

# Natural Product Discovery through Improved Functional Metagenomics in *Streptomyces*

Hala A. Iqbal, Lila Low-Beinart, Joseph U. Obiajulu, and Sean F. Brady\*

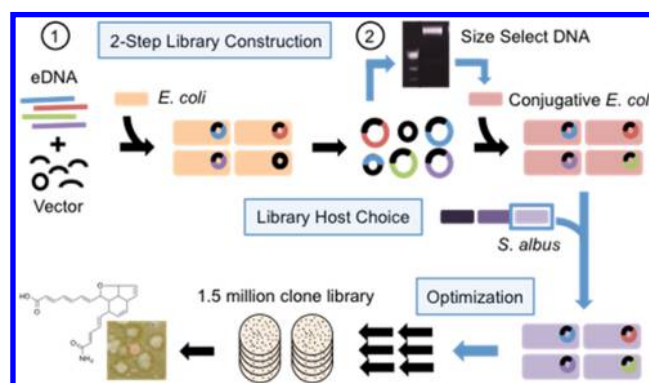
Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, New York 10065, United States

Supporting Information

**ABSTRACT:** Because the majority of environmental bacteria are not easily culturable, access to many bacterially encoded secondary metabolites will be dependent on the development of improved functional metagenomic screening methods. In this study, we examined a collection of diverse *Streptomyces* species for the best innate ability to heterologously express biosynthetic gene clusters. We then optimized methods for constructing high quality metagenomic cosmid libraries in the best *Streptomyces* host. An initial screen of a 1.5 million-membered metagenomic library constructed in *Streptomyces albus*, the species that exhibited the highest propensity for heterologous expression of gene clusters, led to the identification of the novel natural product metatricycloene (1). Metatricycloene is a tricyclic polyene encoded by a reductive, iterative polyketide-like gene cluster. Related gene clusters found in sequenced genomes appear to encode a largely unexplored collection of structurally diverse, polyene-based metabolites.

Cultured bacteria have historically been a very rich source of natural products with diverse structures and biological activities, despite making up as little as 1% of environmental microbiomes.<sup>1</sup> Sequencing efforts aimed at cataloging the biosynthetic potential of the uncultured majority have revealed the existence of large numbers of unexplored natural product gene clusters in most environmental microbiomes.<sup>2</sup> Functional metagenomics<sup>3</sup> offers a means of investigating natural products encoded by uncultured bacteria. In these studies, DNA extracted directly from an environmental sample (environmental DNA, eDNA) is cloned into an easily cultured bacterium. The resulting clones are screened for phenotypes associated with the production of small molecules (e.g., color, antibacterial activity, cytotoxicity).<sup>4,5</sup>

A prerequisite for identifying metabolites using functional metagenomics is the successful heterologous expression of eDNA-derived gene clusters in a host bacterium. Despite the fact that it is known to be a poor heterologous expression host, *Escherichia coli* has been used for the vast majority of such studies. Only a handful of functional metagenomic studies have used hosts other than *E. coli*.<sup>6–8</sup> As with *E. coli*, the hosts used in these studies were primarily selected for ease of growth and genetic tractability. An ideal host for functional metagenomics, on the other hand, would display high natural propensity to support the heterologous expression of diverse secondary



**Figure 1.** Improving functional metagenomics. Libraries are made in *E. coli* using a two-step process that removes small clones. High quality libraries are then conjugated into privileged expression hosts (*S. albus*) for phenotypic screening.

metabolite gene clusters as well as high transformation frequency to allow for the facile construction of large metagenomic libraries.

Because the ability to heterologously express gene clusters obtained by horizontal gene transfer could provide a selective advantage to some environmental bacteria, we hypothesized that (1) naturally privileged heterologous expression strains exist in the global microbiome and (2) functional metagenomic discovery efforts would benefit from identifying these gifted strains. Here, we report on improving functional metagenomics screening through the identification of natively privileged heterologous hosts and optimization of large metagenomic library construction methods (Figure 1). Implementation of these improvements has led to the discovery of carotenoids and the biosynthetically and structurally interesting polyene, metatricycloene (1) (Figure 2). The value of these improvements is highlighted by the fact that almost 20 years after the term metagenomics was first coined,<sup>3</sup> metatricycloene is one of the first biosynthetically complex natural products to be identified using functional metagenomic screening methods.<sup>4</sup>

Considering their history as prolific producers of diverse, therapeutically relevant metabolites, *Streptomyces* are an appealing source of heterologous expression hosts.<sup>9</sup> Nevertheless, previous attempts at functional metagenomic screening in *Streptomyces*, which have been conducted using exclusively

Received: March 21, 2016

Published: July 22, 2016

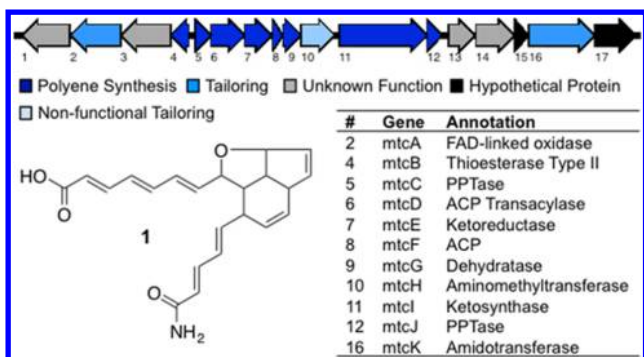


Figure 2. Metatrycycloene (1) structure and gene cluster.

*Streptomyces lividans* as a host, have largely been unsuccessful at identifying secondary metabolites.<sup>10–14</sup>

Genetically engineered strains of *Streptomyces* have been used to improve heterologous expression of individual gene clusters.<sup>15,16</sup> While these strains represent potential hosts for large-scale metagenomic library screening, we believe that the first step to developing the eventual best hosts is the identification of natively privileged heterologous expression strains. This would then be followed by the genetic engineering of these naturally privileged strains into even better hosts. In an effort to identify natively privileged *Streptomyces* species that would be more productive hosts for large metagenomic library screening efforts, we surveyed a collection of 39 *Streptomyces* strains for conjugation and heterologous expression efficiencies (Figure 3). These strains represent a phylogenetically diverse sampling of both commonly used and lesser-known species.

Target strains were interrogated for heterologous expression efficiency using a collection of 97 eDNA cosmid clones containing minimal type II polyketide synthases (PKS).<sup>17</sup> We used type II PKS gene clusters in this study because a large number of the final metabolites and shunt products produced by type II PKS biosynthesis are colored,<sup>18</sup> allowing for the easy visual identification of colonies heterologously expressing natural products. Colored phenotypes were observed with only four species, highlighting the inherent challenges of heterologous expression. *S. lividans*, *S. coeruleus*, and *S. viridochromogenes* each yielded one colored hit. *S. albus* yielded a total of 9 hits: 6 unique hits from phenotype screening and 3 from the subsequent examination of fermentation broths. Due to its superior heterologous expression capabilities and reasonable conjugation efficiency (Table S1), *S. albus* was selected for use as a metagenomic library host.

*E. coli* remains the most efficient host to use for constructing cosmid-based libraries from crude eDNA. For functional metagenomic screening using other hosts, libraries are first constructed in *E. coli* and then shuttled into a target expression host. In our experience, all *E. coli*-based eDNA cosmid libraries contain some empty vector and small-insert clone contamination. The appearance of small-insert clones in a large *E. coli*-based eDNA library is likely inevitable, because there will always be some clones that are not stable and are truncated during library propagation. Since small clones conjugate at up to 10,000-fold higher efficiencies than full length cosmids,<sup>19</sup> even a tiny amount of small clone contamination (<0.1–0.01%) will come to dominate a library following conjugation from *E. coli* into the expression host. We developed a two-step library construction protocol to address this issue (Figure 1). Libraries are first constructed and propagated in *E. coli* to allow for the truncation of any unstable clones. Cosmid DNA is then isolated in bulk from the primary library and size selected by gel electrophoresis. Following the removal of smaller clones, the collection of full-length stable cosmids is electroporated back into *E. coli*, and the resulting high-quality library is mated into the desired expression host.

For this study, we constructed a 1.5 million membered DNA cosmid library from Texas desert soil, using standard eDNA cosmid cloning methods.<sup>20</sup> This library was cleaned of small clones and conjugated into *S. albus* using mating conditions that we extensively optimized for shuttling large (>1 × 10<sup>6</sup> membered) metagenomic cosmid libraries into *S. albus* (Figure S1A–D). The quality of our final *S. albus* library was verified by restriction mapping of exconjugants. All clones we examined contained large insert cosmids with unique restriction maps (Figure S1E,F). This library is almost 2 orders of magnitude larger than any previously reported metagenomic library hosted in *Streptomyces*.<sup>12</sup>

The production of color by bacterial cultures is often an indication of small molecule biosynthesis. Color can therefore be used as a simple assay for identifying eDNA clones containing natural product biosynthetic gene clusters.<sup>6,7</sup> Though examining clones with colored phenotypes will undoubtedly identify known classes of metabolites, since color has been the target of many previous natural product isolation studies, it also provides the opportunity to identify novel metabolites from the rarely explored biosynthetic diversity present in soil metagenomes. As an initial exploration of the natural product-encoding potential of this *S. albus*-based metagenomic library, we screened exconjugants directly from

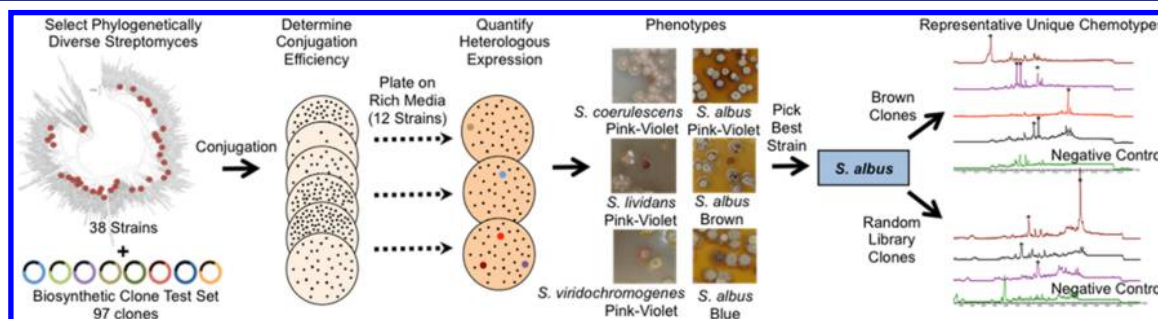
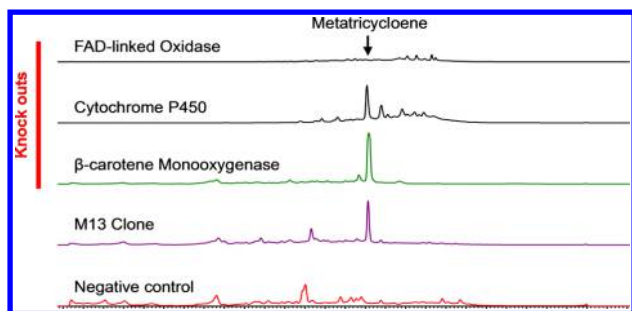


Figure 3. Identifying *Streptomyces* species for use in functional metagenomics. Phylogenetically diverse *Streptomyces* species were tested for conjugation and heterologous expression efficiency using a set of metagenomic clones containing minimal PKS genes. Exconjugants from highly transforming strains were plated on rich media to look for color phenotypes. Broth extracts were examined by HPLC to identify species inducing the most heterologous expression events.



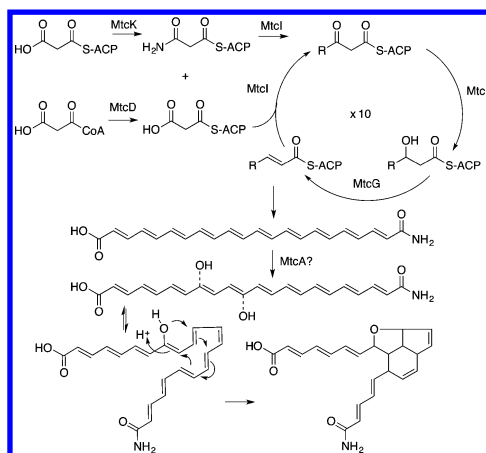
**Figure 4.** HPLC UV (270 nm) traces of culture broth extracts from M13 and M13 oxidase gene knockout mutants.

library mating plates for colored phenotypes. We picked 48 of the most deeply tinted colonies that appeared on these plates for further analysis. Upon retransformation of the cosmids isolated from these colored clones, 16 clones reproduced a color phenotype in *S. albus*. Dereplication by restriction mapping revealed 12 unique eDNA clones that displayed either yellow or orange (6), red (2), or brown (4) phenotypes (Figure S2). Each clone was fully sequenced, revealing that the eight yellow, orange or red clones contained carotenoid biosynthesis gene clusters with an assortment of tailoring genes (Figure S3). Based on LCMS-UV analysis of the acetone extracts of cell pellets and comparisons to commercial standards, the carotenoids produced by this set of clones were found to include lycopene,  $\beta$ -carotene, rhodopsin, dihydroxycopene, and three more highly oxidized carotenoids (Table S3). We did not investigate these clones in more detail because carotenoids have been widely explored in culture-based studies. It is worth noting that, apart from  $\beta$ -carotene, none of the carotenoids have been reported in functional metagenomic screens using other hosts.<sup>6</sup>

LCMS analysis of organic extracts from cultures of one of the brown colored clones (M13) revealed a single, major, clone-specific peak (Figure 4). This metabolite was purified (1.1 mg/L) from large-scale cultures and its structure was determined using 1 and 2D NMR as well as HRMS data. The final structure of (1) consists of a tricyclic core appended with two polyene substituents, one terminating in a carboxylic acid and the other terminating in an amide. We have named this metabolite metatricycloene. While metatricycloene did not show cytotoxicity against any bacteria, fungi, or human cells we tested, it was of interest to us because its structure does not closely resemble that of any described natural product.

Approximately one-third of the open reading frames present on M13 have their closest relative in the genome of a *Kibdelosporangium* species (Table S2).<sup>21</sup> The NCBI database contains only a small number of unique 16S rRNA sequences from cultured *Kibdelosporangium* strains. While a number of natural products have been identified from bacteria in this genus, it remains a rarely cultured Actinobacterial genera.<sup>21</sup>

Bioinformatic analysis of the open reading frames on M13 revealed a collection of polyketide/fatty acid synthesis-like genes (Figure 2, Table S2). This includes a gene predicted to encode a rarely described didomain ketosynthase (KS:KS\*) comprised of two FabF-like domains (MtcI), an acyl carrier protein (ACP) (MtcF), malonyl-CoA-ACP transacylase (AT) (MtcD), a 3-ketoacyl-ACP ketoreductase (KR) (MtcE), and an acyl-ACP dehydratase (DH) (MtcG). Based on the genes present on M13 and the structure of metatricycloene, we proposed the biosynthetic scheme shown in Figure 5.



**Figure 5.** Proposed biosynthetic scheme for (1). Dashed bonds indicate potential locations for MtcA oxidation.

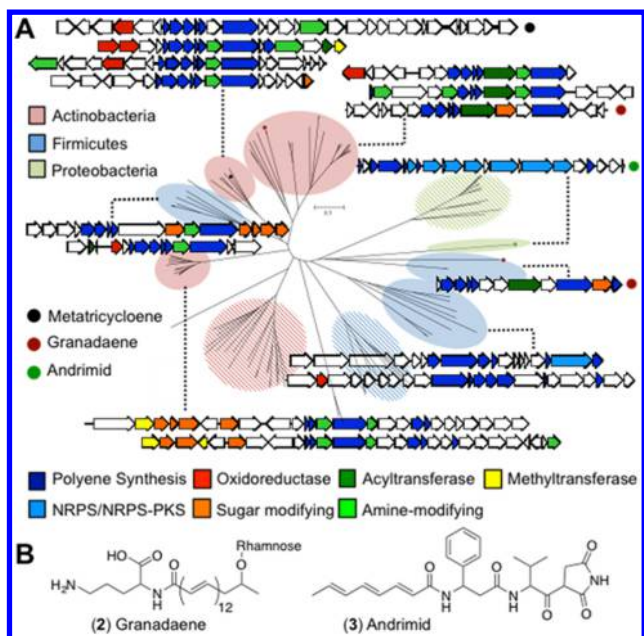
In our proposal, the terminal amide arises from a malonyl starter unit that is modified by the predicted amidotransferase MtcK. MtcK is 66% identical to OxyD, which carries out the same reaction in oxytetracycline biosynthesis. The ACP and KS:KS\* proteins are predicted to catalyze 10 rounds of chain elongation using malonyl CoA. Only the first KS domain in MtcI contains a complete complement of catalytic residues. This KS domain is therefore predicted to be responsible for chain elongation. The second KS domain (KS\*) is missing a conserved active site histidine (H303, Figure S6), and we therefore believe it functions as a chain length factor (CLF). Contrary to what is seen in type II aromatic polyketide biosynthesis where three distinct proteins (ACP, KS, and CLF) are used to generate a nonreduced polyacetate, the polyketide precursor in metatricycloene biosynthesis is reduced to a polyene by predicted ketoreductase (MtcE) and dehydratase (MtcG) enzymes (Figure 5). In our scheme, the polyene is then oxidized and undergoes one C–O and two C–C cyclizations to give the tricyclic ring system.

To better understand the biosynthesis of metatricycloene, we saturated the M13 cosmid with transposons. Cosmids containing transposons in genes predicted to encode three different oxidation enzymes were individually transformed back into *S. albus* and assayed for the ability to encode the production of metatricycloene. Only in the case of the FAD-linked oxidase knockout (MtcA) was (1) no longer observed in culture broth extracts (Figure 4). It remains to be seen whether this oxidase is solely involved in the oxidation of metatricycloene or whether it also “guides” the cyclization of (1), as metatricycloene is the only major clone specific metabolite seen in M13 cultures.

Metatricycloene biosynthesis is related to that of the aryl polyenes.<sup>22</sup> Aryl polyenes, however, arise from the use of distinct KS and CLF proteins, similar to what is seen in aromatic PKS biosynthesis. Reductive biosynthetic systems using KS:KS\*-like didomain proteins have been characterized from culture-based bacterial screening efforts in only a few instances. Two examples are granadaene (2) and andrimid (3) (Figure 6B).<sup>23,24</sup> In granadaene biosynthesis, the polyene is decorated with ornithine and the deoxy sugar rhamnose. In andrimid biosynthesis, the polyene is incorporated into a hybrid nonribosomal peptide/polyketide.

A search of sequenced genomes for KS:KS\* didomain homologues revealed a number of KS:KS\* genes that are





**Figure 6.** (A) Phylogenetic tree of KS:KS\* proteins. Representative gene clusters associated with KS:KS\* genes from different clades. Striped clades contain clusters that do not have complete sets (KS:KS\*, DH, and KR) of polyene encoding genes. (B) Characterized molecules that use didomain ketosynthase proteins.

adjacent to DH, KR, and ACP genes, as well as unique sets of genes potentially encoding tailoring enzymes. KS:KS\* genes group by phylogenetic origin and are often found in species that have rarely been explored for the production of bioactive secondary metabolites, such as rare Actinobacteria and Firmicutes. The diverse collection of potential tailoring genes found around KS:KS\* homologues suggests that they may be involved in the biosynthesis of overlooked polyene-based natural products (Figure 6A). For example, within the KS:KS\* clade in which MtcI falls, there are multiple gene clusters that appear to encode different tailoring enzymes including diverse oxidases. Due to the underexplored nature of this reductive class of polyketides, the KS:KS\* gene should be a productive probe for identifying gene clusters that encode novel natural products in sequenced genomes and metagenomes.

Although the limitations of *S. albus* as a metagenomic screening host remain to be seen, our discovery of both isoprenoid (carotenoid) and polyketide (metatricycloene) encoding gene clusters suggests that it will prove useful for identifying gene clusters that are more biosynthetically diverse than the type II PKS clusters that we used for strain selection. The identification of metatricycloene and multiple carotenoid biosynthetic gene clusters are rare examples of identifying complex biosynthetic gene clusters using functional metagenomics.<sup>4</sup> Based on the work presented here, a larger-scale exploration of bacteria for gifted heterologous expressers, coupled with genome engineering of naturally privileged heterologous expression hosts and larger-insert eDNA cloning methods should help to make functional metagenomics a more routine tool for identifying novel natural products.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b02921.

Additional experimental details, NMR data, gene table, TE knockout data, and colony images (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*sbrady@rockefeller.edu

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Carolina Adura Alcaino and HTSRC for their help. This work was supported by Grant U01 GM110714.

## ■ REFERENCES

- (1) Torsvik, V.; Goksoyr, J.; Daae, F. L. *Appl. Environ. Microbiol.* **1990**, *56*, 782.
- (2) Charlop-Powers, Z.; Owen, J. G.; Reddy, B. V.; Ternei, M. A.; Brady, S. F. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 3757.
- (3) Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. *Chem. Biol.* **1998**, *5*, R245.
- (4) Iqbal, H. A.; Feng, Z.; Brady, S. F. *Curr. Opin. Chem. Biol.* **2012**, *16*, 109.
- (5) Banik, J. J.; Brady, S. F. *Curr. Opin. Microbiol.* **2010**, *13*, 603.
- (6) Craig, J. W.; Chang, F. Y.; Brady, S. F. *ACS Chem. Biol.* **2009**, *4*, 23.
- (7) Craig, J. W.; Chang, F. Y.; Kim, J. H.; Obiajulu, S. C.; Brady, S. F. *Appl. Environ. Microbiol.* **2010**, *76*, 1633.
- (8) Martinez, A.; Kolvek, S. J.; Yip, C. L.; Hopke, J.; Brown, K. A.; MacNeil, I. A.; Osburne, M. S. *Appl. Environ. Microbiol.* **2004**, *70*, 2452.
- (9) de Lima Procopio, R. E.; da Silva, I. R.; Martins, M. K.; de Azevedo, J. L.; de Araujo, J. M. *Braz. J. Infect. Dis.* **2012**, *16*, 466.
- (10) Lussier, F. X.; Chambenoit, O.; Cote, A.; Hupe, J. F.; Denis, F.; Juteau, P.; Beaudet, R.; Shareck, F. *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 1321.
- (11) Martinez, A.; Kolvek, S. J.; Hopke, J.; Tiong Yip, C. L.; Osburne, M. S. *Appl. Environ. Microbiol.* **2005**, *71*, 1638.
- (12) McMahon, M. D.; Guan, C.; Handelsman, J.; Thomas, M. G. *Appl. Environ. Microbiol.* **2012**, *78*, 3622.
- (13) Wang, G. Y.; Graziani, E.; Waters, B.; Pan, W.; Li, X.; McDermott, J.; Meurer, G.; Saxena, G.; Andersen, R. J.; Davies, J. *Org. Lett.* **2000**, *2*, 2401.
- (14) Courtois, S.; Cappellano, C. M.; Ball, M.; Francou, F. X.; Normand, P.; Helynck, G.; Martinez, A.; Kolvek, S. J.; Hopke, J.; Osburne, M. S.; et al. *Appl. Environ. Microbiol.* **2003**, *69*, 49.
- (15) Gomez-Escribano, J. P.; Bibb, M. J. *Microb. Biotechnol.* **2011**, *4*, 207.
- (16) Komatsu, M.; Uchiyama, T.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2646.
- (17) Feng, Z.; Kallifidas, D.; Brady, S. F. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 12629.
- (18) Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380.
- (19) Hanahan, D. *J. Mol. Biol.* **1983**, *166*, 557.
- (20) Brady, S. F. *Nat. Protoc.* **2007**, *2*, 1297.
- (21) Ogasawara, Y.; Torrez-Martinez, N.; Aragon, A. D.; Yackley, B. J.; Weber, J. A.; Sundararajan, A.; Ramaraj, T.; Edwards, J. S.; Melancon, C. E., 3rd *Genome Announce.* **2015**, *3*, e00114-15.
- (22) Schoner, T. A.; Gassel, S.; Osawa, A.; Tobias, N. J.; Okuno, Y.; Sakakibara, Y.; Shindo, K.; Sandmann, G.; Bode, H. B. *ChemBioChem* **2016**, *17*, 247.
- (23) Whidbey, C.; Harrell, M. I.; Burnside, K.; Ngo, L.; Becraft, A. K.; Iyer, L. M.; Aravind, L.; Hitti, J.; Adams Waldorf, K. M.; Rajagopal, L. *J. Exp. Med.* **2013**, *210*, 1265.
- (24) Jin, M.; Fischbach, M. A.; Clardy, J. *J. Am. Chem. Soc.* **2006**, *128*, 10660.