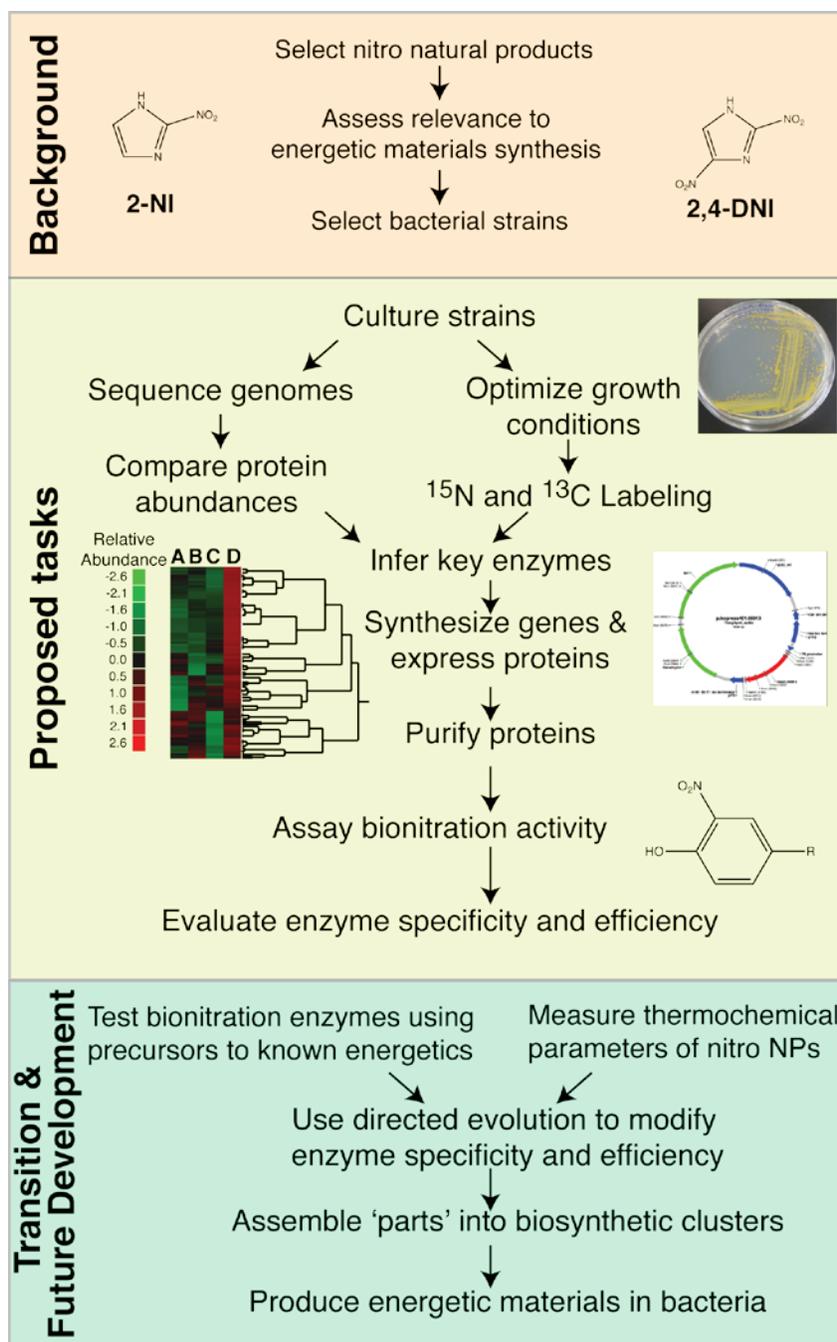
Figure 7. Sample workflow for the identification and characterization of bionitration enzymes in this project

MATERIALS AND METHODS

Chemicals

N-nitroglycine (NNG) was purchased from AKos Consulting & Solutions (Deutschland GmbH, Germany). NMR and MS analyses are reported in Appendix A. L-[*amide*-^{15}N] Glutamine (98%+), L-[*guanidino*-^{15}N$_2$]arginine·HCl (98%+), L-[*amide*-^{15}N]asparagine (98%+), ^{15}NH$_4$Cl (99%), and K^{15}NO$_3$ (99%) were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). [^{13}C$_2$]Glycine (97-99%), [2-^{13}C,^{15}N]glycine (99% ^{13}C, 98% ^{15}N), and 4-amino[*phenyl*-^{13}C$_6$] benzoic acid (99%) were purchased from Sigma-Aldrich (St. Louis, MO). Specialty chemicals 2-aminoimidazole sulfate, naladixic acid, L-(+)-arabinose, L-lysine, and L-arginine were purchased from Alfa Aesar. 2-Nitroimidazole, apramycin sulfate, L-argininamide, and L-glutamine were purchased from Sigma-Aldrich. Pyridoxal-5'-phosphate and D-arginine were purchased from USB. L-Homoarginine was purchased from MP Bio. All other reagents were purchased through common distributors.

Bacterial strains and cultivation

Strains and their sources are listed in Table S1 of Appendix A. To monitor the production of 2NI, strain ATCC 27428 was grown at 30 °C in a 10 L bioreactor with mechanical stirring at 200 rpm in 700 mL of complex medium containing 1% malt extract, 0.4% yeast extract, and 0.4% glucose (GYM) (pH adjusted to 7.4 with sodium hydroxide). At appropriate intervals samples of cells were harvested by centrifugation (2600 *g* for 20 min), washed with an equal volume of 50 mM sodium phosphate buffer (pH 7.4), and stored at -80 °C. Samples from the 20 and 40 h time points were later thawed and used for proteomic analyses. Growth was measured by the dry weight of washed and filtered cells.

The *Streptomyces noursei* JCM 4701 strain derived from strain 8054-MC$_3$ (50) was provided by the RIKEN BRC through the National Bio-Resource Project of the MEXT, Japan. To monitor the production of NNG and for proteome analysis, strain JCM 4701 was grown at 30 °C in 2 L flasks with mechanical shaking at 160 rpm in 600 mL of medium containing 1% malt extract, 0.4% yeast extract, and 0.4% glucose (GYM) (pH adjusted to 7.4 with sodium hydroxide). At appropriate intervals samples of cells were harvested by centrifugation (2600 *g* for 20 min), washed with an equal volume of 50 mM sodium phosphate buffer (pH 7.4), and stored at -80 °C. Samples from the 26 and 67 h time points were later thawed and used for proteomic analyses. Growth was measured by the dry weight of washed and filtered cells. Cultures of related *Streptomyces* strains were obtained from the USDA Agricultural Research Service Culture Collection (NRRL) and grown as described in Supplemental Materials (63).

Salegentibacter spp. were grown in modified B1 medium (56) in flasks or a 2-L bioreactor at 25 °C. The medium contained per liter: NZ Amine (2.5 g), beef extract (3.8 g), yeast extract (1.0 g), KNO$_3$ (1.5 g), and marine salts (33.3 g) at pH 8.0. The bioreactor (Bio Controller ADI 1010 (Applikon Biotechnology, Delft, TheNetherlands) was operated at a working volume of 1.5 liters, aerated with 0.15 liters of air per minute and agitated at 200 rpm.

Streptomyces vitaminophilus were was grown at 28°C in 250 mL flasks with mechanical shaking at 250 rpm in 100 mL seed medium (64) (1% glucose, 1% soluble starch, 0.5% polypeptone, 0.3% yeast extract, 0.2% soy flour, 0.2% beef extract, and 0.2% $CaCO_3$ (pH 7.0) as previously described (1). The strain was also maintained on 2% agar plates without $CaCO_3$ and 0.2% soytone was used in place of soy flour.

Analytical methods

Analysis of 2-aminoimidazole and 2-nitroimidazole production

Culture filtrates were acidified and analyzed for 2-AI and 2-NI with an Agilent 1260 Infinity HPLC. Chromatography was performed at 30 °C on a Synergi 4 µm Hydro-RP 80 Å (250 mm × 4.6 mm i.d.; Phenomenex) column using injection volumes of 10-20 µL. Compounds were eluted using a mobile phase consisting of 5% acetonitrile in 0.1% H_3PO_4 at a flow-rate of 0.8 mL/min. Alternatively, an isocratic method using 30% methanol in 0.1% H_3PO_4 at a flow-rate of 0.5 mL/min was used. 2-AI was detected by its absorbance at 210-214 nm, and 2-NI was detected by its absorbance at 324 nm, relative to standard samples.

Liquid chromatography-high resolution mass spectrometry (LC-HRMS) was used to detect 2-NI and intermediates (Table S2 in Appendix A). This method used a Dionex u3000 HPLC in line with an Orbitrap Pro mass spectrometer (Thermo Fisher). Samples of filtered supernatants were loaded onto a 300-nL fused silica sample injection loop (50 µm I.D.) plumbed into the 6-port switching valve of the Orbitrap Pro, operated in positive ion mode. Samples were injected onto an in-house pulled nanospray emitter (150 µm I.D. fused silica) filled with 25 cm reversed-phase (RP) material (Kinetex, 5 µm, C18; Phenomenex) and separated by isocratic elution (65% water, 25% acetonitrile, 10% isopropanol with 5 mM glacial acetic acid – all solvents LC-MS grade) over 10 min at 500 nL/min (split flow via microtee employed to achieve nanoliter flow rates). Extracted ion chromatograms (5-10 ppm window) were screened for distinct peaks corresponding to target analytes. Additional *Streptomyces* spp. were screened for 2-NI production using high-resolution LC-MS performed at the Metabolomics Center of the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign (K. Ju, personal communication).

Analysis of *N*-nitroglycine production

Culture filtrates were analyzed for NNG with an Agilent 1260 Infinity HPLC (Memphis, TN). Chromatography was performed at 30 °C on a Synergi 4 µm Hydro-RP 80 Å (250 mm × 4.6 mm i.d.; Phenomenex, Torrance, CA) column using injection volumes of 20 µL. Compounds were eluted using a mobile phase consisting of 10% methanol in 0.1% H_3PO_4 at a flow-rate of 0.5 mL/min, and NNG was detected by its absorbance at 230 nm. Nitrite concentrations were determined using the Griess assay. Protein concentrations were determined with a Pierce BCA protein assay reagent kit (Rockford, IL).

NNG production was also monitored by high-resolution liquid chromatography-mass spectrometry (LC-MS) as described above, in negative ionization mode. Eluting metabolites were desolvated by negative nanospray ionization (NSI) and measured by the Orbitrap Pro operating at 15,000 resolution. Negative ions were monitored in the 100-180 *m/z* range. Samples

were run in technical duplicate with water blanks between sets. Extracted-ion chromatograms of peaks corresponding to the molecular ion or a characteristic ion fragment formed by neutral loss of HNO_2 were used to detect NNG and each potential labeled variant of NNG (Xcalibur v2.2; 5 ppm mass tolerance, 7-point boxcar smoothing). *Streptomyces* spp. related to *S. noursei* were screened for NNG production using high-resolution LC-MS performed at the Metabolomics Center of the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign (K. Ju, personal communication).

Analysis of nitrophenol production

Nitrophenols produced by *Salegentibacter* sp. T436 were extracted from culture supernatants using ethyl acetate and evaporated to dryness. Resuspended samples were applied to TLC plates for separation. Nitrophenols were also detected by HPLC with UV detection. Negative ion mode LC-HRMS at the University of Tennessee Biological and Small Molecule Mass Spectrometry Core facility.

Whole cell biotransformations of 2-AI oxidation

E. coli strains expressing *S. eurocidicus* genes were grown to mid-logarithmic growth phase in LB medium supplemented with appropriate antibiotics at 30 °C, with moderate shaking at 148 rpm. Expression was induced by the addition of 0.07 to 0.2% L-arabinose, and cells were harvested by centrifugation after an additional 10 hr of growth. Cells were washed in an equal volume of 50 mM sodium phosphate buffer (pH 7.4), and resuspended in buffer. To monitor production of 2-AI, 1 to 3 mM L-arginine was added, and a 1-mL mixture was incubated in glass tubes (18 mm by 15 cm) with shaking at 30 °C. To monitor production of 2-NI, freshly prepared L-cysteine and iron(II) ammonium sulfate were added at 0.2 mM concentrations to mid-log phase cells along with L-arabinose inducer. Washed cells were mixed with 7 mM D-glucose or succinate and 0.7 mM 2-AI was added. Biotransformation reactions were terminated by the addition of 0.07% H_3PO_4 and centrifugation to remove cell material for HPLC analysis.

Isotope incorporation into *N*-nitroglycine

A 5-mL starter culture of *S. noursei* JCM 4701 grown for 24 h was inoculated into 100 mL of GYM medium in a 250 mL flask and incubated at 30 °C with mechanical shaking at 160 rpm. After 48 h of growth, cells were harvested by centrifugation (1600 *g*, 10 min) and washed three times with sodium phosphate (50 mM) buffer at pH 7.4. The cells were then re-suspended in four times the volume of the wet cell pellet in phosphate buffer. The cell suspension was then divided into 1-mL volumes, and the indicated labeled substrate was added to a final concentration of 1 mM. Cells were incubated with labeled substrates for 24 h at 30 °C with mechanical shaking at 160 rpm in 18 by 150 mm glass culture tubes incubated vertically. Cultures were then centrifuged and supernatants filtered through 0.22 µm filters prior to LC-MS analysis. For the experiments that include the addition of two labeled substrates, the protocol was the same as described above except that after 48 h of growth, cells were suspended in GYM medium, and the indicated substrates were added. Cells were grown in the presence of labeled substrates for an additional 24 h as described above.

Genomic DNA purification

Strain ATCC 27428 was grown in GYM medium. After the medium was autoclaved, 5 mM MgCl$_2$ and 0.5% (w/v) glycine were added prior to inoculation. Cells were harvested after 48 h of growth. Genomic DNA was isolated from harvested cells using the QIAamp DNA Mini and Blood Mini kits (QIAGEN), following the manufacturer's protocols for isolation of genomic DNA from Gram-positive bacteria. The quality and quantity of the DNA was assessed by UV absorbance using a Nanodrop instrument (Thermo Scientific), fluorescence using a Qubit instrument (Life Technologies), and agarose gel electrophoresis.

Strain JCM 4701 was grown in YEME medium containing 0.3% yeast extract, 0.5% Bacto peptone, 0.3% malt extract, 1% glucose, and 34% sucrose broth (65). After the medium was autoclaved, MgCl$_2$ (5 mM) and glycine (0.5%) were added prior to inoculation. Cells were harvested after 48 h of growth. Genomic DNA was isolated from harvested cells using the QIAamp DNA Blood Mini kit (QIAGEN, Valencia, CA), following the manufacturer's protocols for isolation of genomic DNA from Gram-positive bacteria. The quality and quantity of the DNA was assessed by UV absorbance using a Nanodrop instrument (Thermo Scientific), and fluorescence using a Qubit instrument (Life Technologies) and agarose gel electrophoresis.

Salegentibacter spp. were grown in marine broth prior to DNA extraction. Genomic DNAs from the cells were extracted using a *Quick*-DNA fungal/bacterial miniprep kit (Zymo Research) or a Promega Wizard kit following the manufacturers' protocols for Gram-positive bacteria. The quality and quantity of the DNA was assessed by UV absorbance using a Nanodrop instrument (Thermo Scientific), and fluorescence using a Qubit instrument (Life Technologies) and agarose gel electrophoresis

Gene cloning and expression

Candidate genes identified through bioinformatics analysis of nucleotide sequences of fosmids showing NNG transformation were subcloned into the *E. coli* expression vector pBAD/HisA (Invitrogen) using overlapping oligonucleotide primers, genomic DNA templates, Q5 High-Fidelity DNA polymerase, and the NEBuilder HiFi DNA Assembly Cloning kit (New England Biolabs) according to the manufacturer's instructions. The resulting plasmid vectors (Table S3 in Appendix A) were transformed into *E. coli* NEB5α or LMG194 strains for heterologous expression and screening.

Oligodeoxyribonucleotide primers (IDT) were used to amplify DNA sequences from *S. eurocidicus* ATCC 27428 genomic DNA. Purified products were assembled with the backbone of *E. coli* expression vector pBAD-HisA (Invitrogen) using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). Gene sequences optimized for *E. coli* codon usage were synthesized as gBlock gene fragments (IDT) and assembled into expression vector pET-19b (Novagen), as described above. *Streptomyces* vector pSOK806 (66) and a PCR product of *S. eurocidicus aznABCDE* genes were used to construct integration vector pDG734. Sanger sequencing confirm construction of the recombinant vectors, performed at The University of Tennessee Genomics Core (Knoxville). Vectors containing native *S. eurocidicus* sequences were transformed into *E. coli* Rosetta 2 (DE3) cells (Novagen) to enhance expression.

Conjugations

Intergeneric conjugation from *Escherichia coli* ET12567 (pUZ8002) pDG734 into *Streptomyces* strains was performed as described previously (65) with the following modifications. Donor cells were grown in LB containing 25 mg/L kanamycin, 25 mg/L chloramphenicol, 50 mg/L apramycin, and 10 mM MgCl$_2$. A mixture of donor cells and *Streptomyces* spores was spread on ISP4 agar plates supplemented with 10 mM MgCl$_2$ and incubated at 30 ℃. The following day, 1 mg apramycin in 0.5 ml water was evenly overlaid on the plate. Isolated colonies were re-streaked on ISP4 or GYM agar containing 50 mg/L apramycin to isolate exconjugants.

Genome sequencing

Genomic DNA was used to prepare libraries for sequencing using Illumina short-read technology, Pacific Biosciences long-read technology, or a hybrid combination of both. Genome sequence accession numbers in GenBank are listed in Table S1 of Appendix A.

The *S. eurocidicus* ATCC 27428 genome sequence was assembled from Illumina MiSeq paired-end reads and Pacific Biosciences single molecule, real-time DNA sequencing data using the SPAdes genome assembler (67) and the SMRT Analysis scaffolding protocol (68).

The *S. noursei* JCM4701 genome was sequenced using an Illumina TruSeq library prepared as described in the manufacturer's low-throughput protocol (Part# 15005180 RevA). Illumina MiSeq sequencing produced 11,283,395 reads with a mean sequence quality of 36 (Phred score) and 70% G+C, calculated using FastQC (ver. 0.11.2; Babraham Bioinformatics). These data provided a mean coverage > 300. The paired sequences were trimmed of adapters and low-quality bases using cutadapt (ver. 1.7.1) (69). Pacific Biosciences single molecule, real-time DNA sequencing produced 141,764 filtered reads with an N50 = 6370 bp and read quality of 0.808. These data provided a mean coverage of 41.

The *S. noursei* JCM 4701 genome sequence was assembled from Illumina MiSeq paired-end reads and Pacific Biosciences single molecule, real-time DNA sequencing data using the SPAdes genome assembler (67) and the SMRT Analysis scaffolding protocol (68). The SPAdes genome assembler (ver. 3.6.0) was used to correct and assemble paired Illumina reads with a k-mer length of 127 (67). The resulting 79 contigs were combined using the RS_AHA_Scaffolding protocol of Pacific Biosciences's SMRT Analysis software (ver. 2.3.0) to produce 8 scaffolds (68). The scaffolded assembly was corrected with Pilon software (ver. 1.3) (70) using paired-end sequences assembled using PEAR (ver. 0.9.6) (71). The QUAST (ver. 3.0) and CheckM (ver. 1.0.5) software assessed *de novo* genome assembly quality (72, 73).

Genome annotation and comparative genomics

Coding DNA sequences (CDS) were identified and annotated by the NCBI Prokaryotic Genome Annotation Pipeline, the RAST server (74) and Prodigal software (ver. 2.6.2) (75). All software was executed using default parameters unless otherwise specified. Reciprocal best hits were calculated using the USEARCH algorithm (76) and custom Perl scripts to identify putative orthologs through pairwise proteome comparisons at 65-75% amino acid identity levels. For

strain JCM4701, the UCLUST algorithm implemented in USEARCH (ver. 8.0.1623) was used to cluster the predicted protein sequences at 98% sequence identity (76), and the 9468 centroid sequences formed a library for protein MS spectral identification. Protein domains were identified using HMMER (hmmer.org) with the Pfam database. Potentially secreted proteins were identified using the SignalP program (77), and transmembrane proteins were predicted using TMHMM (78). Prophage genes were identified using the PHAST web server (79). Putative secondary metabolite gene clusters were identified using the antiSMASH web server (ver. 3) (80). Hidden Markov model searches were performed using the HMMER (ver. 3.1b2) software (Eddy, 2011) with profiles from the PFAM database (ver. 28.0) (81). Average nucleotide identity between genomes was calculated using BLAST and Perl scripts (82). A phylogeny of *Streptomyces* spp. closely related to strain ATCC 27428 was inferred by maximum likelihood methods using 5-gene multilocus sequence tag analysis (83). A scoring matrix weighted by gene conservation was created in Microsoft Excel to prioritize gene targets for evaluation.

Proteomic and transcriptomic analyses

S. eurocidicus **ATCC 27428**

Proteins were extracted from cell pellets, digested with trypsin, and analyzed using MudPIT LC-MS/MS as described previously (84). Peptide fragmentation spectra (MS/MS) were searched against a *S. eurocidicus* ATCC 27428 proteome database concatenated with common contaminants and reversed sequences to control false-discovery rates using MyriMatch v.2.1 (85). Peptide spectrum matches (PSM) were filtered by IDPicker v.3 (86) to ensure a peptide-level false discovery rate (FDR) < 1% and assigned matched-ion intensities (MIT) based on observed peptide fragment peaks. PSM MITs were summed on a per-peptide basis and only those uniquely matching a particular protein were moved onto subsequent analysis with InfernoRDN (http://omics.pnl.gov/software/*InfernoRDN*). Peptide intensity distributions were log2-transformed, normalized across biological replicates by LOESS, and standardized by median absolute deviation and median centering across samples. Peptide abundance trends were aggregated to the protein-level via RRollup with the following parameters: 2 minimum peptides per protein, 50% minimum occurrence of peptide across samples, 3 peptide minimum occurrence for scaling, outliers removed. For ANOVA, values missing from the protein abundance matrix were imputed with a normal distribution of low-end values, as determined by Perseus analysis software (87), using a downshift of 1.8 and width of 3. Sample-to-sample variation was visualized by principal component analysis, Pearson's correlation, and hierarchically clustered using the Ward agglomeration method to generate a heat map of protein abundance trends normalized by z-score. Protein abundances were then compared across culture sampling points and significant differences highlighted.

S. noursei **JCM 4701**

Proteins were extracted from *S. noursei* cell pellets via sonic disruption (Branson) in SDS lysis buffer (4% SDS, 100 mM Tris-HCl, pH 8) and quantified by BCA assay (Pierce). The resulting crude protein extracts were then reduced with 25 mM dithiothreitol and boiled. Three milligrams of crude protein extract was precipitated with 20% trichloroacetic acid and pelleted by

centrifugation. Protein pellets were then cold acetone-washed, air-dried, and resolubilized in 8 M urea, 100 mM Tris-HCl, 5 mM DTT, pH 8. Once the pellets were fully resolubilized, samples were incubated at room temperature (RT) for 1 h then adjusted to 15 mM iodoacetamide to alkylate cysteines thus preventing disulfide bond reformation. Denatured, reduced, and alkylated proteins were then digested to peptides with sequencing-grade trypsin (Promega). Peptides were then salted (200 mM NaCl), acidified (0.1% formic acid), and filtered through a 10 kDa MWCO spin-filter. Tryptic peptides were quantified by BCA assay (Pierce) and 25 µg loaded via pressure cell onto a biphasic MudPIT column for online 2D HPLC separation (strong-cation exchange and reversed-phase) and concurrent analysis via nanospray MS/MS using a hybrid LTQ-Orbitrap XL mass spectrometer (Thermo Scientific). Mass spectral analysis and peptide searching was performed as described above.

RNAseq analysis of *Salegentibacter* sp. T436 gene expression

Cells were collected at the indicated time points and stored frozen in RNA*later* RNA stabilization reagent (QIAGEN). Cells were pelleted by centrifugation of the thawed solution and resuspended in 250 µL of SET buffer (50 mM Tris-HCl pH 8.0, 50 mM EDTA, and 25% w/v sucrose) by pipetting and vortexing.

For RNA isolation cells in RNA*later* were pelleted and supernatant was removed. Cell pellets were resuspended in 250 µL of lysozyme (20 mg/ml) and incubated at 37 °C for 8 mins with intermittent vortexing. The aqueous phase containing RNA was further processed beginning with the addition of QIAGEN buffer RLT with a QIAGEN RNeasy mini kit, following the manufactures instructions, including the on-column DNaseI treatment. Quantity of RNA was measured with a NanoDrop ND-1000 spectrophotometer (Thermo Scientific-NanoDrop Technologies) and quality was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies) on a RNA 6000 Nano Chip. High quality total RNA (RIN>7.8) was depleted of rRNA using Ribo-Zero rRNA Removal Kit for bacteria (Epicentre-Illumina) following the manufacturer's protocol. The depleted sample was purified on an RNA Clean & Concentrator-5 (Zymo Research) following the manufacturer's protocol.

Depleted RNA was used for RNA-Seq library preparation with the Epicentre ScriptSeq Complete kit for Bacteria following the manufacturer's protocol (EPILIT345 Rev.A). Agencount AMPure beads (Beckman Coulter) were used to purify the cDNA, and unique indexes were added during 13 cycles of library amplification. The final RNA-Seq libraries were purified with Agencount AMPure beads (Beckman Coulter) and quantified with a Qubit fluorometer (Life Technologies). The library quality was assessed on a 2100 Bioanalyzer with a DNA 7500 DNA Chip (Agilent). Final libraries were pooled and diluted to a working concentration. Fifty cycles of single end sequencing was completed on an Illumina HiSeq 2500 platform using V4 chemistry (HudsonAlpha Genomic Services laboratory, Huntsville, AL).

Protein purification and iron-sulfur cluster reconstitution
S. eurocidicus proteins were heterologously expressed in *E. coli* with N-terminal polyhistidine tags and purified by nickel affinity chromatography using standard methods (88).

RESULTS AND DISCUSSION

Biosynthesis of 2-nitroimidazole

The growth conditions for *S. eurocidicus* ATCC 27428[T] production of 2-NI were optimized, and a bioreactor system was implemented for the cultivation of dispersed cells under controlled conditions. Analytical HPLC and positive ion mode HR-MS methods confirmed 2-NI production using a commercially synthesized standard for reference. In addition to 2-NI (*m/z* 114.0292 observed (+1 charge state), 114.0298 expected, -5.3 ppm), the cells also produced the antibiotics eurocidin E (*m/z* 780.4126 observed (+1 charge state), 780.4165 expected, -5.0 ppm) and abundant tertiomycin (*m/z* 437.7465 observed (+2 charge state), 437.7434 expected, -7.1 ppm) (89). As reported previously, L-arginine was converted by cell-free extract of *S. eurocidicus* to 2-AI, but no 2-NI was produced. This activity was retained in ammonium sulfate-precipitated fractions of lysate. Only intact cells converted a substantial amount of 2-AI to 2-NI.

We sequenced the ATCC 27428 genome to identify potential bionitration genes and build a foundation for proteomic analysis. A hybrid, polished assembly of both short-read and long-read data included 7,931,574 bases on 18 contigs (Table 2). At least 5993 coding DNA sequences were identified in the genome, but none had significant sequence similarity to known *N*-oxidation enzymes that form nitro groups such as AurF or PrnD (90, 91). A homolog of the Rub8N nitrosynthase (92) was identified with 35% amino acid identity, although this family also includes numerous acyl-CoA dehydrogenase enzymes.

Table 2. Genome sequence assembly for 2-NI biosynthesis

Organism	*S. eurocidicus* ATCC 27428	*S. albireticuli* NRRL B-1670
GenBank Accession No.	LGUI000000	NSJV00000000
Sequencing method	Illumina TruSeq paired-end and PacBio	Illumina Nextera paired-end
Mean depth of coverage (fold)	220	434
Total length (bp)	7,931,574	8,454,574
# contigs	18	714
GC[1] (%)	72.6	72.4
N50	801,731	21,466
L50	4	117
CDS[1]	5993	7445

[1] Abbreviations: GC, guanine + cytosine base content, CDS, coding DNA sequences

The most closely related cultured organisms to this strain according to multi-locus sequence analysis include *Streptomyces eurocidicus* ATCC 19551 and *Streptomyces albireticuli* NRRL B-1670[T] (aka *S. eurocidicus* ATCC 19721) (83). Batch cultures of all three strains released 2-NI in growth media, confirmed by HPLC and liquid chromatography high-resolution mass spectrometry (LC-HRMS). Therefore, we assumed that 2-AI and 2-NI biosynthesis genes were conserved in their genomes. Draft genome sequences of ATCC 19551 (K. Ju, personal communication) and NRRL B-1670 strains (Table 2) were assembled from short-read data. The two *S. eurocidicus* strains share 94.5% average nucleotide identity (ANI), while the *S.*

albireticuli strain shares 91.7% ANI with *S. eurocidicus* ATCC 27428. These three genomes share 4717 homologs with at least 60% amino acid sequence identity.

Several related *Streptomyces* strains did not produce 2-NI in the tested growth conditions, including *S. eurocidicus* NRRL B-1677, *S. lydicus*, *S. varsoviensis*, *S. mobarensis*, *S. roseoverticullatus,* and *S. chattanoogensis*. The B-1677 strain appears more similar to *S. roseoverticullatus* than *S. eurocidicus* (83). Genes conserved in both 2-NI producing and non-producing strains were considered less likely to be responsible for 2-NI biosynthesis. However, it is possible that some of these strains contain cryptic 2-NI biosynthetic genes that were not expressed.

S. eurocidicus cells were grown in 10-L bioreactors for proteomic analysis to identify proteins whose abundance correlated with 2-NI biosynthesis (Figure 8). 2-NI began to accumulate in the supernatant after 20 h of growth when cells entered stationary phase (Figure 9). Cells were removed from the bioreactors to measure 2-AI oxidase activity in biotransformations; maximum activity was observed at 42 h. Based on this profile, cells were harvested from bioreactors containing either complex or defined media after 20 h, before the onset of 2-AI oxidase activity and 2-NI accumulation, for comparison with cells collected at 46 h during 2-NI production (Figure 9). Proteins extracted from duplicate samples were digested and analyzed using MudPIT LC-MS/MS. Peptide spectra were matched against proteins predicted from the genome sequence. ANOVA identified proteins with significant differential abundance at the two sample points from among the 2,420 quantifiable proteins. The abundances of 468 proteins increased at least two-fold in complex media, compared to 218 in defined media. 402 of these genes were conserved in genomes of all three 2-NI producing strains, of which 182 were missing from the genomes of a majority of non-producing strains.

Figure 8. Growth of S. eurocidicus in a bioreactor
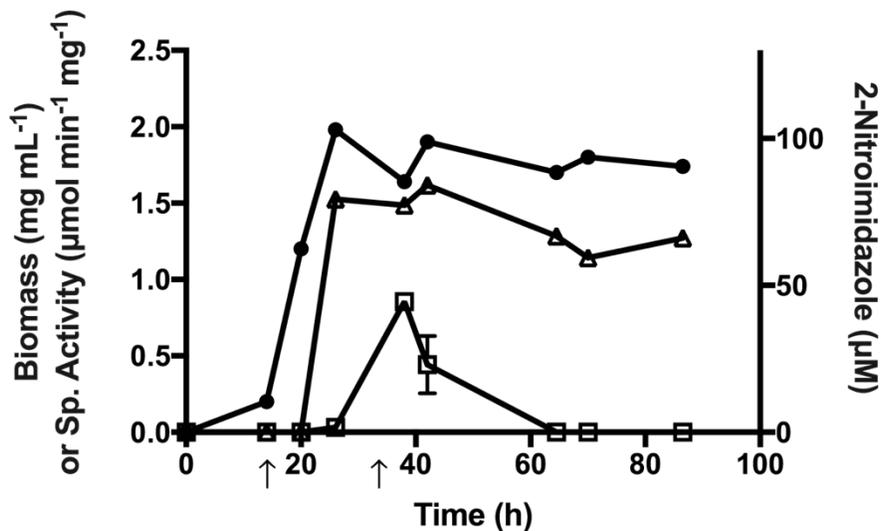
Figure 9. Growth curve of S. eurocidicus in a bioreactor

Dry biomass (circles) increased, followed by 2-AI oxidase activity (squares) and the accumulation of 2-NI (right axis) in stationary phase cells.
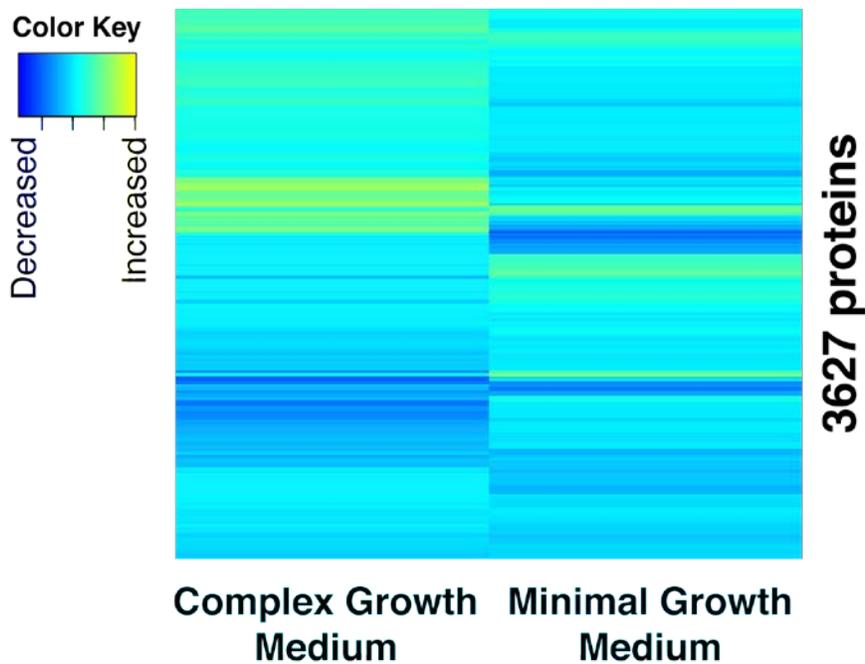
Figure 10. Heat map of changes in relative protein abundance

Proteins with increased abundance (yellow) correlate with 2-nitroimidazole production. Other proteins decreased in abundance (dark blue).

From this set of conserved overexpressed genes, an apparent 3.7-kbp operon with five genes (designated *aznABCDE* with GenBank identifiers AF335_07605 to AF335_07625) was selected for further investigation. All five proteins increased in relative abundance during growth on complex medium by an average of 5.8-fold. Nevertheless, ANOVA *p*-values exceeded the

significance target of ≤ 0.01 due to variability. These proteins also decreased in relative abundance during 2-NI production in minimal medium, and homologous genes were found in the genome of *S. vitaminophilus*, which does not produce 2-NI. The first gene in this cluster encodes a protein with 34% amino acid sequence identity to a recently discovered L-arginine α-deaminase, γ-hydroxylase involved in enduracididine biosynthesis (93). This enzyme catalyzes the formation of compound **2**, our first proposed reaction in 2-AI biosynthesis (Figure 11). With this ambiguous evidence, we launched experiments to express the gene cluster in a platform organism, *Escherichia coli*, characterize heterologous protein activity, and iteratively refine the proposed biosynthetic pathway that was originally proposed by Nakane *et al.* (Figure 3).
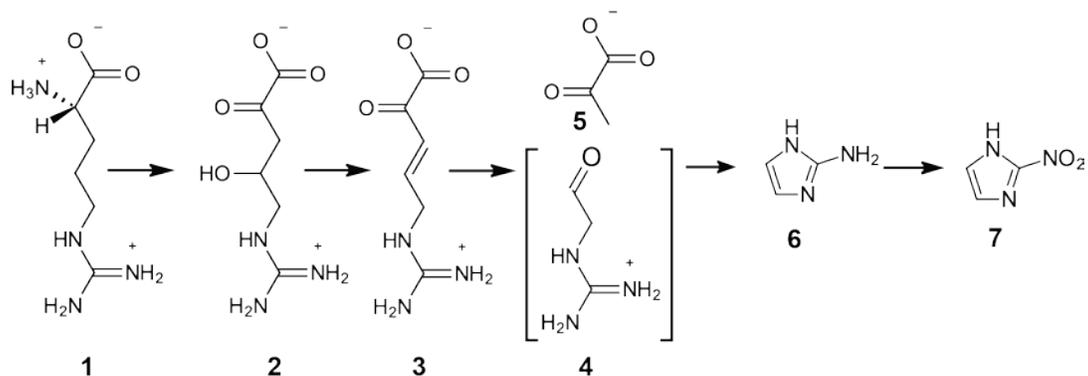
Figure 11. Revised biosynthesis of 2-nitroimidazole from L-arginine

Table 3. Intermediates and products formed by cells expressing parts of the *aznABCDE* pathway

Strain	Genes	Substrate	2-AI (µM)			Compounds detected by HRMS			
			Whole-cell biotrans-formations (yield)	Cell-free lysates	2	3	4	6	7
E. coli pBAD-HisA	None	Arg	-[1]	-	-	-	+	-	-
E. coli pDG714	*aznA*	Arg	BD		+	+	+	+	-
E. coli pDG726	*aznAB*	Arg	97 (3%)	184 ± 4.1	+	+	+	+	-
E. coli pDG728	*aznABC*	Arg	144 (5%)	511 ± 6.9	+	+	+	+	-
E. coli pDG712	*aznABCDE*	Arg	127 (4%)	533 ± 44.9	+	+	-	+	-
E. coli pDG728 + *E. coli* pDG716	*aznABC*, *aznDE*	Arg	ND	ND	+	+	+	+	+
E. coli pBAD-HisA	None	2AI + Glc	ND[2]	ND	-	-	+	+	-
E. coli pDG716	*aznDE*	2AI + Glc	ND	ND	-	-	+	+	+
E. coli pDG716	*aznDE*	2AI + Succ	ND	ND	-	-	-	+	+

[1] BD, below detection

[2] ND, not determined.

We expressed the *aznABCDE* cluster in *Escherichia coli*, from plasmid vector pDG712 with an arabinose-inducible promoter. An aerobic reaction mixture containing induced, washed cells converted 5% of the L-arginine substrate to 2-AI, detected by HPLC (Table 3). No 2-NI was detected in these reactions. LC-HRMS analysis confirmed the presence of 2-AI, as well as significant amounts of intermediates **2** and **3** and traces of intermediate **4** but not 2-NI. To evaluate activity in a heterologous host similar to *S. eurocidicus*, we expressed *aznABCDE* from a constitutive *ermE**p promoter in vector pDG734 that was integrated into the chromosomes of *Streptomyces lividans* TK24 and *Streptomyces noursei* JCM4701 (94). Both recombinant strains produced 2-AI from L-arginine in biotransformation reactions, but not 2-NI.

Fragments of the five-gene cluster were separately expressed in *E. coli* to distinguish the proteins' roles in 2-AI biosynthesis using biotransformation assays. Cells containing vector pDG714 expressed the *aznA* gene. Although no 2-AI was detected by HPLC analysis of reactions containing these cells incubated with L-arginine, LC-HRMS analysis identified

miniscule traces of 2-AI, with more substantial amounts of intermediate **2** and traces of intermediates **3** and **4** (Table 3). These results suggest the AznA protein catalyzes the same function as the homologous L-arginine α-deaminase, γ-hydroxylase enzyme. The AznB protein has no sequence or structural similarity to characterized proteins. *E. coli* cells expressing both *aznAB* genes from vector pDG726 produced similar amounts of intermediates **2, 3, 4** and 2-AI, detected by LC-HRMS. Finally, the AznC protein is homologous to the lysine biosynthesis enzyme 4-hydroxy-tetrahydrodipicolinate synthase. Cells expressing *aznABC* genes from vector pDG728 produced substantial amounts of 2-AI with 8.8% yield, demonstrating that these three genes are sufficient for 2-AI biosynthesis (Table 3 and Figure 12). The relative abundances of intermediates **2** and **3** decreased by 72% and 57% in strains expressing *aznABC* compared to *aznA* alone, reflecting increased transformation to 2-AI. There was no trace of intermediates **2, 3** or 2-AI in reactions containing control cells with empty pBAD-HisA vector.
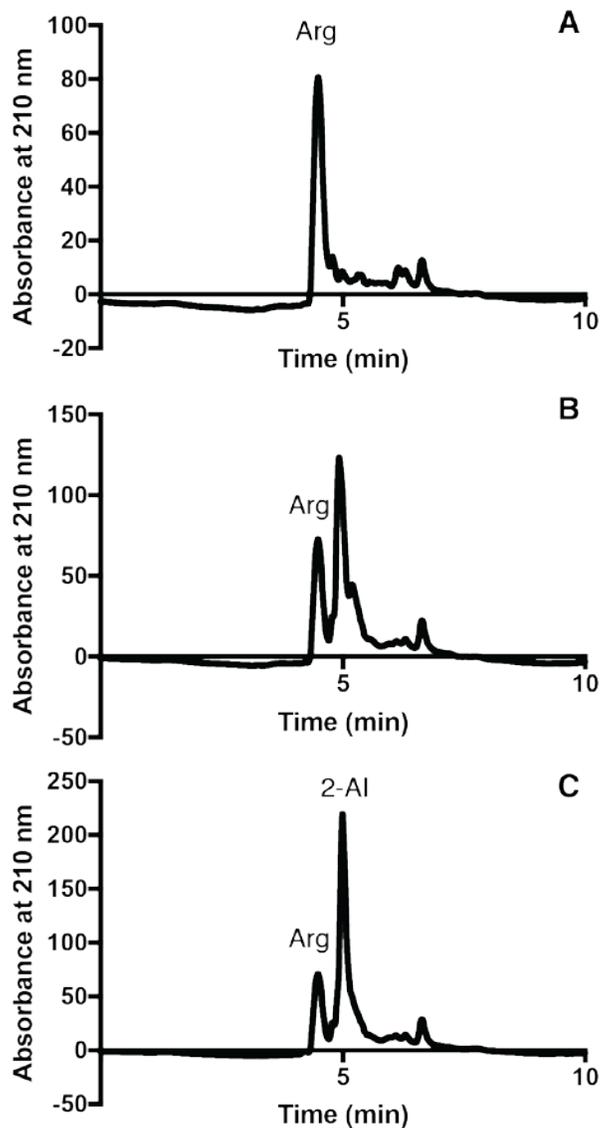
Figure 12. The aznABC genes are sufficient for 2-AI biosynthesis

Part A: Cells containing control vector do not transform L-arginine (Arg). Part B: Cells expressing aznA convert some Arg into a new product consistent with compound 2. Part C: Cells expressing aznABC convert Arg to 2-AI.

The intermediates 4-hydroxyarginine and compound **4** proposed by Nakane *et al.* (Figure 3) may not contribute significantly to 2-AI biosynthesis. Ultra-trace amounts of a compound with the exact mass of 4-hydroxyarginine were identified in all reactions containing arginine at similar low abundances, including the controls. Therefore, 4-hydroxyarginine is not a likely pathway intermediate. The compound 4-guanidinobutyraldehyde (**4**) may be a side-product rather than a free intermediate. It was detected at similarly low levels in most reaction mixtures. These results suggest that the AznA protein catalyzes the deamination and hydroxylation of L-arginine to form compound **2.** The function of AznB is unclear, but it could catalyze the reversible dehydration of **2** to form **3.** Finally, AznC could catalyze the retro-aldol condensation of **3** and cyclization of the

covalently bound intermediate **4** to form 2-AI (**6**) (Figure 13). We could not detect the pyruvate product in any reaction. It is probably decomposed by the cells, and it is poorly ionized by positive ion mode MS.
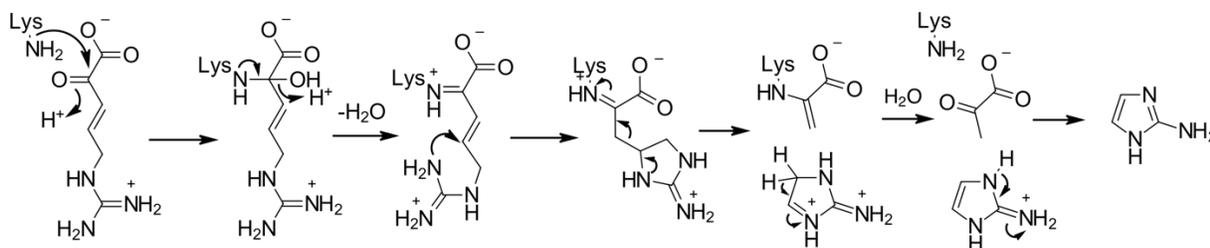
Figure 13. Proposed cyclization activity of AznC

Compound 3 forms a Schiff base with a lysine side chain of AznC. The intermediate cyclizes and undergoes a β-elimination to form 2-aminoimidazole and pyruvate.

Heterologously expressed AznA protein was purified and tested *in vitro* to confirm its role in 2-NI biosynthesis. AznA with an amino-terminal decahistidine tag (His_{10}-AznA) was purified by affinity chromatography, and 22 μM enzyme was incubated with pyridoxal 5'-phosphate cofactor and 3 mM L-arginine substrate. HPLC analysis indicated that 45% of the arginine was converted into a new compound that eluted in a later peak. No 2-AI was detected. No conversion was observed when D-arginine, L-homoarginine, L-lysine or L-glutamine was tested as substrate, but a new peak appeared when L-argininamide was provided. These results demonstrate this enzyme's stereochemical specificity and substrate side-chain requirement.

The *aznDE* genes encode hypothetical proteins from the heme oxidase and Rieske iron-sulfur protein families, which could require different growth conditions to produce activated enzymes in their heterologous host. Amino-terminal tagged His_{10}-AznD protein was heterologously expressed and purified by nickel affinity chromatography. Another *E. coli* strain expressed AznE protein with its predicted native start site. These cells were lysed by sonication and then treated with 2-mercaptoethanol, iron(II) ammonium sulfate and sodium sulfide under argon to reconstitute iron-sulfur centers (95). Diluted, clarified lysate from the reconstitution reaction was mixed with 3 μM AznD protein, 4 mM dithiothreitol (DTT) and 2 mM 2-AI and incubated for 14 h under oxic conditions. HPLC analysis of the reaction mixture detected 1.3 μM 2-NI (0.07% yield), consistent with single turnover of the AznD or AznE proteins (Figure 14). No significant 2-NI peak was identified in reactions without AznD, AznE, 2-AI or DTT. These results demonstrated that the AznD and AznE proteins together oxidized 2-AI, although an unidentified cofactor may be required for catalytic activity.
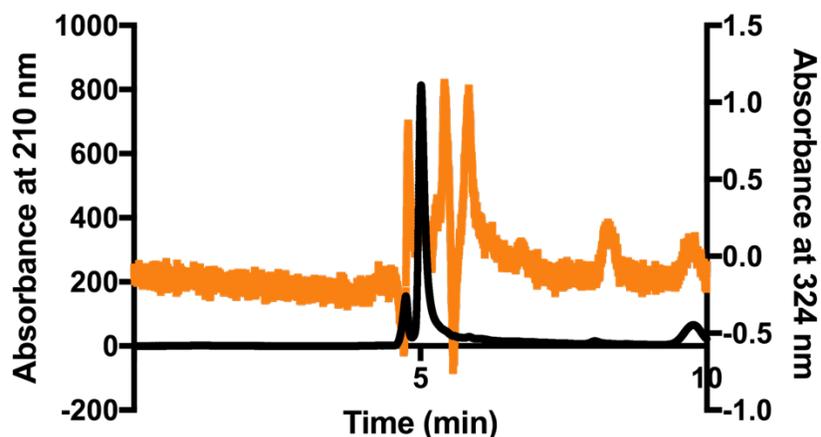
Figure 14. Activity of purified AznD protein

Reconstituted protein mixed with reconstituted extract containing AznE converted 2-AI to 2-NI (observed by absorbance at 324 nm, orange trace, right axis).

We expressed the *aznDE* genes together in *E. coli* from vector pDG716 in the presence of L-cysteine and iron(II) ammonium sulfate. In aerobic biotransformation reactions containing 670 μM 2-AI and an electron source (6.6 mM D-glucose or succinate), these cells catalyzed the formation of 6.0 ± 1.7 μM 2-NI (n=6, 0.9% yield). Cells oxidized 10% of the succinate to fumarate, and these cells produced 62% more 2-NI than glucose-treated cells. LC-HRMS analysis confirmed the identity of hydroxylaminoimidazole (**7**) and the 2-NI product, but did not identify a nitrosoimidazole intermediate (Table 3). *E. coli* control cells containing the empty pBAD-HisA vector did not catalyze this transformation. A mixture of cells expressing *aznABC* genes with cells expressing *aznDE* genes (as described above) catalyzed the complete transformation of L-arginine to 2-AI (297 μM, 6.2% yield) and 2-NI (7.3 μM, 0.15% yield) (Figure 15). The heterologous expression results confirmed that AznD and AznE proteins catalyze 2-AI oxidation *in vivo* in the presence of an electron source.

26
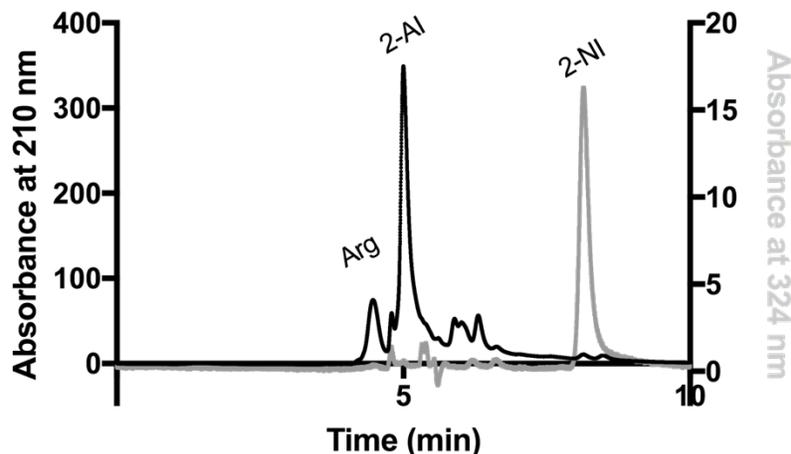
Figure 15. Complete biosynthesis of 2-NI

A mixture of cells expressing aznABC genes with cells expressing aznDE genes catalyzed the conversion of L-arginine to 2-AI and the oxidation of 2-AI to form 2-nitroimidazole (detected at 324 nm, grey trace, right axis).

Reactions containing [guanidino-$^{15}N_2$]-L-arginine and a mixture of cells expressing *aznABC* and *aznDE* produced labeled intermediates confirming the pathway shown in Figure 11. Compounds **1, 2, 3, 6**, **7**, and **8** containing two ^{15}N atoms were all identified using LC-HRMS with extracted ion chromatograms. The respective molecular ions [MH$^+$], their expected masses, observed *m/z* peak, and calculated mass error were L-arginine (**1**), expected 177.1130, observed 177.1130 *m/z*, 0 ppm; compound **2**, expected 192.0763, observed 192.0762 *m/z*, -0.5 ppm; compound **3**, expected 174.0657, observed 174.0658 *m/z*, +0.2 ppm; 2-AI (**6**), expected 86.04969, observed 86.0503 *m/z,* +7.1 ppm; (**7**), expected 102.04461, observed 102.0450 *m/z*, +3.8 ppm, and 2-NI (**8**), expected 174.0657, observed 174.0658 *m/z*, +2.0 ppm.

Discussion

This project successfully identified five proteins required for the conversion of L-arginine to 2-NI. We applied comparative genomic analysis, systems biology with semi-quantitative proteomic analysis, gene expression and protein biochemistry, and analytical biochemistry methods to discover a new series of biochemical reactions and revise the previously proposed biosynthetic pathway. By evaluating multiple lines of evidence, we were able to identify the gene candidates despite inconsistencies in gene distribution and differential protein abundance.

Other projects have constructed gene libraries of DNA from bacteria and transformed libraries into an expression host to screen or select for activity or natural product formation. For example, we used this approach to identify the novel *N*-nitroglycine lyase enzyme from *Variovorax* sp. JS1663 (88). A fosmid library was constructed from this bacterium's genomic DNA, and four fosmid clones transformed into an *E. coli* strain expressed the new lyase activity. We subsequently constructed 4 subclones to test the function of specific genes within the overlapping region of the fosmid clones. In this case, we used a simple colorimetric screen (Griess reaction) for nitrite formation that was amenable to high-throughput screening methods. The system required the heterologous expression of only one 560-bp gene, and the *Variovorax*

codon usage and native gene structure was sufficiently similar to *E. coli* to permit low-level expression without optimization.

Compared to this conventional screening approach, our systems biology strategy proved efficient to target 2-nitroimidazole biosynthesis. This pathway required three genes (2.5 kbp) to produce 2-AI and five genes (3.7 kbp) for the complete production of 2-NI. To ensure a 99% probability that this *S. eurocidicus* DNA would be incorporated into a fosmid with 30 kbp DNA insert size would require 1215 clones. The fastest specific assay for 2-AI and 2-NI used an HPLC method that required 15 minutes per sample; therefore, several weeks of uninterrupted HPLC analysis would have been necessary for screening. Importantly, there was no guarantee that the genes would be expressed in *E. coli* or that the products would be secreted at a sufficient concentration for detection. In fact, when we cloned the 3.7 kbp DNA fragment containing *aznABCDE* genes in *E. coli* pDG712 with a strong promoter, this strain produced only 2-AI (Table 3). Similar results were observed expressing the cluster in different *Streptomyces* strains. Only when we grew cells expressing *aznDE* under microaerophilic conditions supplemented with nutrients to promote iron-sulfur cluster formation and supplied an electron donor substrate did we observe 2-AI oxidase activity. While combinatorial screening remains an important tool for natural product identification, it is most effective for targeting simple systems with few interacting genes and facile assays.

Biosynthesis of *N*-nitroglycine
To investigate the biosynthesis of *N*-nitroglycine (NNG) by *Streptomyces noursei* JCM 4701, we combined several experimental approaches. The genomes of several NNG-producing strains were sequenced to identify conserved gene candidates for nitration enzymes. Stable isotope incorporation experiments were developed to determine the origins of carbon and nitrogen atoms in NNG. Finally, semi-quantitative proteomics was applied to identify proteins whose abundance correlated with NNG production activity in the cells. Analytical HPLC and HR-MS methods were developed to detect NNG in the supernatants of *S. noursei* cultures, using the chemically synthesized compound as a reference standard.

A hybrid assembly strategy combined short-read Illumina and long-read PacBio DNA sequences to produce a permanent draft genome sequence of *S. noursei* JCM 4701. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LJSN00000000. The final assembly with 8 scaffolds contains 10,123,402 bp. It has a G+C content of 71.3% an N50 of 3,830,602 bp, and an L50 of 2 scaffolds. These scaffolds include 79 contigs containing 10,116,605 bp, with an N50 of 231,639 bp and an L50 of 9 contigs. From a search for *Streptomyces* lineage-specific marker genes using the CheckM program, this genome sequence has 99.8% completeness, less than 2% contamination (due to paralogous genes) and no strain heterogeneity. Primary methylation modifications identified by PacBio sequencing included CTCGm^{6}AG, CGGm^{6}ATCT, and GCGm^{6}AGGNC. A consensus of several genome annotation tools identified 9,468 coding DNA sequences. Profile hidden Markov model searches identified 69 putative cytochrome P450 enzymes, 3 flavin-dependent monooxygenases, 19 flavin-containing amino oxygenases, 10 peroxidases, and > 250 FAD-binding oxidoreductases, which could catalyze bionitration chemistry. The genome also contains a complete SV1-type prophage and complete gene clusters for the biosynthesis of aureothin and nystatin. A genome

sequence for the related strain *S. noursei* ATCC 11455 was subsequently deposited in GenBank with accession CP011533 (Sergei Zotchev, personal communication).

S. vitaminophilus cells produced substantial amounts of NNG in media containing L-arginine or amino acid hydrolysates (Figure 16). NNG was produced near the end of the exponential growth phase, detected by HPLC or LC-HRMS (Figure 17). Samples of culture medium removed at suitable intervals were filtered and analyzed using HPLC. After 100 h of growth, the cells produced more than 150 µM NNG in nutrient-rich medium. A subsequent experiment conducted for more than 500 h revealed that NNG concentrations did not decrease, indicating that it is a stable metabolite that does not change over time.
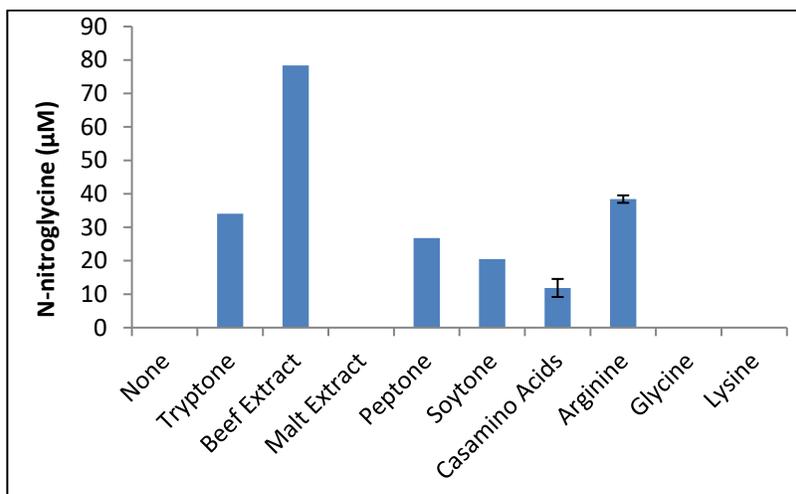
Figure 16. NNG production in various growth media

S. noursei cells produced the highest titers of NNG in media supplemented with organic nutrients.
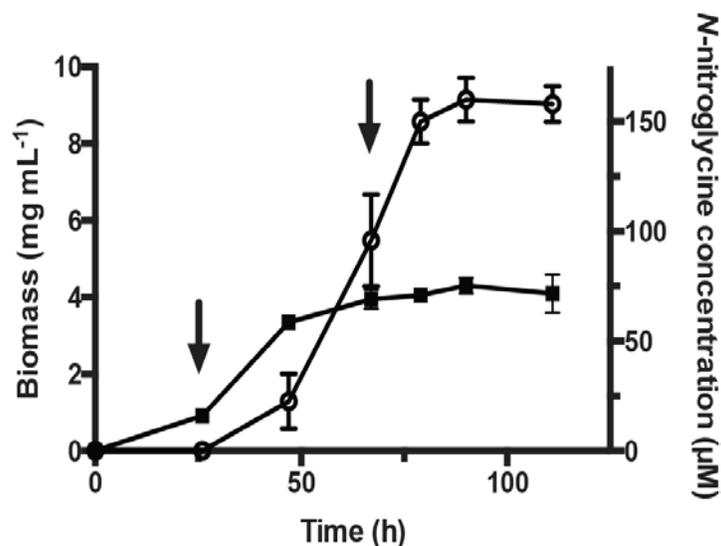
Figure 17. Growth curve of S. noursei cells with NNG production

N-nitroglycine accumulated in culture medium during the late growth stage of S. noursei cells. Black squares indicate cell biomass as dry weight on the left axis. Open circles indicate NNG concentrations in culture fluids on the right axis. Data points represent the mean and error bars show the range of values for duplicate samples. Arrows indicate time points when samples were removed for proteomic analysis.

We screened closely related *Streptomyces* spp. for NNG production to assess the phylogenetic distribution of biosynthetic genes. Related strains were identified using multilocus sequence analysis (D. Labeda, personal communication) and (83). A selection of these strains was grown on six different media and analyzed for NNG using high-resolution LC-MS. Both *S. noursei* strains JCM 4701 and ATCC 11455 (or ISP-5126) produced NNG, in addition to the closely related strains *Streptomyces albulus* NRRL B-5386 and *Streptomyces yunnanensis* NRRL B-24306 (K. Ju, personal communication). Draft genome sequences available for several of the strains were used to constrain the potential set of bionitration enzymes. We assume that NNG-producing strains contain orthologs of the requisite bionitration enzyme, which may not be conserved in related, non-producing strains.

Potential precursors of NNG labeled with stable isotopes were fed to *S. noursei* JCM 4701 cells harvested at the end of their exponential growth phase. All samples contained more unlabeled NNG than labeled product, due to production from residual endogenous precursors. Adding unlabeled arginine to the cells did not cause an increase in NNG production after 12 h relative to control experiments. After 24 h, arginine-fed cells produced up to 30% more NNG than control experiments, measured by the relative peak areas of extracted ion chromatograms. Therefore, 24 h incubations with labeled precursors were required for significant isotope incorporation. Label incorporation was identified by LC-HRMS analysis. Isotopically labeled NNG was not identified after incubation with 4-amino[*phenyl*-$^{13}C_6$]benzoate, [$^{13}C_6$] 4-aminobenzoic acid, L-[*amide*-^{15}N]asparagine, L-[*amide*-^{15}N]glutamine, K$^{15}NO_3$, or $^{15}NH_4Cl$ (data not shown). Incubation with [$^{13}C_6$] 4-aminobenzoic acid produced isotopically labeled *p*-aminobenzamide but no labeled NNG.

On the other hand, incubations of washed cells with L-[*guanidino*-$^{15}N_2$]arginine, [$^{13}C_2$]glycine, or [2-^{13}C,^{15}N]glycine all resulted in accumulation of isotopically labeled NNG (Experiments 1-3 in

Table 4). Incubations combining L-[*guanidino*-^{15}N$_2$]arginine with either [^{13}C$_2$] or [2-^{13}C,^{15}N]glycine in actively growing cells resulted in triple-labeled NNG (Experiments 4-5 in Table 4 and Figure 18). Peaks corresponding to other permutations of labeled NNG were also detected, including singly- or doubly-labeled variants, indicating that labeled substrates combined with unlabeled endogenous precursors to produce partially labeled NNG (Table 1). The peak corresponding to [^{13}C$_1$]- NNG is probably a natural isotope peak from the unlabeled NNG rather than mixing of the label in glycine (96). The result showing triple label incorporation indicates that glycine is the backbone precursor to NNG and the amine moiety, while one of the terminal guanidinium N atoms of arginine is incorporated into the nitro moiety of the nitramine (Figure 18). Isotope incorporation was not detected after shorter incubation times, which may be due to the large endogenous pool of precursors or limited uptake of the labeled precursors.

Table 4. Incorporation of isotopically labeled precursors into *N*-nitroglycine

Expt	Isotopically Labeled Substrates[1]	Fully Labeled NNG Product	Unlabeled NNG[2]	[13C1]-NNG[3]	[13C2]-NNG	[15N1]-NNG	[13C2,15N1]-NNG	[13C1,15N1]-NNG	[13C1,15N2]-NNG
1	glycine (^{13}C labeled carboxyl/CH$_2$)	NNG (^{13}C$_2$)	▓		▓				
2	glycine (^{13}C CH$_2$, ^{15}N NH$_2$)	NNG (^{13}C, ^{15}N)	▓					▓	
3	arginine (^{15}N labeled guanidino)	NNG (^{15}N)	▓			▓			
4	arginine (^{15}N) + glycine (^{13}C)	NNG (^{13}C$_2$, ^{15}N)	▓	▓	▓	▓	▓		
5	arginine (^{15}N) + glycine (^{13}C, ^{15}N)	NNG (^{13}C, ^{15}N)	▓	▓		▓		▓	▓

[1] Atoms substituted with ^{13}C are colored blue, while ^{15}N substituted atoms are colored red in arginine and green in glycine. Incorporation of isotopes is indicated by shaded boxes.

[2] Unlabeled NNG was produced in all samples from endogenous substrates. Adding unlabeled substrates did not result in any peak for any isotopically labeled NNG.

[3] [^{13}C$_1$]-*N*-nitrogycine was detected in samples containing high concentrations of NNG and was attributable to the natural isotope peak of unlabeled NNG.
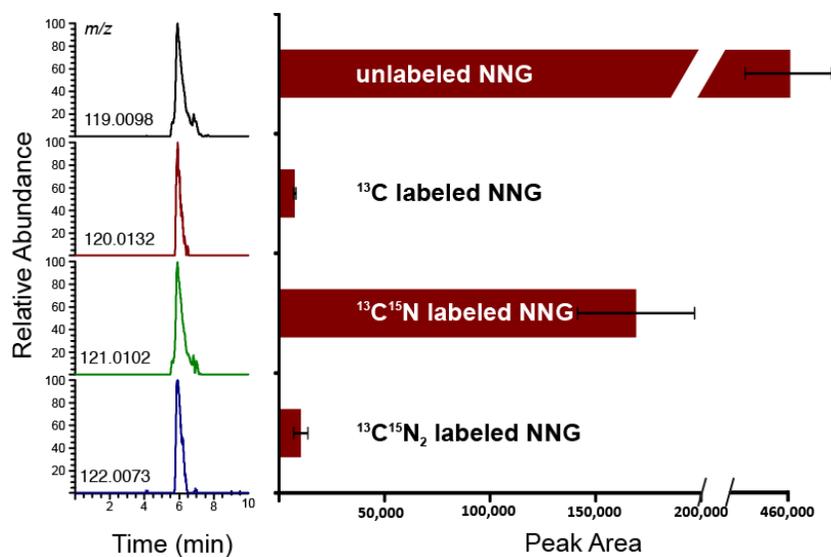
Figure 18. Incorporation of isotopic label into NNG.

Labeled NNG was produced from [2-^{13}C,^{15}N]glycine and L-[guanidino-^{15}N$_2$]arginine isotope incorporation into NNG during growth of S. noursei, measured by high resolution LC-MS. Extracted ion chromatograms (target m/z ± 0.0006 [5 ppm]; left) were integrated to show the abundances of molecular ions corresponding to isotopically substituted NNG. The peak at m/z = 120.0132 probably corresponds to the natural isotope peak of unlabeled NNG.

In order to gain insight into the protein determinants of NNG synthesis, the *S. noursei* proteome was analyzed both before and during the production phase, describe above. Peptide spectrum matches (PSM) were filtered to maintain a peptide-level FDR ≤ 1% per sample. This resulted in the identification of 3,478 *S. noursei* proteins (≥2 peptides per protein), with an average of 2,630 (RSD = 13%) identified per sample with protein-level FDRs ranging from 1.9 to 3.3%. For peptide and protein quantification, only peptides uniquely assigned to specific proteins were considered. In total, 37,239 peptides mapping to 2,794 proteins were quantifiable. As shown in Figure S1, peptide abundance is log-normally distributed and generally consistent across all samples, even before normalization and standardization (Figure S2). From a global protein abundance perspective, samples from the 26-h time point are substantially different from the 67 h time point along the time axis of the principal component analysis (PCA; Figure S3). This difference is maintained both before and after missing value imputation. Pearson correlation analysis corroborates the PCA plots whereby differences between the time points (71% average correlation, RSD = 2%) far surpass differences observed between biological replicates (94% average correlation, RSD = 1%). These data indicate that large protein abundance differences are observed between time points, thus providing ample targets for further inquiry with regard to NNG production by *S. noursei*.

ANOVA was performed to identify proteins exhibiting significant differential abundance at the two sample points from among the 2,794 quantifiable proteins. Considering only proteins with a p-value ≤ 0.01 and at least a two-fold change in abundance (≥ 1 or ≤ -1 in log$_2$ space), 975 proteins exhibited differential abundance. Of these differentially abundant proteins, 598 showed significantly increased abundance at 67 h, the point at which *S. noursei* actively synthesizes NNG. Hierarchical cluster (HC) analysis of differentially expressed proteins, as depicted by the heat map in Figure 19 formed two main clusters of proteins (rows) whose abundance is either increased or decreased across time points. The HC analysis also showed appropriate clustering of

sample replicates (columns) into their respective time point groups. Many of these changes in protein abundance could be due to the complex *Streptomyces* developmental cycle. However, only a handful of these proteins correspond to *Streptomyces coelicolor* proteins that were identified as differentially abundant at successive growth phases (97).
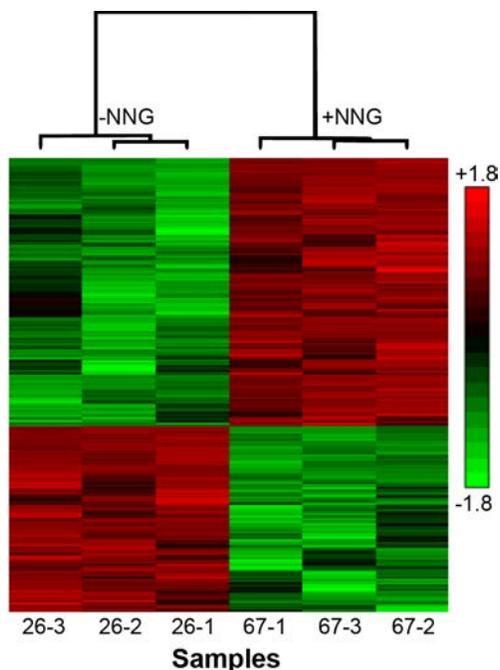
Figure 19. Heat map illustrates S. noursei proteins whose abundance changed significantly between 26 and 67 h samples

Relative abundances with imputed values were log$_2$ transformed and range from low (green) to high (red). Triplicate samples of cells grown for 26 or 67 h show high reproducibility.

Multiple lines of evidence are required to prioritize gene candidates for NNG biosynthesis. Among 2788 confidently identified proteins from the JCM 4701strain, 2102 (75%) are encoded in genomes of all four *Streptomyces* spp. that produce NNG. Of the 598 proteins described above that significantly increased in relative abundance at 67 h, 418 (70%) are conserved in all the NNG-producing strains. Only 67 proteins in this group are not found in six closely related *Streptomyces* spp. that do not produce NNG. Although this final criterion may be too restrictive, this limited gene set includes numerous hypothetical proteins and several homologs of putative oxidases enzymes that could be involved in NNG biosynthesis.

Gene clusters were analyzed in the *S. noursei* JCM 4701 genome to distinguish genes involved in the biosynthesis of known secondary metabolites from genes that could encode novel NNG biosynthetic proteins. The AntiSMASH 3.0 server (80) predicted 41 biosynthetic clusters in this genome, including genes for nonribosomal peptide synthases, type I, II, and III polyketide synthases, siderophores, terpenes, lantipeptides, bacteriocins, linaridins, ectoine, and butyrolactone. Also predicted were antibiotic gene clusters similar to albonoursin, cyclohexamide, aureothin, and nystatin (98). Nystatin biosynthesis has been studied in detail using *S. noursei* ATCC 11455 (99). The program also predicts a cluster for 4-hydroxy-3-nitrosobenzamide biosynthesis (100). Feeding *S. noursei* JCM 4701 labeled *p*-aminobenzoate

(*p*ABA) resulted in the production of labeled *p*-aminobenzamide (data not shown). Proteins responsible for the synthesis may be part of the 4-hydroxy-3-nitrosobenzamide cluster. Three benzamide synthase-like proteins were 30-fold more abundant during the late growth stages of *S. noursei*. *S. noursei* JCM 4701 also contains homologs of all the genes required for aureothin biosynthesis. We have not confirmed the production of aureothin by *S. noursei* JCM 4701 but all the proteins were on average 7-fold more abundant during the late stages of growth. In *S. thioluteus*, a *p*-nitrobenzoic acid (*p*NBA) intermediate in aureothin production is synthesized from the oxidation of *p*ABA (90), where the amino group attached to a benzene ring is oxidized by AurF. Feeding studies performed using ring-labeled *p*ABA did not result in the production of labeled or unlabeled *p*NBA in *S. noursei.*

Discussion

Among the handful of *Streptomyces* spp. that produce *N*-nitroglycine, *S. noursei* JCM 4701 produces this nitramine at the highest titer. The label incorporation results indicate that glycine is the backbone precursor to NNG, while the terminal guanidinium N atoms of arginine provide the nitrogen of the nitro group. During 24 h incubations, ^{15}N was not incorporated into NNG from asparagine, glutamine, nitrate or ammonium. *S. noursei* produced asparaginase and glutaminase enzymes, which were 2- and 10-fold less abundant during the late growth stage when NNG was produced (Table S2). Proteomic analysis also detected enzymes for the *de novo* biosynthesis of glycine and arginine. The relative abundance of arginine biosynthetic enzymes decreased by an average of 5-fold during the late growth stage, while enzymes to degrade arginine increased by 44-fold. These results are consistent with a shift in metabolic activity from primary to secondary metabolism over 67 h growth, and they support the role of L-arginine as the nitrogen donor for NNG nitro group biosynthesis. While we cannot exclude the possibility that a degradation product of arginine such as urea or citrulline provides the nitrogen, the direct oxidation of L-arginine is the simplest explanation. Several alternative mechanisms could explain how the arginine is oxidized, releasing a reactive nitrogen species that is used to nitrate glycine to produce NNG under physiological conditions. These possible mechanisms are discussed below.

The observed nitrogen incorporation from the terminal guanidinium nitrogen atoms of arginine suggests the activity of a nitric oxide synthase (NOS) enzyme (Figure 6A) (101-103). In thaxtomin biosynthesis, a truncated NOS (TxtD) produces nitric oxide (•NO) from L-arginine for the nitration of a tryptophanyl group (101). However, bacterial or eukaryotic *nos* homologs were not identified in *S. noursei.* Either NOS is not required to mediate this nitration reaction or the organism has a novel NOS protein. Cytochrome P450 and NOS enzymes share similar heme cofactors, thiolate ligands, and reaction intermediates. P450s activate molecular oxygen to catalyze a vast array of oxidative transformations including *N*-hydroxylations and *N*-oxidations but normally do not contribute directly to •NO formation (104, 105). Nevertheless, microsomal P450 enzymes can catalyze the second half-reaction of NOS, the oxidation of an *N*-hydroxyguanidine to produce a mixture of nitrogen oxides (106). Twenty-nine P450s were identified in the *S. noursei* proteome, and half of those were relatively more abundant during the late growth stage, with a median 12-fold increase. Therefore, it is possible that this organism has evolved a new NOS activity.

Alternatively, the cells could use a non-heme enzyme to catalyze arginine oxidation. Sequences similar to PrnD (*N*-oxygenase involved in aminopyrrolnitrin biosynthesis (95)) and AurF (*N*-oxygenase involved in aureothin biosynthesis (22)) were evident in the genome of *S. noursei* JCM 4701, but they are not present in *S. noursei* ATCC 11455 that also produce NNG. Thus the PrnD and AurF homologs are not candidate enzymes. *S. noursei* also contains homologs of genes required for cremeomycin biosynthesis. In this pathway, a flavin monooxygenase CreE oxidizes L-aspartate to nitrosuccinate, from which lyase CreD releases nitrous acid (107). Although not previously seen, a similar flavin monooxygenase could oxidize the terminal guanidine nitrogen atom, and the resulting N^ω-hydroxylarginine or N^ω-nitroarginine could then be cleaved to release nitrous acid. However, the substrate nitrogen atom presumably must be deprotonated before oxidation, and the guanidinium moiety is more difficult to deprotonate than the amino group of aspartate due to delocalization of the positive charge.

Nitric oxide can also be produced from nitrite by nitrite reductase enzymes, as well as metalloproteins including xanthine oxidase/xanthine dehydrogenase (102, 108, 109), aldehyde oxidase (108), sulfite oxidase, and others. *S. noursei* cells produced assimilatory nitrate and nitrite reductase enzymes at 425- and 12-fold higher abundance at the late growth stage. The *S. noursei* genome contains homologs of xanthine oxidases/dehydrogenases and aldehyde oxidases, which were not identified by proteomic analysis. Aerobic feeding experiments using $^{15}NO_3^-$ did not result in the incorporation of the isotope into NNG. Therefore, the cells are unlikely to use intermediates produced from nitrate or nitrite as nitration agents for NNG biosynthesis.

If a NOS-like enzyme produces ·NO from L-arginine in *S. noursei*, then a reactive nitrogen species could add to the glycine backbone, producing NNG. Several direct nitration reactions have been proposed to use the radical species ·NO or nitrogen dioxide (·NO$_2$) in biological systems. In these reactions, the substrate is converted to a stable radical by the abstraction of an electron before the addition of a radical nitrogen species leading to nitration. This may be the primary mode of 3-nitrotyrosine formation *in vivo* (36). A similar mechanism has been proposed for the biosynthesis of thaxtomin. Biochemical studies showed that a P450 oxygenase (TxtE) catalyzes the direct nitration of L-tryptophan using ·NO to form L-4-nitrotryptophan (33). *S. noursei* does not have a closely related homolog of *txtE*; however, it encodes numerous distantly related P450 enzymes. Similar chemistry was used to describe an adventitious nitration reaction catalyzed by an iron-dependent halogenase (SyrB2) using nitrite (110). No homolog of *syrB2* was identified in *S. noursei*. If ·NO reacts directly with a glycine radical cation, it could produce an *N*-nitrosoglycine intermediate that could be oxidized to form NNG; alternatively reaction with ·NO$_2$ could produce NNG.

Nitric oxide can be oxidized to form more reactive nitrogen species. For the bionitration of a primary amine, the cells could use a nitrosonium cation (NO$^+$) to form an *N*-nitrosoglycine intermediate that could be oxidized to form NNG (Figure 6B). Also, an electrophilic or charge-transfer reaction using nitronium ion (NO$_2^+$) could produce NNG directly from glycine. The NO$^+$ species reacts rapidly in water, producing NO$_2^+$, which also reacts with water to produce NO$_3^-$ at neutral pH (111). Therefore, enzymes would need to closely couple the production of these reactive species with electrophilic nitration to efficiently catalyze selective bionitration and avoid non-specific nitration reactions. This mechanism may be associated with peroxynitrite-mediated nitration of tyrosine to form 3-nitrotyrosine in the presence of transition metals (112). NO$_2^+$ has

also been proposed as an intermediate in the diazotization of an arylamine for cremeomycin biosynthesis discussed above (107).

A biosynthetic cluster responsible for the production of NNG is not obvious in the AntiSMASH predictions, and it is not known how many proteins are required to produce this nitramine. Comparative genome analysis and differential proteomic analysis constrained the set of candidate proteins for NNG biosynthesis to 5% of the predicted CDS in *S. noursei*. After excluding proteins within this set that were also present in *Streptomyces* strains that do not produce NNG, numerous candidate proteins remain. Our initial attempts to disrupt genes in the JCM 4701 strain did not create stable insertion mutations in the targeted genes. Future studies can build on this work to investigate the functions of the putative oxidases that could catalyze novel biochemistry of nitramine formation and expand our biosynthetic toolkit.

Biosynthesis of nitrophenols

The *Salegentibacter* sp. T436 strain was uniquely reported to produce nitrophenolic compounds. Our strategy to characterize the mechanism of this bionitration included confirming previously published results, reducing the complexity of culture media to enable reproducible growth experiments, determining the source of nitrogen for bionitration, sequencing the genomes of several *Salgentibacter* spp. for comparative genomic analysis, and applying transcriptomic analysis to identify genes differentially expressed during bionitration activity.

Salegentibacter sp. T436 cells were grown on complex media reported previously to stimulate the production of nitroaromatic compounds. Growth rates and cell densities were comparable in these media. We have developed HPLC methods to detect these molecules and have identified several new peaks that appear after 120 hours incubation and are unique to the supernatant of these cultures. High-resolution LC-MS/MS analyses of these cultures (both filtered supernatant and a supernatant extract) have identified more than seven compounds exhibiting fragmentation patterns with characteristic neutral losses of either HNO_2 (loss of 47.0007 Da) or NO_2 (45.9929 Da). These molecules were unique to samples containing the bacteria (compared to media controls) and were identified by relatively conservative metrics (ppm < 10 and neutral loss peak intensity > 5% of the intact nitro-containing molecule).

Alternative growth media were evaluated for the cultivation of *Salegentibacter* sp. T436 to reduce the complexity of the B1 medium used previously (56) and identify nutrients that contribute to nitro compound production. Nitrate was required for efficient production of nitroaromatics. The cells also required a marine salts basal medium for growth. Omitting N-Z amine, soy meal, or yeast extract each reduced nitro compound concentrations by 50-70%. Omitting seaweed extract had little effect. Based on these results, we omitted soy meal and seaweed extract and reduced yeast extract from media for most experiments, without substantially affecting growth rate or final cell density (Figure 20). Removing undefined ingredients reduced the complexity of the medium and confirmed that the T436 bacteria were responsible for producing the nitrated compounds, probably from aromatic amino acid precursors.
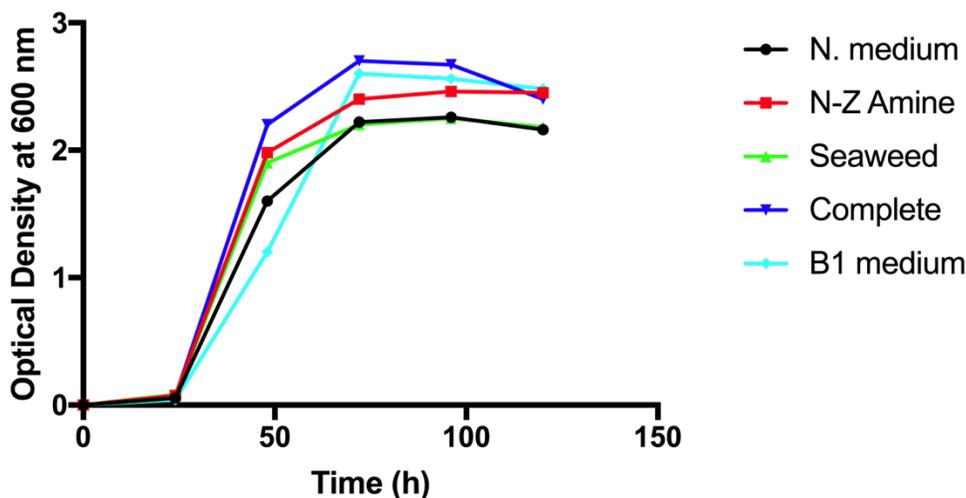
Figure 20. Growth of Salegentibacter sp. T436 with different nutrient supplements.

Basal nitrogen medium (N. medium, black) was supplemented with N-Z amines (red), seaweed extract (green) or a combination of N-Z amine, seaweed extract and beef extract (complete, blue). For reference, a growth curve of cells in B1 medium (cyan) is shown.

Ethyl acetate efficiently extracted nitrophenolic products from the growth media of *Salegentibacter* sp. T436 cultures, together with other colored products. When the organic extract was evaporated to dryness, a concentrated oil or brown precipitate formed. The resuspended extract was separated by thin layer chromatography (TLC) to resolve several major UV-absorbing, yellow-colored peaks (Figure 21). Compounds eluted from the TLC plate were characterized by negative ion mode LC-HRMS at the University of Tennessee Biological and Small Molecule Mass Spectrometry Core facility (Figure 22). The most abundant nitrated products were 4-nitrophenol, 4-nitrobenzoic acid, 4-hydroxy-3-nitrophenyl propionic acid, 4-hydroxy-3-nitrophenyl acetic acid, 4-hydroxy-3-nitrobenzoic acid, and 3-nitrotyrosine.
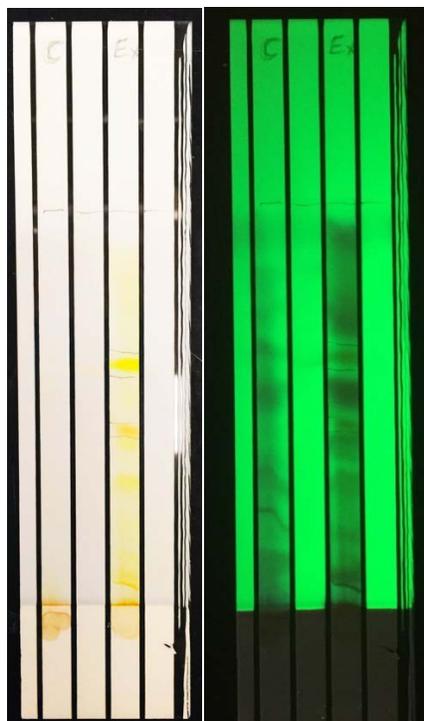
Figure 21. Thin layer chromatography separation of T436 extract

A comparison of extracts from uninoculated growth medium (lane C) and T436 culture supernantant (lane Ex) shows several yellow peaks (left) that absorb UV light (right).
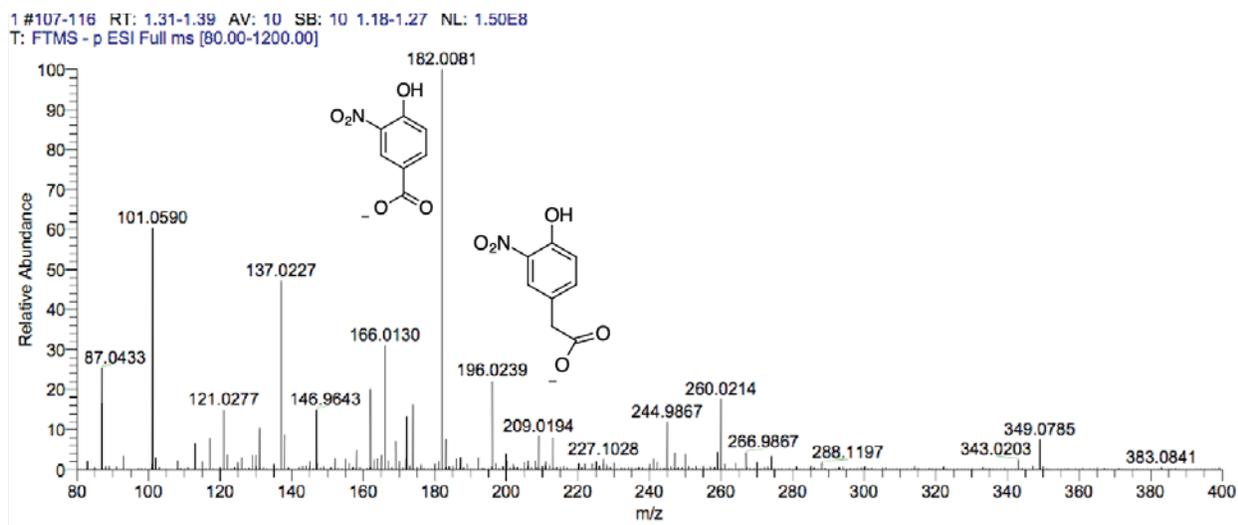
Figure 22. Mass spectrum showing two nitrophenolic compounds from T436 cultures

Nitrophenolic compounds including 4-hydroxy-3-nitrobenzoate and 4-hydroxy-3-nitrophenylacetate (shown) were identified in culture supernatants.

Growth experiments were performed to determine whether cells incorporated nitrate from the medium into nitrophenolic products. T436 cells were grown in a bioreactor using standard medium supplemented with isotopically labeled $Na^{15}NO_3$. The dissolved oxygen concentration began to decrease rapidly after 20 hours of growth, when nitrite began to accumulate in the supernatant. A supernatant sample was collected after 24 hours, and the extract was analyzed by LC-HRMS (Table 5). Nitrophenolic compounds were detected by UV absorbance and negative ion Fourier transform mass spectrometry. Both unlabeled (^{14}N) and labeled (^{15}N) compounds were detected in extracted ion chromatograms with a mass error less than 5 ppm. The relative abundance of labeled compounds ranged from 8 to 60% of the unlabeled compound. 3-Nitrotyrosine was also detected, although ^{15}N was not significantly incorporated. Nitrotyrosine, hydroxynitrophenylacetic acid, hydroxynitrophenylpropionic acid, and hydroxynitrophenyl-benzoic acid were all detected in supernatant at concentrations from 10-40 μM after 54 hours, using HPLC with UV absorbance. These results demonstrated that the cells incorporate nitrate into nitrophenolic products.

Table 5. Nitro compounds detected in supernatants of *Salegentibacter* sp. T436 cells grown on complex medium with added $Na^{15}NO_3$

Compound	Unlabeled			^{15}N-Labeled		
	$[M-H^+]^-$ Predicted (*m/z*)	Observed (*m/z*)	Mass error (ppm)	$[M-H^+]^-$ Predicted (*m/z*)	Observed (*m/z*)	Mass error (ppm)
4-Nitrophenol	138.0202	138.0196	-4.3	139.0173	139.0167	-4.3
4-Hydroxy-3-nitrobenzoic acid	182.0100	182.0095	-2.7	183.0071	183.0064	-3.8
4-Hydroxy-3-nitrophenylacetic acid	196.0257	196.0251	-3.1	197.0227	197.0221	-3.0
4-Hydroxy-3-nitrophenylpropionic acid	210.0413	210.0407	-3.1	211.0384	211.0378	-2.8
3-Nitrotyrosine	225.0522	225.0535	+5.8	ND[1]	ND	ND

[1] ND, Not detected.

Phenolic compounds were added to cultures as substrates to test the cells' ability to nitrate exogenous compounds. Aromatic compounds (200 μM) were added to actively growing cultures of T436 cells (optical density at 600 nm = 0.3) and incubated for 25 h to assess bionitration activity. Products were extracted in ethyl acetate and analyzed by HPLC with detection by UV absorbance. Assignments were made relative to commercial standards. The five phenolic compounds shown in Table 6 were nitrated at a position ortho to the hydroxyl substituent on the aromatic ring according to the general reaction shown in Figure 23. Although traces of each nitrophenol were detected in incubations without added substrates, concentrations increased at least 10-fold in cultures with added substrates. In contrast, benzoic acid was not a substrate for bionitration, demonstrating specificity for phenolic substrates. Based on these results, we conclude that the exogenous phenolic substrates were nitrated by the cells.

Table 6. Bionitration of phenolic compounds added to *Salegentibacter* sp. T436 cultures

Substrate	Product	Conversion
Anisole	2-Nitroanisole	0.3%
Tyrosine	3-Nitrotyrosine	0.3%
4-Hydroxyphenylacetate	4-Hydroxy-3-nitrophenylacetate	4%
4-Hydroxybenzoate	4-Hydroxy-3-nitrobenzoate	12%
Phenol	4-Nitrophenol	12%
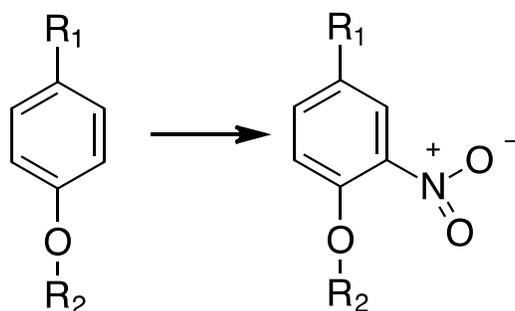
Figure 23. General reaction for bionitration of phenolic compounds by Salengentibacter sp. T436

The full genome sequence of *Salegentibacter* sp. T436 and a permanent draft sequence of *Salegentibacter mishustinae* were completed, each comprising approximately four million base pairs (Table 7). The two strains share 85% average nucleotide identity (ANI). 2674 orthologous proteins share 90% average amino acid identity. The T436 strain reduced nitrate to nitrite and produces nitrophenols. In contrast, *S. mishustinae* cells did not efficiently reduce nitrate and did not product nitrophenols. The T436 genome encodes pathways for the reduction of nitrate and nitrite, consistent with the observed reduction of nitrate to nitrite. These genes are missing in the genome of *S. mishustinae*. No known nitration proteins were predicted in the T436 genome; therefore, the mechanism of nitration will be novel.

Table 7. *Salegentibacter* spp. genome sequence assembly

Organism	*Salegentibacter* sp. T436	*S. mishustinae* DSM 23404
GenBank Accession No.	CP012872	LLKN00000000
Sequencing method	Illumina TruSeq paired-end and PacBio	Illumina TruSeq paired-end and PacBio
Mean depth of coverage (fold)	240 (Illumina)	426 (Illumina)
Total length (bp)	4,086,550	3,792,722
# contigs	1	5
GC (%)	37	37
CDS	3792	3523

Transcriptomic analysis was performed twice to identify T436 genes that were differentially expressed during nitrate reduction, and the production of nitrite and nitrophenols. Cells were grown in a 2-L bioreactor using standard medium (Figure 24). After ~24 hours growth, the

dissolved oxygen (DO) concentration decreased, accompanied by an increase in nitrite formation, followed by the release of nitrophenol (NP) and hydroxynitrobenzoic acid (HNBA). Cells were sampled and harvested at time points from 2 to 69 hours for RNAseq analysis.
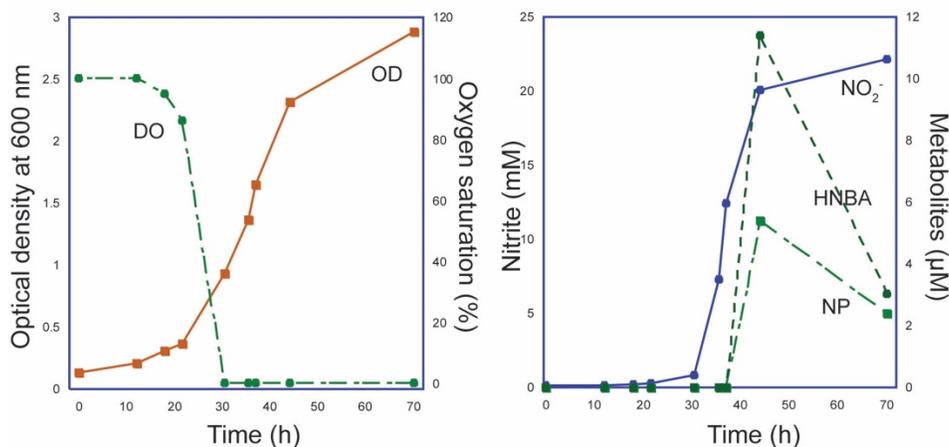
Figure 24. Growth of T436 cells for RNAseq experiments

Cell biomass increased (indicated by optical density (OD), while dissolved oxygen (DO) decreased during growth (Left). The products nitrite, HNBA and NP accumulated during later growth stages.

In the first RNAseq experiment, 3450 genes were expressed, representing 91% of the CDS. Only 51 genes showed greater than twofold increases in relative expression between the initial and final time points. Fifteen of those genes do not have apparent orthologs in the *S. mishustinae* genome, which does not produce nitrite or nitrophenols. One cluster of contiguous genes on the chromosome, containing five of these genes, encodes a putative nitrate/nitrite major facilitator superfamily transporter and nitrate reductase enzyme subunits, in addition to molybdopterin cofactor assembly proteins. This gene cluster is not found in *S. mishustinae*. A second cluster including two differentially expressed genes encodes genes involved in NAD synthesis, and lies adjacent to genes encoding a nitrite reductase. None of these are conserved in *S. mishustinae*. Finally, five of the most over-expressed genes encode subunits of cytochrome c oxidase, which can reduce nitrite to form nitric oxide (·NO) in addition to its usual role in respiration (113). These genes are conserved in *S. mishustinae*.

The second RNAseq experiment identified 3613 expressed genes. RNAseq data from the *Salegentibacter* sp. T436 growth experiment were analyzed using differential expression analysis, principal components analysis and hierarchical clustering (J. Whitham, ORNL). Genes that were overexpressed in late growth stages coincident with nitrite and nitrophenol production were found in two hierarchical clusters (designated 1.2.2 and 2.2.2 in Figure 25). These overexpressed genes included those identified in the first experiment, providing a consistent signature.
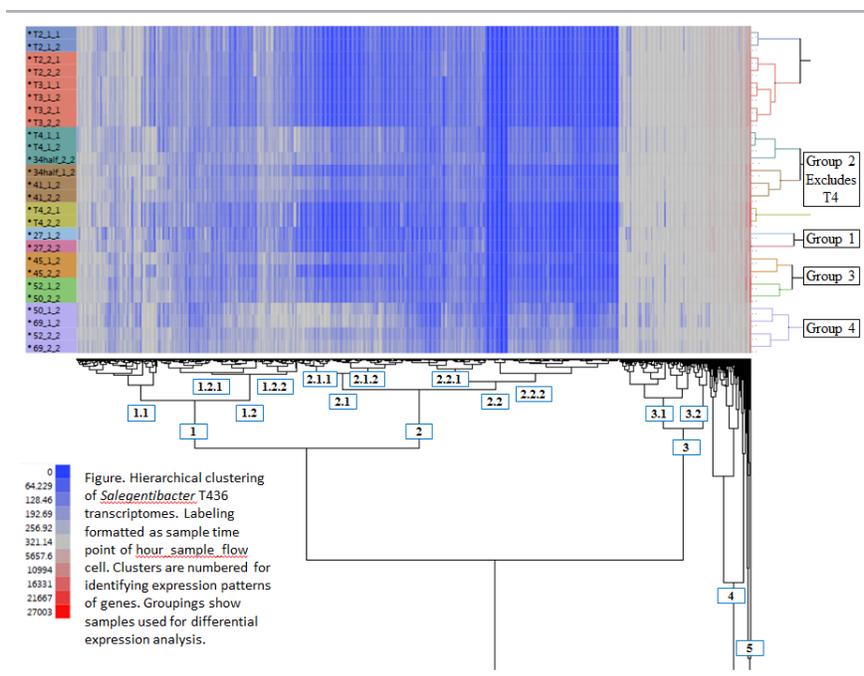
Figure 25. Clustered gene expression data from an RNAseq experiment

Graph by J. Whitham (ORNL)

Discussion

Our studies comparing the bionitrating strain *Salegentibacter* sp. T436 with the non-nitrating strain *S. mishustinae* suggest that T436 has evolved a broad specificity bionitration mechanism. T436 cells apparently grow aerobically on carbohydrates and amino acids. Tyrosine degradation produces the major phenolic precursors for the nitrophenols. The T436 genome encodes canonical pathways for both tyrosine biosynthesis and degradation (homogentisate pathway) (114, 115). Yet additional aromatic degradation pathways and enzymes remain to be discovered (116, 117).

When oxygen becomes limiting in the cultures, the cells reduce nitrate forming significant amounts of nitrite. Our isotopic labeling experiments demonstrated that nitrogen from nitrate was incorporated into nitrophenol products. Nitrous acid can directly nitrate phenol and hydroxyphenylacetic acid in aqueous solutions at low pH (118-120). It was also proposed to be the precursor the diazotization precursor for cremeomycin biosynthesis by *Streptomyces cremeus* (107). However, the pKa for nitrous acid (~3.4) is much lower than typical cellular conditions, and its tendency to decompose to a mixture of NO_2 and $\cdot NO$ suggests nitrous acid may not be the direct substrate for bionitration.

Cells can produce other reactive nitrogen oxide species from nitrite (121). The cytochrome c oxidase enzyme reduces nitrite to $\cdot NO$ under hypoxic conditions (113). Nitrous oxide often nitrates activated, substituted phenols at the free ortho position adjacent to the hydroxyl group, consistent with the T436 products (122). Reaction of $\cdot NO$ with superoxide ($O_2^{\cdot-}$) produces peroxynitrite ($ONOO^-$), which reacts faster with phenols and tyrosine (forming 3-nitrotyrosine)

(123, 124). One of the T436 substrates, 4-hydroxyphenylacetic acid, was reported to be nitrated by reactive nitrogen species in rats following direct injection (125). Although additional experiments are required to confirm this mechanism and identify the nature of the reactive nitrogen oxide species that reacts with phenols, this system shows promise. A biocatalyst with the broad specificity could be most useful for the green chemistry bionitration of various synthetic or biosynthetic precursors to energetic materials.

Biosynthesis of pyrrolomycin

The natural precursors for pyrrolomycin nitration are complex aromatic compounds (Figure 6) that cannot be readily purchased, isolated, or synthesized. Therefore, we tested structurally similar compounds as potential substrates for bionitration by *Streptomyces vitaminophilus. S. vitaminophilum* (formerly *Actinosporangium vitaminophilum*) cells were grown under conditions known to produce pyrrolomycins. Small aromatic compounds were added to the cultures to evaluate them as surrogate substrates for the nitration reaction. Only the addition of hydroquinone led to the appearance of new compounds. However, this new, yellow compound does not appear to be a nitro compound.

We sequenced the genome of *Streptomyces vitaminophilus* ATCC 31673, developing a permanent draft sequence that included 6,549,812 bp with a G+C content of 71.9%. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LLZU00000000. The assembled genome comprises 39 contigs with an N50 of 249,406 bp and an L50 of 8 contigs. Coding DNA sequences (CDS) were identified and annotated by the NCBI Prokaryotic Genome Annotation Pipeline. The genome was predicted to contain 5941 CDS, 56 tRNA, two 16S rRNA, two 23S rRNA, and four 5S rRNA genes.

A 227-kbp contig includes the entire 56 kbp pyrrolomycin biosynthetic gene cluster previously deposited to GenBank (Accession EF140901.1) (1). There are no recognizable homologs of nitric oxide synthases or previously characterized *N*-oxygenases. However, the genome encodes 15 cytochrome P450 homologs, two assimilatory nitrate reductases and one nitrite reductase, which could be involved in bionitration.

Previous studies have shown that nitric oxide synthase inhibitors do not adversely affect nitrated pyrrolomycin biosynthesis in *Streptomyces fumanus* suggesting a completely novel bionitration reaction (39). AntiSMASH 3.0 predicted 27 biosynthetic gene clusters including genes for nonribosomal peptide synthases, type I, II, and III polyketide synthases, pyrrolomycins, siderophores, terpenes, lantipeptides, and even a lassopeptide. *Streptomyces vitaminophilus* ATCC 31673 produces the pyrrolomycins A-D, of which pyrrolomycins A and B contain a nitro group.

Discussion

Few enzymes involved in nitro group formation have been identified. The availability of the *Streptomyces vitaminophilus* genome and pyrrolomycin biosynthetic gene cluster sequences will facilitate the future identification of bionitration enzymes and add to the knowledge about nitro group formation. Due to the complexity of the pyrrolomycin biosynthetic pathway, the lack of suitable organic substrates for assaying this nitration reaction, and the genetic intractability of

this bacterium we decided to discontinue this task in order to focus resources on the more promising tasks described above.

CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH/IMPLEMENTATION

This project has identified new pathways and enzymes for the biosynthesis of nitro compounds, substantially expanding the range of known nitration enzymology. Future research and development efforts can extend our discoveries of 2-nitroimidazole biosynthetic apparatus to characterize the enzymes' specificities and engineer new functionality to produce a diversified product mix. The 2-aminoimidazole oxidase enzyme may require an additional cofactor that could significantly enhance activity in heterologous hosts. These heterologous expression systems should be tested in larger scale batch or continuous reactors, where enhanced process control can be used to optimize activity. Finally, applications for 2-aminoimidazole as a precursor to chemical synthesis of energetic materials should be explored.

The biosynthetic pathway for *N*-nitroglycine identified here and the set of candidate proteins for its production should be the basis for future genetic or biochemical analysis to identify responsible enzymes. The use of the newly discovered *N*-nitroglycine hydrolase enzyme could accelerate discovery by providing a rapid screen. Additionally, glycine analogs should be tested as precursors to determine the specificity of *S. noursei* enzymes for the future production of energetic compounds.

Our studies of *Salegentibacter* sp. T436 nitroaromatic compound biosynthesis strongly supports the hypothesis that this nitration mechanism has broad specificity. Additional phenolic compounds should be tested as substrates, and its nitrophenolic products should be evaluated as precursors for the synthesis of novel energetic materials. Heterologous expression of *Salegentibacter* genes whose expression correlates with nitration could resolve the mechanism of this remarkable nitration process. Future experiments with larger scale reactors and sophisticated process control and on-line measurements could significantly enhance bionitration efficiency.

LITERATURE CITED

1.   Zhang X, Parry RJ. 2007. Cloning and characterization of the pyrrolomycin biosynthetic gene clusters from *Actinosporangium vitaminophilum* ATCC 31673 and *Streptomyces* sp. Strain UC 11065. Antimicrob Agents Chemother 51:946-957.
2.   Olah GA, Malhotra R, Narang SC. 1989. Nitration : methods and mechanisms. VCH Publishers, New York, N.Y.
3.   U.S. Army. 1990. Military Explosives. TM 9-1300-214, Department of the Army, Washington, DC.
4.   Chavez DE. 2014. The Development of Environmentally Sustainable Manufacturing Technologies for Energetic Materials, p 235-258. *In* Brinck T (ed), Green Energetic Materials. John Wiley & Sons, New York.

5.   Badgujar DM, Talawar MB, Mahulikar PP. 2016. Review on Greener and Safer Synthesis of Nitro Compounds. Propellants, Explos Pyrotech 41:24-34.
6.   Millar RW, Arber AW, Hamid J, Endsor RM. 2007. Elimination of Redwater Formation from TNT Manufacture. SERDP Project WP-1408, SERDP, Arlington, VA.
7.   Katritzky AR, Scriven EFV, Majumder S, Akhmedov RG, Akhmedov NG, Vakulenko AV. 2005. Direct nitration of five membered heterocycles. ARKIVOC 2005:179-191.
8.   Parry R, Nishino S, Spain J. 2011. Naturally-occurring nitro compounds. Nat Prod Rep 28:152-167.
9.   Keasling JD. 2008. Synthetic biology for synthetic chemistry. ACS Chem Biol 3:64-76.
10.   Werpy T, Petersen G. 2004. Top value added chemicals from biomass: Volume 1-Results of screening for potential candidates from sugars and synthesis gas. Office of the Biomass Program, Department of Energy,
11.   Frost JW. 2010. Manufacture of TATB and TNT from biosynthesized phoroglucinols. WP-1582-FP, SERDP,
12.   Achkar J, Xian M, Zhao H, Frost JW. 2005. Biosynthesis of phloroglucinol. J Am Chem Soc 127:5332-5333.
13.   Niu W, Molefe MN, Frost JW. 2003. Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. J Am Chem Soc 125:12998-12999.
14.   Straessler NA. 2010. Synthesis of Trinitroaromatics Using Alternative Mixed Acid Nitration Conditions. Synthetic Commun 40:2513-2519.
15.   Winkler R, Hertweck C. 2007. Biosynthesis of nitro compounds. Chembiochem 8:973-977.
16.   Spain JC. 1995. Biodegradation of nitroaromatic compounds. Annu Rev Microbiol 49:523-555.
17.   Ju K-S, Parales RE. 2010. Nitroaromatic compounds, from synthesis to biodegradation. Microbiol Mol Biol Rev 74:250-272.
18.   Husserl J, Hughes JB, Spain JC. 2012. Key Enzymes Enabling the Growth of Arthrobacter sp. Strain JBH1 with Nitroglycerin as the Sole Source of Carbon and Nitrogen. Appl Environ Microbiol 78:3649-3655.
19.   Fida TT, Palamuru S, Pandey G, Spain JC. 2014. Aerobic biodegradation of 2,4-dinitroanisole by *Nocardioides* sp. strain JS1661. Appl Environ Microbiol 80:7725-7731.
20.   Lee J-K, Zhao H. 2007. Identification and characterization of the flavin:NADH reductase (PrnF) involved in a novel two-component arylamine oxygenase. J Bacteriol 189:8556-8563.
21.   Krebs C, Matthews ML, Jiang W, Bollinger JM. 2007. AurF from *Streptomyces thioluteus* and a Possible New Family of Manganese/Iron Oxygenases. Biochemistry 46:10413-10418.
22.   Zocher G, Winkler R, Hertweck C, Schulz GE. 2007. Structure and Action of the N-oxygenase AurF from *Streptomyces thioluteus*. J Mol Biol 373:65-74.
23.   Choi YS, Zhang H, Brunzelle JS, Nair SK, Zhao H. 2008. In vitro reconstitution and crystal structure of *p*-aminobenzoate *N*-oxygenase (AurF) involved in aureothin biosynthesis. Proc Natl Acad Sci USA 105:6858-6863.
24.   Li N, Korboukh VK, Krebs C, Bollinger JM. 2010. Four-electron oxidation of *p*-hydroxylaminobenzoate to *p*-nitrobenzoate by a peroxodiferric complex in AurF from *Streptomyces thioluteus*. Proc Natl Acad Sci USA 107:15722-15727.

25. Fries A, Bretschneider T, Winkler R, Hertweck C. 2011. A Ribonucleotide Reductase-Like Electron Transfer System in the Nitroaryl-Forming N-Oxygenase AurF. Chembiochem 12:1832-1835.

26. Winkler R, Richter MEA, Knüpfer U, Merten D, Hertweck C. 2006. Regio- and chemoselective enzymatic *N*-oxygenation in vivo, in vitro, and in flow. Angew Chem Int Ed Engl 45:8016-8018.

27. Fries A, Winkler R, Hertweck C. 2010. Structural and biochemical basis for the firm chemo- and regioselectivity of the nitro-forming N-oxygenase AurF. Chem Commun 46.

28. Knoot CJ, Kovaleva EG, Lipscomb JD. 2016. Crystal structure of CmlI, the arylamine oxygenase from the chloramphenicol biosynthetic pathway. JBIC Journal of Biological Inorganic Chemistry 21:589-603.

29. Parry RJ, Li W. 1997. An NADPH:FAD oxidoreductase from the valanimycin producer, *Streptomyces viridifaciens*. J Biol Chem 272:23303-23311.

30. Bruender NA, Thoden JB, Holden HM. 2010. X-ray structure of KijD3, a key enzyme involved in the biosynthesis of D-kijanose. Biochemistry 49:3517-3524.

31. Johnson EG, Krasnoff SB, Bignell DRD, Chung W-C, Tao T, Parry RJ, Loria R, Gibson DM. 2009. 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. Mol Microbiol 73:409-418.

32. Sudhamsu J, Crane BR. 2009. Bacterial nitric oxide synthases: what are they good for? Trends Microbiol 17:212-218.

33. Barry SM, Kers JA, Johnson EG, Song L, Aston PR, Patel B, Krasnoff SB, Crane BR, Gibson DM, Loria R, Challis GL. 2012. Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. Nat Chem Biol 8:814-816.

34. Yu F, Li M, Xu C, Wang Z, Zhou H, Yang M, Chen Y, Tang L, He J. 2013. Structural insights into the mechanism for recognizing substrate of the cytochrome P450 enzyme TxtE. PLoS ONE 8:e81526.

35. Souza J, Peluffo G, Radi R. 2008. Protein tyrosine nitration -Functional alteration or just a biomarker? Free Radic Biol Med 45:357-366.

36. Radi R. 2004. Nitric oxide, oxidants, and protein tyrosine nitration. Proc Natl Acad Sci USA 101:4003-4008.

37. Rubbo H, Radi R. 2008. Protein and lipid nitration: Role in redox signaling and injury. Biochim Biophys Acta 1780:1318-1324.

38. Carter GT, Nietsche JA, Goodman JJ, Torrey MJ, Dunne TS, Siegel MM, Borders DB. 1989. Direct biochemical nitration in the biosynthesis of dioxapyrrolomycin. A unique mechanism for the introduction of nitro groups in microbial products. J Chem Soc, Chem Commun:1271-1273.

39. Ratnayake AS, Haltli B, Feng X, Bernan VS, Singh MP, He H, Carter GT. 2008. Investigating the biosynthetic origin of the nitro group in pyrrolomycins. J Nat Prod 71:1923-1926.

40. Seki Y, Nakamura T, Okami Y. 1970. Accumulation of 2-aminoimidazole by *Streptomyces eurocidicus*. J Biochem 67:389-396.

41. Simpson RL, Coon CL, Foltz MF, Pagoria PF. 1995. A new insensitive explosive that has moderate performance and is low cost: 2,4-dinitroimidazole. UCRL-ID-119675, Lawrence Livermore National Laboratory,

42. Novikov SS, Khmel'nitskii LI, Lebedev OV, Sevast'yanova VV, Epishina LV. 1970. Nitration of imidazoles with various nitrating agents. Chem Heterocycl Compd (NY) 6:465-469.

43. Ravi P, Gore G, Tewari S, Sikder A. 2012. A DFT study of aminonitroimidazoles. J Mol Model 18:597-605.

44. Gao H, Shreeve JnM. 2011. Azole-based energetic salts. Chem Rev 111:7377-7436.

45. Singh RP, Verma RD, Meshri DT, Shreeve JnM. 2006. Energetic nitrogen-rich salts and ionic liquids. Angew Chem Int Ed Engl 45:3584-3601.

46. Smith MW, Cliff MD. 1999. NTO-Based Explosive Formulations: A Technology Review. Defence Science & Technology Organisation, Salisbury, Australia.

47. Qu Y, Spain JC. 2011. Catabolic pathway for 2-nitroimidazole involves a novel nitrohydrolase that also confers drug resistance. Environ Microbiol 13:1010-1017.

48. Lancini GC, Kluepfel D, Lazzari E, Sartori G. 1966. Origin of the nitro group of azomycin. Biochim Biophys Acta 130:37-44.

49. Nakane A, Nakamura T, Eguchi Y. 1977. A novel metabolic fate of arginine in *Streptomyces eurocidicus*. Partial resolution of the pathway and identification of an intermediate. J Biol Chem 252:5267-5273.

50. Miyazaki Y, Kono Y, Shimazu A, Takeuchi S, Yonehara H. 1968. Production of nitraminoacetic acid by *Streptomyces noursei* 8054-MC$_3$. J Antibiot (Tokyo) 21:279-282.

51. Hantzsch A, Metcalf WV. 1896. Ueber Nitraminessigsäure. Ber Dtsch Chem Ges 29:1680-1685.

52. Frankel MB, Klager K. April 4, 1961 1961. Nitramino esters. USA patent 2,978,495.

53. Dagley IJ, Kony M, Walker G. 1995. Properties and impact sensitiveness of cyclic nitramine explosives containing nitroguanidine groups. J Energ Mater 13:35-56.

54. Holstead C, Lamberton AH. 1952. 347. Nitramines and nitramides. Part III. The formation of nitric acid by the action of sulphuric acid. J Chem Soc.

55. Crocker F, Indest K, Fredrickson H. 2006. Biodegradation of the cyclic nitramine explosives RDX, HMX, and CL-20. Appl Microbiol Biotechnol 73:274-290.

56. Al-Zereini W, Schuhmann I, Laatsch H, Helmke E, Anke H. 2007. New aromatic nitro compounds from *Salegentibacter* sp. T436, an Arctic sea ice bacterium: taxonomy, fermentation, isolation and biological activities. J Antibiot (Tokyo) 60:301-308.

57. Schuhmann I, Yao CBF-F, Al-Zereini W, Anke H, Helmke E, Laatsch H. 2009. Nitro derivatives from the Arctic ice bacterium *Salegentibacter* sp. isolate T436. J Antibiot (Tokyo) 62:453-460.

58. Badgujar DM, Talawar MB, Mahulikar PP. 2015. Review on Greener and Safer Synthesis of Nitro Compounds. Propellants, Explos Pyrotech:n/a-n/a.

59. Ezaki N, Shomura T, Koyama M, Niwa T, Kojima M, Inouye S, Itō T, Niida T. 1981. New chlorinated nitro-pyrrole antibiotics, pyrrolomycin A and B (SF-2080 A and B). J Antibiot (Tokyo) 34:1363-1365.

60. Koyama M, Kodama Y, Tsuruoka T, Ezaki N, Niwa T, Inouye S. 1981. Structure and synthesis of pyrrolomycin A, a chlorinated nitro-pyrrole antibiotic. J Antibiot (Tokyo) 34:1569-1576.

61. Conder GA, Zielinski RJ, Johnson SS, Kuo MS, Cox DL, Marshall VP, Haber CL, DiRoma PJ, Nelson SJ, Conklin RD, Lee BL, Geary TG, Rothwell JT, Sangster NC. 1992. Anthelmintic activity of dioxapyrrolomycin. J Antibiot (Tokyo) 45:977-83.

62.   Bellamy AJ, Ward SJ, Golding P. 2002. A new synthetic route to 1,3,5-triamino-2,4,6-trinitrobenzene (TATB). Propellants, Explos Pyrotech 27:49-58.

63.   Ju K-S, Gao J, Doroghazi JR, Wang K-KA, Thibodeaux CJ, Li S, Metzger E, Fudala J, Su J, Zhang JK, Lee J, Cioni JP, Evans BS, Hirota R, Labeda DP, van der Donk WA, Metcalf WW. 2015. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. Proc Natl Acad Sci USA 112:12175-12180.

64.   Ezaki N, Shomura T, Koyama M, Niwa T, Kojima M, Inouye S, Ito T, Niida T. 1981. New chlorinated nitro-pyrrole antibiotics, pyrrolomycin A and B (SF-2080 A and B). J Antibiot (Tokyo) 34:1363-5.

65.   Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich.

66.   Phelan RM, Sekurova ON, Keasling JD, Zotchev SB. 2015. Engineering Terpene Biosynthesis in Streptomyces for Production of the Advanced Biofuel Precursor Bisabolene. ACS Synthetic Biology 4:393-399.

67.   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-477.

68.   English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA. 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS ONE 7:e47768.

69.   Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17:10-12.

70.   Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9:e112963.

71.   Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614-620.

72.   Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

73.   Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043-1055.

74.   Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass E, Kubal M, Meyer F, Olsen G, Olson R, Osterman A, Overbeek R, McNeil L, Paarmann D, Paczian T, Parrello B, Pusch G, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75.

75.   Hyatt D, Chen G-L, LoCascio P, Land M, Larimer F, Hauser L. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

76.   Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.

77.   Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783-795.

78.   Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567-580.

79.   Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res.

80.   Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237-W243.

81.   Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222-D230.

82.   Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126-19131.

83.   Labeda DP, Dunlap CA, Rong X, Huang Y, Doroghazi JR, Ju K-S, Metcalf WW. 2017. Phylogenetic relationships in the family *Streptomycetaceae* using multi-locus sequence analysis. Antonie van Leeuwenhoek 110:563-583.

84.   Giannone RJ, Wurch LL, Heimerl T, Martin S, Yang Z, Huber H, Rachel R, Hettich RL, Podar M. 2015. Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. ISME J 9:101-114.

85.   Tabb DL, Fernando CG, Chambers MC. 2007. MyriMatch: Highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome Res 6:654-661.

86.   Ma Z-Q, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, Tabb DL. 2009. IDPicker 2.0: Improved Protein Assembly with High Discrimination Peptide Identification Filtering. J Proteome Res 8:3872-3881.

87.   Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Meth 13:731-740.

88.   Mahan KM, Zheng H, Fida TT, Parry RJ, Graham DE, Spain JC. 2017. Iron-Dependent Enzyme Catalyzes the Initial Step in Biodegradation of *N*-Nitroglycine by *Variovorax* sp. Strain JS1663. Appl Environ Microbiol 83:e00457-17.

89.   Osato T, Ueda M, Fukuyama S, Yagishita K, Okami Y, Umezawa H. 1955. Production of tertiomycin (a new antibiotic substance), azomycin and eurocidin by *S. eurocidicus*. J Antibiot (Tokyo) 8:105-109.

90.   He J, Hertweck C. 2004. Biosynthetic Origin of the Rare Nitroaryl Moiety of the Polyketide Antibiotic Aureothin: Involvement of an Unprecedented N-Oxygenase. J Am Chem Soc 126:3694-3695.

91.   Kirner S, van Pée K-H. 1994. The Biosynthesis of Nitro Compounds: The Enzymatic Oxidation to Pyrrolnitrin of Its Amino-Substituted Precursor. Angew Chem Int Ed Engl 33:352-352.

92.   Hu Y, Al-Mestarihi A, Grimes CL, Kahne D, Bachmann BO. 2008. A Unifying Nitrososynthase Involved in Nitrosugar Biosynthesis. J Am Chem Soc 130:15756-15757.

93.   Han L, Schwabacher AW, Moran GR, Silvaggi NR. 2015. Streptomyces wadayamensis MppP Is a Pyridoxal 5′-Phosphate-Dependent l-Arginine α-Deaminase, γ-Hydroxylase in the Enduracididine Biosynthetic Pathway. Biochemistry 54:7029-7040.

94.   Sekurova ON, Brautaset T, Sletta H, Borgos SEF, Jakobsen ØM, Ellingsen TE, Strøm AR, Valla S, Zotchev SB. 2004. In vivo analysis of the regulatory genes in the nystatin biosynthetic gene cluster of *Streptomyces noursei* ATCC 11455 reveals their differential control over antibiotic biosynthesis. J Bacteriol 186:1345-1354.

95.   Lee J, Simurdiak M, Zhao H. 2005. Reconstitution and characterization of aminopyrrolnitrin oxygenase, a rieske *N*-oxygenase that catalyzes unusual arylamine oxidation. J Biol Chem 280:36719-36727.

96.   Kikuchi G, Motokawa Y, Yoshida T, Hiraga K. 2008. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. Proc Jpn Acad, Ser B 84:246-263.

97.   Manteca A, Sanchez J, Jung HR, Schwämmle V, Jensen ON. 2010. Quantitative proteomics analysis of *Streptomyces coelicolor* development demonstrates that onset of secondary metabolism coincides with hypha differentiation. Mol Cell Proteomics 9:1423-1436.

98.   Fjærvik E, Zotchev SB. 2005. Biosynthesis of the polyene macrolide antibiotic nystatin in *Streptomyces noursei*. Appl Microbiol Biotechnol 67:436-443.

99.   Fjærvik E, Zotchev S. 2005. Biosynthesis of the polyene macrolide antibiotic nystatin in *Streptomyces noursei*. Appl Microbiol Biotechnol 67:436-443.

100.   Noguchi A, Kitamura T, Onaka H, Horinouchi S, Ohnishi Y. 2010. A copper-containing oxidase catalyzes C-nitrosation in nitrosobenzamide biosynthesis. Nat Chem Biol 6:641-643.

101.   Kers JA, Wach MJ, Krasnoff SB, Widom J, Cameron KD, Bukhalid RA, Gibson DM, Crane BR, Loria R. 2004. Nitration of a peptide phytotoxin by bacterial nitric oxide synthase. Nature 429:79-82.

102.   Zhang Z, Naughton D, Winyard PG, Benjamin N, Blake DR, Symons MCR. 1998. Generation of Nitric Oxide by a Nitrite Reductase Activity of Xanthine Oxidase: A Potential Pathway for Nitric Oxide Formation in the Absence of Nitric Oxide Synthase Activity. Biochem Biophys Res Commun 249:767-772.

103.   Moncada S, Higgs A. 1993. The L-Arginine-Nitric Oxide Pathway. N Engl J Med 329:2002-2012.

104.   Gorren ACF, Mayer B. 2007. Nitric-oxide synthase: A cytochrome P450 family foster child. Biochim Biophys Acta 1770:432-445.

105.   Zhu Y, Silverman RB. 2008. Revisiting heme mechanisms. A perspective on the mechanisms of nitric oxide synthase (NOS), heme oxygenase (HO), and cytochrome P450s (CYP450s). Biochemistry 47:2231-2243.

106.   Mansuy D, Boucher J-L. 2002. Oxidation of *N*-hydroxyguanidines by cytochromes P450 and NO-synthases and formation of nitric oxide. Drug Metabolism Reviews 34:593-606.

107.   Sugai Y, Katsuyama Y, Ohnishi Y. 2015. A nitrous acid biosynthetic pathway for diazo group formation in bacteria. Nat Chem Biol 12:73-75.

108.   Maia LB, Pereira V, Mira L, Moura JJG. 2014. Nitrite Reductase Activity of Rat and Human Xanthine Oxidase, Xanthine Dehydrogenase, and Aldehyde Oxidase: Evaluation of Their Contribution to NO Formation in Vivo. Biochemistry 54:685-710.

109. Maia LB, Moura JJG. 2015. Nitrite reduction by molybdoenzymes: a new class of nitric oxide-forming nitrite reductases. J Biol Inorg Chem 20:403-433.
110. Matthews ML, Chang W-c, Layne AP, Miles LA, Krebs C, Bollinger Jr JM. 2014. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. Nat Chem Biol 10:209-215.
111. Hughes MN. 1999. Relationships between nitric oxide, nitroxyl ion, nitrosonium cation and peroxynitrite. Biochim Biophys Acta 1411:263-272.
112. Tsikas D, Duncan MW. 2014. Mass spectrometry and 3-nitrotyrosine: Strategies, controversies, and our current perspective. Mass Spectrom Rev 33:237-276.
113. Hematian S, Garcia-Bosch I, Karlin KD. 2015. Synthetic Heme/Copper Assemblies: Toward an Understanding of Cytochrome c Oxidase Interactions with Dioxygen and Nitrogen Oxides. Acc Chem Res 48:2462-2474.
114. Sparnins VL, Chapman PJ. 1976. Catabolism of L-tyrosine by the homoprotocatechuate pathway in gram-positive bacteria. J Bacteriol 127:362-366.
115. Arias-Barrau E, Olivera ER, Luengo JM, Fernández C, Galán B, García JL, Díaz E, Miñambres B. 2004. The Homogentisate Pathway: a Central Catabolic Pathway Involved in the Degradation of l-Phenylalanine, l-Tyrosine, and 3-Hydroxyphenylacetate in Pseudomonas putida. J Bacteriol 186:5062-5077.
116. Blakley ER. 1977. The catabolism of L-tyrosine by an *Arthrobacter* sp. Can J Microbiol 23:1128-39.
117. Pometto AL, Crawford DL. 1985. L-Phenylalanine and L-tyrosine catabolism by selected *Streptomyces* species. Appl Environ Microbiol 49:727-729.
118. Al-Obaidi U, Moodie RB. 1985. The nitrous acid-catalysed nitration of phenol. J Chem Soc Perk T 2:467-472.
119. Pannala AS, Mani AR, Rice-Evans CA, Moore KP. 2006. pH-dependent nitration of para-hydroxyphenylacetic acid in the stomach. Free Radic Biol Med 41:896-901.
120. Rubio MA, Lissi E, Olivera N, Reyes JL, López-Alarcon C. 2014. Reactions of p-Substituted Phenols with Nitrous Acid in Aqueous Solution. International Journal of Chemical Kinetics 46:143-150.
121. Maia LB, Moura JJG. 2014. How Biology Handles Nitrite. Chem Rev 114:5273-5357.
122. Yenes S, Messeguer A. 1999. A study of the reaction of different phenol substrates with nitric oxide and peroxynitrite. Tetrahedron 55:14111-14122.
123. Goldstein S, Merényi G. 2008. The Chemistry of Peroxynitrite: Implications for Biological Activity, p 49-61. *In* Robert KP (ed), Methods Enzymol, vol 436. Academic Press.
124. Daiber A, Mehl M, Ullrich V. 1998. New Aspects in the Reaction Mechanism of Phenol with Peroxynitrite: The Role of Phenoxy Radicals. Nitric Oxide 2:259-269.
125. Mani AR, Pannala AS, Orie NN, Ollosson R, Harry D, Rice-Evans CA, Moore KP. 2003. Nitration of endogenous para-hydroxyphenylacetic acid and the metabolism of nitrotyrosine. Biochem J 374:521-527.

APPENDIX A. SUPPORTING DATA

Chemicals. *N*-nitroglycine or 2-(nitroamino)-acetic acid was purchased from AKos Consulting & Solutions (AKOS006375173). White solid. The NMR and MS spectra were consistent with the structure. ^{1}H NMR (400 MHz, DMSO-d_6, δ): 13.06 (br, COOH, 1H), 12.25 (s, NH, 1H), 4.07 (d, *J*= 4 Hz, CH_2, 2H). ^{1}H NMR (400 MHz, D_2O + DSS, δ): 4.33 (s, CH_2, 2H). ^{13}C NMR (100 MHz, DMSO-d_6, δ): 169.1 (OC=O), 46.0 (CH_2). ^{13}C NMR (100 MHZ, D_2O + DSS, δ): 174 (OC=O), 49 (CH_2).

HRMS-ESI(-) (*m/z*): [M-H]$^-$ 119.0098 calcd for $C_2H_4N_2O_4$, 119.0093 found (+4.5 ppm). [M-HNO_2]$^-$ 72.0091 calcd for $C_2H_2NO_2$, found 72.0089 (+5.6 ppm). The NNG molecular ion [M-H]$^-$ underwent neutral loss of HNO_2 in collision induced dissociation (CID) to produce a fragment ion of dehydroglycine or 2-iminoacetate (*m/z* 72.0091 expected, 72.0089 observed, +2.8 ppm)

Table S1. Strains, sources and genome sequences

Species and Genus	Strain	Source	Nitro product(s)	GenBank Genome Sequence Accession No.
Escherichia coli	DH5α		N.A.	NA
Escherichia coli	LMG194	Invitrogen	N.A.	
Escherichia coli	Rosetta 2 (DE3)	Novagen	N.A.	
Escherichia coli	ET12567 (pUZ8002)	R. Parry	N.A.	
Salegentibacter sp	T436	Georg-August University	Nitrophenols	
Salegentibacter mishustinae	DSM	DSM	N.D.	LLKN00000000
Streptomyces albireticuli	B-24306	NRRL	2-NI	NSJV00000000
Streptomyces albulus	NRRL B-5386	NRRL	NNG	
Streptomyces blastmyceticus	B-5480	NRRL	N.D.	
Streptomyces chattanoogensis	ISP-5002	NRRL	N.D.	
Streptomyces chattanoogensis	B-2255	NRRL	N.D.	
Streptomyces eurocidicus	ATCC 27428	ATCC	2-NI	LGUI000000
Streptomyces eurocidicus	B-1677	NRRL	N.D.	

Streptomyces rimosus subsp. *Romosus*	ISP-5260	NRRL	N.D.	
Streptomyces noursei	JCM 4701	JCM	NNG	LJSN00000000
Streptomyces noursei	ATCC 11455	ATCC	NNG	CP011533
Streptomyces noursei	B-1714 (ATCC 11455)	NRRL	NNG	
Streptomyces noursei	B-2625 (ATCC 11455)	NRRL	NNG	
Streptomyces noursei	S-1565 (ATCC 11455)	NRRL	NNG	
Streptomyces noursei	ISP-5126 (ATCC 11455)	NRRL	NNG	
Streptomyces (Actinosporangium) vitaminophilum	ATCC 31673	ATCC	Pyrrolomycins A to D	LLZU00000000
Streptomyces yunnanensis	NRRL B-24306	NRRL	NNG	

1 N.A., Not applicable, N.D. Not detected.

The following raw sequence data was submitted to the NCBI Short Read Archive (SRA):
Streptomyces noursei JCM 4701 SRA accessions PRJNA295563
Streptomyces eurocidicus ATCC 27428 SRA accessions PRJNA291301 (has raw reads and references assembly)
Salegentibacter mishustinae DSM 23404 SRA accessions PRJNA297772

Table S2. 2-NI intermediates and compounds identified by LC-HRMS

Compound	Name	Molecular Formula	[MH$^+$] (*m/z*)		Mass measurement error
#			Expected	Observed	ppm
1	L-arginine (Arg)	$C_6H_{14}N_4O_2$	175.11895	175.11878[A]	-0.98
2	2-oxo-4-hydroxy-5-guanidinovalerate	$C_6H_{11}N_3O_4$	190.08223	190.08210	-0.68
3	5-guanidino-2-oxo-3-pentenoate	$C_6H_9N_3O_3$	172.07167	172.07171	+0.23
4	guanidinoacetaldehyde	$C_3H_7N_3O$	102.06619	102.06663	+4.3
5	pyruvate	$C_3H_4O_3$	89.02332	ND[B]	
6	2-aminoimidazole (2-AI)	$C_3H_5N_3$	84.05562	84.05611[A]	+5.8
7	2-hydroxylaminoimidazole	$C_3H_5N_3O$	100.05054	100.05086	+3.2
8	2-nitroimidazole (2-NI)	$C_3H_3N_3O_2$	114.02980	114.0301 [A]	+2.6
	4-hydroxyarginine	$C_6H_{11}N_3O_4$	191.11387	191.11370	-0.89

Table S3. Plasmid vectors

Vector	Description	Source or reference
pBAD-HisA	Expression vector	Invitrogen
pDG714	*aznA* from *S. eurocidicus* in pBAD-HisA	This work
pDG726	*aznAB* from *S. eurocidicus* in pBAD-HisA	This work
pDG728	*aznABC* from *S. eurocidicus* in pBAD-HisA	This work
pDG712	*aznABCDE* from *S. eurocidicus* in pBAD-HisA	This work
pDG716	*aznDE* from *S. eurocidicus* in pBAD-HisA	This work
pDG734	*aznABCDE in pSOK806 with ErmE* promoter*	This work
pDG742	*His10-AznA*	This work
pDG754	*aznE* (*E. coli* codon-optimized)	This work
pET-19b	*E. coli* expression vector	Novagen
pSOK806	*Streptomyces* spp. integration vector	(66)
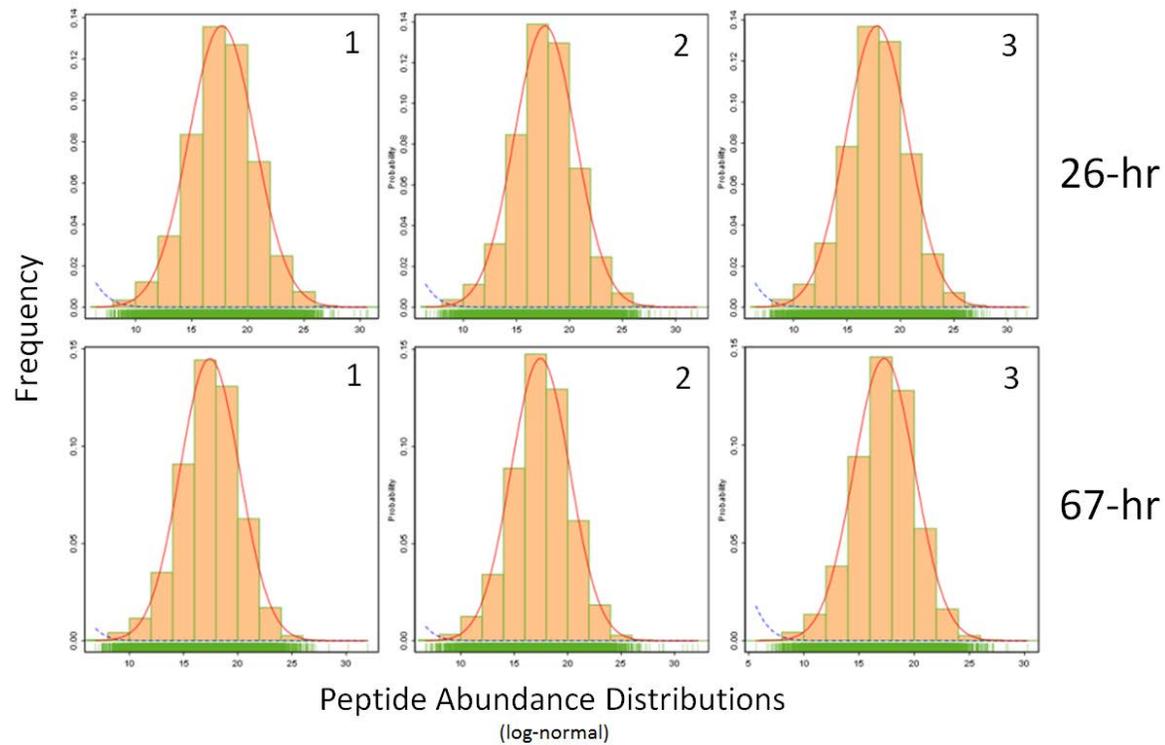
Figure S1. Histogram analysis of raw peptide abundances.

As assessed by MIT, histograms indicate that each LC-MS/MS run produced data that is generally (i) log-normally distributed (red line = reference to the log-normal distribution) and (ii) similar across samples. Distributions of this type allow for parametric statistical testing, such as ANOVA and/or Student's T-test, to assess quantitative differences amongst proteins identified in each time point.
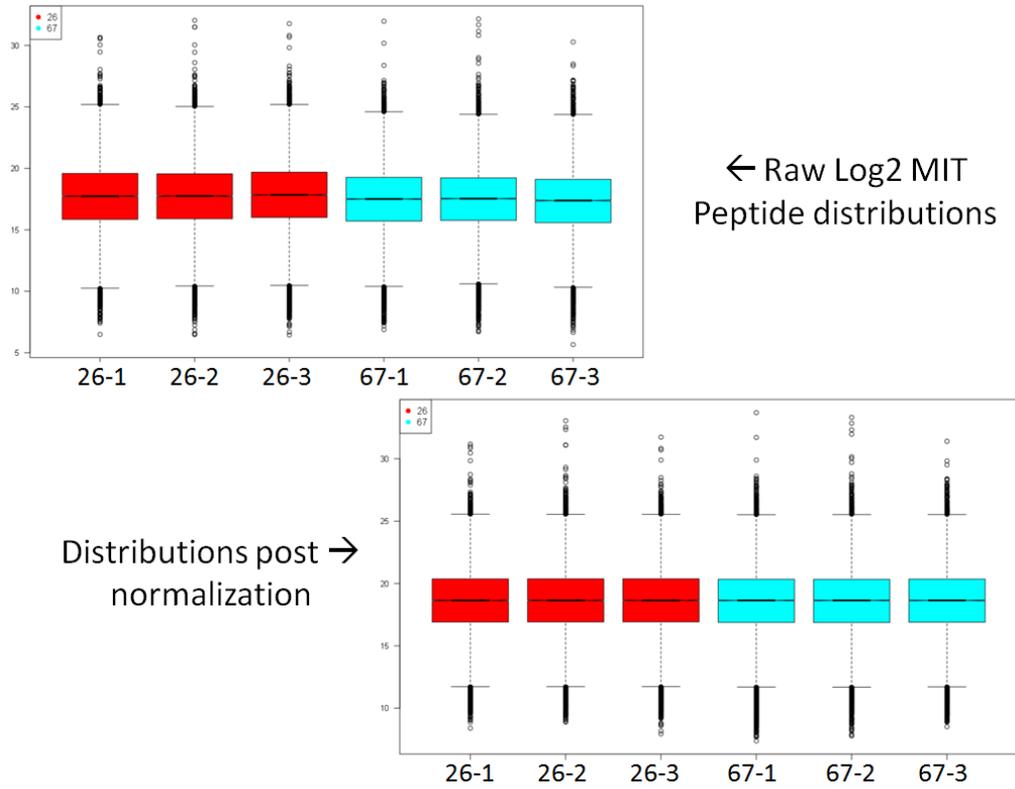
Figure S2. Effect of normalization on peptide distributions.

Box plots depicting log-normal peptide abundance distributions across samples (26 h = red; 67 h = blue) both before (top) and after (bottom) normalization by InfernoRDN (LOESS, MAD, & median centering). Raw abundance distributions were already modestly comparable and thus only small-scale adjustments were required before continuing with the quantitative analysis.

PCA Plot (84.5%)
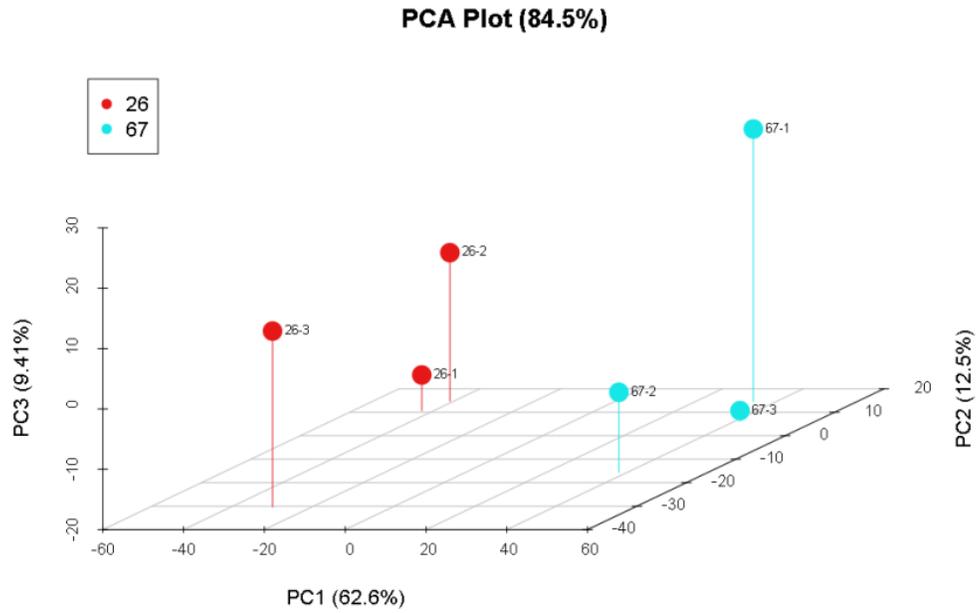
Figure S3. Principle component analysis shows a reproducible difference between sampling timepoints.

PCA depicts the S. noursei data dimensions that are differentiable across the samples. As shown, the main axis of differentiability is along the PC1 axis whereby corresponding replicates group according to the time point sampled (26 h = red; 67 h = blue). This indicates major quantitative differences at the protein-level between the two sets of data.

APPENDIX B. LIST OF SCIENTIFIC/TECHNICAL PUBLICATIONS

Published peer-reviewed articles

Mahan KM, Klingeman DM, Hettich RL, Parry RJ, Graham DE. 2016. Draft Genome Sequence of *Streptomyces vitaminophilus* ATCC 31673, a Producer of Pyrrolomycin Antibiotics, Some of Which Contain a Nitro Group. Genome Announc 4:e01582-01515.

Mahan KM, Zheng H, Fida TT, Parry RJ, Graham DE, Spain JC. 2017. Iron-Dependent Enzyme Catalyzes the Initial Step in Biodegradation of *N*-Nitroglycine by *Variovorax* sp. Strain JS1663. Appl Environ Microbiol 83:e00457-00417.

Conference or Symposium Abstracts

Mahan, K., Giannone, R., Klingeman, D.M., Close, D., Hettich, R.L., Graham, D., Fida, T., Spain, J. and Parry, R. Production of *N*-nitroglycine by *Streptomyces noursei* JCM 4701. Society for Industrial Microbiology and Biotechnology Annual Meeting and Exhibition. Philadelphia, PA.

Graham, D., Mahan, K., Giannone, R., Klingeman, D.M.., Close, D., Hettich, R.L., Gulvik, C., Spain, J., and Parry, R. 2015. Nitration in the biosynthesis of azomycin (2-nitroimidazole) by *Streptomyces eurocidicus*. Society for Industrial Microbiology and Biotechnology Annual Meeting and Exhibition. Philadelphia, PA.

Graham, D., Mahan, K., Gulvik, C., Giannone, R., Klingeman, D.M., Hettich, R.L., Parry, R., and Spain, J. 2015. Novel nitration reactions in the biosynthesis of 2-nitroimidazole by *Streptomyces eurocidicus*. Natural Product Discovery & Development in the Post-Genomic Era. San Diego, CA.

Mahan, K, Fida, T., Giannone, R., Hettich, R.L., Klingeman, D.M., Parry, R., Spain, J. and Graham, D. 2015. Production of *N*-nitroglycine by *Streptomyces noursei*. Natural Product Discovery & Development in the Post-Genomic Era. San Diego, CA.

APPENDIX C. OTHER SUPPORTING MATERIALS

None.