Directed natural product biosynthesis gene cluster capture and expression in the model bacterium Bacillus subtilis

Citation
Li, Yongxin, Zhongrui Li, Kazuya Yamanaka, Ying Xu, Weipeng Zhang, Hera Vlamakis, Roberto Kolter, Bradley S. Moore, and Pei-Yuan Qian. 2015. “Directed natural product biosynthesis gene cluster capture and expression in the model bacterium Bacillus subtilis.” Scientific Reports 5 (1): 9383. doi:10.1038/srep09383. http://dx.doi.org/10.1038/srep09383.

Published Version
doi:10.1038/srep09383

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:27662204

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Directed natural product biosynthesis gene cluster capture and expression in the model bacterium *Bacillus subtilis*

Yongxin Li1,2, Zhongrui Li1, Kazuya Yamanaka2,3, Ying Xu1, Weipeng Zhang1, Hera Vlamakis4, Roberto Kolter4, Bradley S. Moore2,5 & Pei-Yuan Qian1

1KAUST Global Collaborative Research, Division of Life Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, 2Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093, United States, 3JNC Corporation, Yokohama Research Center, 5-1 Okawa, Kanazawa-ku, Yokohama, Kanagawa 2368605, Japan, 4Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, United States, 5Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093, United States.

Bacilli are ubiquitous low G+C environmental Gram-positive bacteria that produce a wide assortment of specialized small molecules. Although their natural product biosynthetic potential is high, robust molecular tools to support the heterologous expression of large biosynthetic gene clusters in *Bacillus* hosts are rare. Herein we adapt transformation-associated recombination (TAR) in yeast to design a single genomic capture and expression vector for antibiotic production in *Bacillus subtilis*. After validating this direct cloning “plug-and-play” approach with surfactin, we genetically interrogated amicoumacin biosynthetic gene cluster from the marine isolate *Bacillus subtilis* 1779. Its heterologous expression allowed us to explore an unusual maturation process involving the N-acyl-asparagine pro-drug intermediates preamicoumacins, which are hydrolyzed by the asparagine-specific peptidase into the active component amicoumacin A. This work represents the first direct cloning based heterologous expression of natural products in the model organism *B. subtilis* and paves the way to the development of future genome mining efforts in this genus.

Next generation sequencing and genome mining technologies have revolutionized the discovery of natural product chemicals and biosynthetic enzymes that help fuel the fields of biotechnology and biomedicine. Based on a recent comprehensive survey of publically available bacterial genomes, three phyla account for the majority of natural product biosynthetic potential, namely Actinobacteria, Proteobacteria, and Firmicutes. While sophisticated molecular biology techniques have been firmly established to connect biosynthetic gene clusters to encoded natural product molecules in Actinobacteria and Proteobacteria through the use of model expression hosts such as *Streptomyces coelicolor* and *Escherichia coli*, Firmicutes, which contain the natural product rich genus *Bacillus*, are lacking comparable molecular tools to support the heterologous expression of large natural product pathways. Here we report the design and implementation of a versatile vector to support the direct capture of *Bacillus* biosynthetic gene clusters from genomic DNA by transformation-associated recombination (TAR) in yeast and heterologous expression in the model host *B. subtilis*.

*B. subtilis* is a low G+C, Gram-positive bacterium that has been commonly used for decades in genetic and biochemical studies of chromosome replication and bacterial sporulation. This species is an attractive option for the heterologous production of natural products for three main reasons. First, the *Bacillus* genus produces a wide assortment of biologically active small molecules, including antibacterial non-ribosomal cyclic lipopeptides of the surfactin and gageotetrin families, polyketides such as macroactin and bacillaene, antitumor polyketide-peptide hybrids like amicoumacin and ieooglucomide, and the discopyrrole alkaloids (Fig. 1; Fig. S1). Second, *B. subtilis* has the capacity for natural genetic competence and subsequent homologous recombination, allowing the introduction of foreign DNA. This feature offers a wide range of available genetic manipulation techniques to facilitate practical biosynthetic efforts of natural products. Indeed, the natural transformation system of *B. subtilis* is so effective that the 3.5-megabase genome of *Synechocystis* PCC6803 was successfully assembled into the *B. subtilis* 168 genome, which served as a cloning vector (also known as the *Bacillus* Genome [BGM] vector). This BGM cloning system has been applied to demonstrate the cloning of the entire mouse mitochondrion and rice chloroplast genomes. And third, *B. subtilis* is non-pathogenic and is...
Generally recognized as a safe production host that can satisfy safety requirements for the industrial productions of drug leads and enzymes\cite{17, 18}.

Despite these technical advantages and that \textit{B. subtilis} is routinely used for protein expression, there are few reports utilizing \textit{B. subtilis} as an expression host system for the production of natural product small molecules. Most reports have used chromosomal transfer or cosmid library expression techniques and have focused on relatively small pathways of ribosomal and non-ribosomal peptide products\textsuperscript{19–27}. However, the limitations of gene cluster cloning via clone libraries\textsuperscript{9} and chromosomal transfer hinder the efficient study of gene clusters from undomesticated producers or large pathways. Furthermore, a critical bottleneck in using \textit{B. subtilis} as a heterologous host is the lack of autonomous plasmids to facilitate cloning, transfer and heterologous expression of large biosynthetic gene clusters\textsuperscript{9}.

We recently developed a new genetic platform for the efficient capture of a silent 67-kb biosynthetic gene clusters directly from genomic DNA via TAR in yeast to support a “plug-and-play” approach to small molecule production\textsuperscript{28}. To date we have captured and expressed high G+C Gram-positive actinomycete pathways for the marinopyrrole taromycin and enterocin\textsuperscript{28, 29} antibiotics and the Gram-negative pseudoalteromonad pathway for the alterochromides\textsuperscript{30}. Herein we adapted this platform to support the capture and expression of low G+C Gram-positive bacilli-based natural products. We validated the method with the prototype \textit{Bacillus} lipo-peptide surfactins and the hybrid polyketide-peptide amicoumacins, which exhibit broad bioactivities, including antibacterial, antifungal and antitumor activities\textsuperscript{31–33}. This work represents a very useful approach to interrogate the function of biosynthetic gene clusters in \textit{Bacillus} through heterologous biosynthesis.

**Results**

**Design and validation of the pCAPB gene cluster capture vectors.**

The gene cluster capture vector pCAP01 consists of three elements that allow direct capture and manipulation in yeast, maintenance and manipulation in \textit{Escherichia coli}, and chromosomal integration and expression of cloned pathways in actinomycetes\textsuperscript{30}. To repurpose this vector for \textit{Bacillus} expression, we first replaced the actinomycete elements with the \textit{Bacillus} element from pBU4\textsuperscript{34} to generate the yeast/\textit{E. coli} shuttle-\textit{B. subtilis} capture vector pCAPB1 in order to support conjugal transfer into various \textit{Bacillus} species (Fig. S2). We evaluated this replication plasmid pCAPB1 with the prototype \textit{Bacillus} lipopeptide surfactin, which is encoded on the 38-kb \textit{srf} locus from \textit{B. subtilis} 1779. Although we successfully captured the \textit{srf} locus from genomic DNA via TAR and constructed the pCAPB1-\textit{srf} vector (Fig. 2, Fig. S3), upon its transfer into five \textit{Bacillus} host strains, the vector was not stable and did not allow for surfactin production.

To overcome the instability issue of pCAPB1, we next designed pCAPB2 based on the \textit{amyE} chromosomal integration plasmid pDR111, which is used in heterologous protein expression experiments in \textit{B. subtilis}\textsuperscript{35, 36}. We incorporated the yeast elements from pCAP01 into a derivative of pDR111 to give the yeast/\textit{E. coli} shuttle-\textit{B. subtilis} chromosome integrative capture vector pCAPB2 (Fig. 2). This plasmid is maintained as a single copy in yeast cells to avoid unintended multiple recombination events during TAR, while it functions at multiple copies in \textit{E. coli} to provide sufficient plasmid DNA materials for transfer to \textit{B. subtilis}. The vector was designed to allow specific integration of cloned gene clusters into the chromosome of \textit{B. subtilis} JH642 via double crossover recombination into the \textit{amyE} gene. We validated the function of pCAPB2 by direct cloning of the \textit{srf} gene cluster from \textit{B. subtilis} 1779 to give the integration plasmid pCAPB2-\textit{srf} (Fig. 2, Fig. S3). Upon its introduction into \textit{B. subtilis} ROM77, in which the native \textit{srf} locus was disrupted (JH642, \textit{srfAA::cat})\textsuperscript{37}, we clearly detected surfactin production at wild-type levels by UPLC-MS (Fig. S4).

**Identification and TAR capture of amicoumacin biosynthesis (ami) genes.**

With the successful completion of the proof-of-principle experiment to directly capture and express a \textit{Bacillus} natural product pathway in \textit{B. subtilis}, we next turned our attention to an uncharacterized \textit{Bacillus} pathway to further showcase the proficiency of the pCAPB2 system. We selected to evaluate the amicoumacins, which are bioactive isocoumarin natural products that were first reported from \textit{Bacillus pumilus} in 1981\textsuperscript{31}. Recent biochemical studies revealed that amicoumacin belongs to a new class of protein synthesis inhibitors that binds to the ribosome\textsuperscript{38, 39}. We re-isolated the amicoumacins from the Red Sea isolate \textit{B. subtilis} 1779 following a bioassay-guided isolation procedure to give amicoumacins A-C (1–3) and \textit{O-methylamicoumacin} B (4) (Fig. 1), based on high-resolution MS and NMR characterization. We suspect that some of the amicoumacin analogues may be byproducts generated from amicoumacin A (1) during the isolation and sample preparation processes. To capture and express the amicoumacin (ami) gene.

**Figure 1 |** Chemical structure of amicoumacins (1–4) and surfactin.

---

\textsuperscript{1} amicoumacin A  \textsuperscript{2} amicoumacin B  \textsuperscript{3} O-methylamicoumacin B

\textsuperscript{R} \textsuperscript{1} = H  \textsuperscript{R} \textsuperscript{2} = NH\textsubscript{2}

\textsuperscript{R} \textsuperscript{1} = H  \textsuperscript{R} \textsuperscript{2} = OH

\textsuperscript{R} \textsuperscript{1} = Me  \textsuperscript{R} \textsuperscript{2} = OH

---

**OH**

**OH**

**NH\textsubscript{2}**

**NH\textsubscript{2}**

**OH**

**R_1**

**R_2**

**COR_2**

**OH**

**OH**

**COOH**

**N**

**N**

**H**

**H**

**HOOC**

**NH**

**NH**

**NH**

**OH**

**COOH**

**surfactin A**
cluster, which we hypothesized to be encoded by a hybrid modular polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS), we sequenced the wild-type producer B. subtilis 1779 genome. Sequence analysis revealed several assembly line biosynthetic gene clusters, including a contiguous region of 47.4 kb containing 16 open reading frames (ORFs) that we predicted were responsible for amicoumacin biosynthesis (Fig. 3A, Table S3). We designated these genes amiA-O. Inspection of this locus identified an NRPS-PKS hybrid protein encoded by amiI, two NRPSs encoded by amiA and amiJ, and three PKS genes encoded by amiK-M. The predicted biosynthesis gene cluster consists of eight modules in total, as shown in Fig. 3, for the incorporation of three amino acids and five malonate residues. Based on the co-linearity rule of assembly line biosynthesis, we suspected that the predicted initiation module encoded by amiA synthesizes a fatty acyl-D-Asn residue reminiscent of lipopeptide natural products, suggesting that the product of the AmiA-M megasynthetase may be a derivative of lipoamicoumacins A–D that we previously reported40. We thus explored the possibility that the immediate product of the ami biosynthetic pathway may not be amicoumacin A but rather a lipitated precursor that may support a pro-drug-like activation mechanism (Fig. 3). The unusual dihydroisocoumarin core structure is likely formed by the terminating AmiJ–M megasynthetase proteins to generate a highly oxygenated polyketide chain that rearranges into the bicyclic dihydroisocoumarin moiety. Feeding experiments with 15N2-L-asparagine and 5,5,5-trifluoro-DL-leucine followed by MS analyses further supported the proposed amicoumacin A biosynthetic pathway (Table S4).

Owing to the projected unusual natural product activation mechanism, we initially interrogated the biosynthetic pathway of amicoumacins within the wild-type producer strain B. subtilis 1779. However, all of our attempts to disrupt target genes in the native strain were unsuccessful, as had been other attempts to previously analyze amicoumacin production41. Therefore, to study the amicoumacin biosynthesis pathway, we targeted the genomic region containing the 47.4-kb ami gene cluster for TAR and heterologous expression in either Bacillus subtilis JH642 for natural products expression studies. (B) Physical map of the gene cluster capture vector pCAPB2 used in TAR direct cloning. The vector consists of three elements that allow direct cloning of pathways in yeast (blue), maintenance and manipulation in E. coli (yellow), and chromosomal integration and expression of cloned pathways in B. subtilis (red). The yeast element consists of ARSH4/CEN6 (replication origin) and TRP1 auxotrophic marker, while the E. coli and the Bacillus elements consist of DNA sequence for integration into the B. subtilis amyE gene, the lac repressor lacI, a spectinomycin resistance gene (specR) for Bacillus and an ampicillin resistance gene for E. coli (ampR). For the construction of a pathway specific capture vector, homology arms corresponding to both ends of the pathway are introduced into the capture arm cloning sites.
Heterologous expression of the amicoumacin biosynthesis genes.

For the heterologous expression of the ami gene cluster, we introduced the integration plasmid pCAPB2-ami into the genome of B. subtilis JH642 in which the phosphopantetheinyl transferase gene sfp has been added. The resultant transformants successfully produced amicoumacins 1–4 at comparative levels to that in B. subtilis 1779, as revealed by UPLC-MS (Fig. 5A) and NMR analyses. We additionally expressed the ami cluster in E. coli BL21(DE3) via the construction of a different capture vector based on pETDuet-1, which also resulted in amicoumacin production, albeit at levels 100-fold less than that in the native B. subtilis 1779 strain or in the B. subtilis JH642 host (Fig. 5B, Fig. S6).

The heterologous production of the amicoumacins in the B. subtilis and E. coli hosts provided unequivocal evidence that the ami locus encodes amicoumacin biosynthesis. With these systems in hand, we further interrogated the function of amiA by λ-Red recombination-mediated PCR targeting in E. coli BW25113. Restriction mapping of the plasmid propagated in E. coli confirmed that the amiA gene was successfully replaced by a gene encoding apramycin resistance to yield pCAPB2-ami (ΔamiA) (Fig. S5). This construct was integrated into the chromosome of the B. subtilis host, whereupon we observed that amicoumacin production was now lost (Fig. 5A).

Pro-drug mechanism of amicoumacin activation.

With the successful heterologous expression of the amicoumacin biosynthetic gene cluster and the ability to readily inactivate individual ami genes, we next explored the molecular and functional relationship between the various structures and the possibility of a pro-drug-like activation strategy. To this end, we first examined amiB, which codes for a D-Asp peptidase homologous to XcnG and ClbP that convert inactive precursors into the antibiotics xenocoumacin and colibactin.
To evaluate if AmiB is similarly involved in activating lipoamicoumacin-like precursors to form mature amicoumacin antibiotics, we mutated the amiiB gene in the pCAPB2-ami plasmid (Fig. S5) and expressed pCAPB2-ami (ΔamiB) in the B. subtilis host. The mutant strain lacking amiB was then analyzed by UPLC-MS analysis (Fig. 5A). Indeed, as predicted, new shunt products were produced instead of amicoumacins in the amiB deletion mutant. These products were isolated and characterized as preamicoumacins A–B (5–6) (Fig. 3B) by comprehensive NMR, MS, and Marley analyses (Table S5, Fig. S7–S8, and Supplemental Experimental Procedures). Preamicoumacinin A resembles lipoamicoumacin A and specifically differs from amicoumacin A by tailoring the amine group at C-10 by comprehensive NMR, MS, and Marley analyses (Table S5, Fig. S7–S8, and Supplemental Experimental Procedures). Preamicoumacin A resembles lipoamicoumacin A and specifically differs from amicoumacin A by tailoring the amine group at C-10 with N-acyl Asn as predicted bioinformatically. The configuration of the N-acyl Asn residue was assigned as D on the basis of the advanced Marley’s methods.67,68. Both 5 and 6 represent derivatives of amicoumacin A extended at the N terminus by D-Asn carrying two different acyl N-acyl chains (Fig. 3B).

With the structures of preamicoumacins A–B in hand, we were able to investigate the biochemical function of the AmiB, which is predicted to be a membrane-associated peptidase. To directly observe the cleavage of preamicoumacins to amicoumacin A, we heterologously expressed AmiB in B. subtilis JH642 for in vivo tests of proteolytic activity against preamicoumacins. Addition of exogenous 5 to B. subtilis carrying the gene amiB resulted in its conversion to 1 (Fig. 5C). While amicoumacin A (1) is active against Staphylococcus aureus (UST950701-005) with an MIC of 5.0 μg mL⁻¹, derivatives 5 and 6 were inactive (Table S6). These results support the biosynthetic scenario whereupon inactive preamicoumacin precursors are first synthesized and then converted to the active component amicoumacin A.

Discussion

Connecting genes to molecules with the help of efficient heterologous expression techniques is beginning to fundamentally change the natural product discovery paradigm. Herein we add the Bacillus antimicrobial compounds surfactin and amicoumacins to the small yet growing list of TAR captured and heterologously expressed microbial compounds, thereby opening up the metabolically rich Bacillus genus to future natural product discovery efforts. In the present study, the TAR-directed capturing of the amicoumacin biosynthesis gene cluster from a marine B. subtilis isolate allowed for its heterologous expression and biosynthetic interrogation in a host B. subtilis strain, which represents a common genetic procedure practiced in other bacterial systems but not before with Bacillus. Our mutational work allowed us to establish an antibiotic “pro-drug” activation pathway in which newly discovered preamicoumacins are converted by the AmiB peptidase into the biologically active isocoumarin antibiotic in a process resembling xenocoumacin and colibactin processing in Gram-negative bacteria but not before observed in Gram-negative bacteria. This study was greatly facilitated with the pCAPB2 expression system that should similarly support the discovery and characterization of new chemical entities and enzymatic processes in other bacilli, which is an active pursuit of our laboratories.

Methods

Strains, fermentation, and isolation of amicoumacin compounds. All strains, plasmids and oligonucleotides used in this study are listed in Table S1–S2. Bacillus subtilis 1779 was isolated from seawater collected from the Red Sea during a 2010 research cruise. Its crude extract showed strong antibacterial activity against S. aureus UST950701-005. Detailed culture conditions and isolation procedures of amicoumacins A–C and O-methylamicoumacin B (1–4) are described in the Supplemental Experimental Procedures. For identification of preamicoumacins A–B, B. subtilis JH642 + sfp that carried pCAPB1-ami (ΔamiB) was cultured in LB media for 24 hours, extracted with ethyl acetate, and isolated by preparative RP-HPLC (60–100% MeCN in 0.1% trifluoroacetic acid, 40 min gradient) to obtain pure compounds. Analytical details of the preamicoumacins are provided in the Supplemental Information.

Genome sequencing, annotation, and bioinformatics analysis of the ami gene cluster. The draft genome of B. subtilis 1779 was sequenced on an Illumina Hiseq2000 to generate 490-fold coverage of the 4.25 Mb genome. A total of 10.5 million pairs of Illumina reads were obtained from a 200 bp paired-end library. Genome assembly was performed with Velvet 1.0.15 with the following custom parameters: hash-length = 55 and coverage cutoff = 30. The draft genome sequences were deposited in GenBank as accession number SR5606752. The bioinformatics program antiSMASH (http://antismash.secondarymetabolites.org/) was initially used to analyze the whole draft genome sequence. The sequence of the orphan 47.4-kb NRPS/PKS hybrid gene cluster amiA-O encoded on contig-29 (95,452–140,246 nt) was further predicted and annotated using Pfam analyses (http://pfam.sanger.ac.uk/) and protein-protein BLAST (http://blast.ncbi.nlm.nih.gov). NRPS A domain specificities were analyzed using online program NRPSpredictor2 (http://nrps.informatik.uni-tuebingen.de/Controller?cmd=SubmitJob)51,52.

Construction of the gene cluster capture vectors pCAPB1 and pCAPB2. Our initial attempts of replication plasmid pCAPB1 generated from our previous capture vector for Streptomyces pCAP01 by replacing Streptomyces element with Bacillus element respectively. To evaluate if AmiB is similarly involved in activating lipoamicoumacin-like precursors to form mature amicoumacin antibiotics, we mutated the amiiB gene in the pCAPB2-ami plasmid (Fig. S5) and expressed pCAPB2-ami (ΔamiB) in the B. subtilis host. The mutant strain lacking amiB was then analyzed by UPLC-MS analysis (Fig. 5A). Indeed, as predicted, new shunt products were produced instead of amicoumacins in the amiB deletion mutant. These products were isolated and characterized as preamicoumacins A–B (5–6) (Fig. 3B) by comprehensive NMR, MS, and Marley analyses (Table S5, Fig. S7–S8, and Supplemental Experimental Procedures). Preamicoumacinin A resembles lipoamicoumacin A and specifically differs from amicoumacin A by tailoring the amine group at C-10 with N-acyl Asn as predicted bioinformatically. The configuration of the N-acyl Asn residue was assigned as D on the basis of the advanced Marley’s methods. Both 5 and 6 represent derivatives of amicoumacin A extended at the N terminus by D-Asn carrying two different acyl N-acyl chains (Fig. 3B).

With the structures of preamicoumacins A–B in hand, we were able to investigate the biochemical function of the AmiB, which is predicted to be a membrane-associated peptidase. To directly observe the cleavage of preamicoumacins to amicoumacin A, we heterologously expressed AmiB in B. subtilis JH642 for in vivo tests of proteolytic activity against preamicoumacins. Addition of exogenous 5 to B. subtilis carrying the gene amiB resulted in its conversion to 1 (Fig. 5C). While amicoumacin A (1) is active against Staphylococcus aureus (UST950701-005) with an MIC of 5.0 μg mL⁻¹, derivatives 5 and 6 were inactive (Table S6). These results support the biosynthetic scenario whereupon inactive preamicoumacin precursors are first synthesized and then converted to the active component amicoumacin A.

Discussion

Connecting genes to molecules with the help of efficient heterologous expression techniques is beginning to fundamentally change the natural product discovery paradigm. Herein we add the Bacillus antimicrobial compounds surfactin and amicoumacins to the small yet growing list of TAR captured and heterologously expressed microbial compounds, thereby opening up the metabolically rich Bacillus genus to future natural product discovery efforts. In the present study, the TAR-directed capturing of the amicoumacin bio-

Figure 4 | Physical maps of the TAR-cloned ami gene cluster and heterologous expression vectors. The 47.4-kb genomic region containing the ami gene cluster (blue) was directly cloned in yeast, yielding pCAPB1-ami and pCAPB2-ami. The pCAPB1 backbone and unrelated gene orf1 on pCAPB1-ami (dotted arrows) were replaced with the ami specific capture vector pCAPE in E. coli cells via λ-red mediated recombination, generating ami gene cluster E. coli expression construct pCAPE-ami.
from plasmid pBU4 was not successful for heterologous expression in *B. subtilis*. To generate the integration vector pCAPB2, the yeast element consisting of ARSH4/CEN6 (replication origin) and TRP1 auxotrophic marker from pCAP01, the *B. subtilis* and the *Bacillus* elements consisting of DNA sequence for integration into the *B. subtilis* *amyE* gene, the lac repressor *lacI* and an IPTG-inducible promoter, a spectinomycin resistance gene (*spec*) for *Bacillus* and an ampicillin resistance gene for *E. coli* (*amp*) from the pDR111 were assembled in *E. coli* Top10. For capture vector of heterologous expression in *E. coli* BL21 (DE3), the phosphopantetheine transferase (PPTase) gene *sfp* was inserted into MCS2 of pETDuet-1 to generate capture vector pCAPE. Detailed information is provided in Supplemental Experimental Procedures.

**Direct cloning of the ami gene cluster using TAR.** Producer strain *B. subtilis* 1779 was grown in LB liquid medium overnight and genomic DNA was isolated from stationary phase cells. Approximately, 20 μg of genomic DNA were digested with 400 U of Scal or SpeI, which did not cut the *ami* or *srf* gene clusters, respectively, in an overnight reaction at 37°C. The *ami* pathway-specific capture vector was constructed by introducing two PCR-amplified 1-kb homology arms corresponding to upstream and downstream regions of the *ami* gene cluster (orf1 and orf3) into capture vector pCAPB1 and pCAPB2 (Fig. S2). To capture the *ami* gene cluster, spheroplast cells of *S. cerevisiae* VL6-48 were transformed with the linearized *ami* pathway-specific capture vector and enzymatically fragmented genomic DNA together. Desired transformants with the captured *ami* gene cluster were selected on synthetic tryptophan dropout agar and identified by PCR. Direct cloning of the *ami* gene cluster was confirmed by restriction mapping to give pCAPB1-ami and pCAPB2-ami. The surfactin gene cluster *srf* was similarly captured following the same protocol to give pCAPB1-srf and pCAPB2-srf. For the *ami* gene cluster expression in *E. coli*, competent cells of *E. coli* BW25113 carrying pIJ790 and pCAPB1-ami were transformed with the linearized capture vector pCAPE to generate pCAPE-ami via λ-Red mediated recombination. The pCAPB1-ami, pCAPB2-ami and pCAPE-ami constructs were obtained and confirmed by restriction mapping after stable propagation through *E. coli* (Fig. 4, Fig. S5). More detailed information is provided in Supplemental Experimental Procedures.

**Heterologous expression of the ami gene cluster.** The construct pCAPB2-ami and its derivatives, which have 1.0-kb homology regions corresponding to the upstream and downstream regions of the *ami* gene *amyE*, were transferred to strain *B. subtilis* JH642 + *sfp* by natural competence transformation. Spectinomycin-resistant and PCR positive clones were routinely grown in LB broth containing spectinomycin (100 μg mL⁻¹) at 30°C overnight. A portion (1.0 mL) of the preculture was inoculated into 100 mL of LB broth and grown for 1 d at 30°C in a 250-mL flask with rotary shaking. For heterologous expression in *E. coli*, construct pCAPE-ami was introduced into BL21 (DE3) cells via electroporation. The positive clones were inoculated and confirmed by restriction mapping and then similarly cultured. The EtOAc extracts from the culture broth were analyzed by reversed-phase UPLC-MS. Detailed information, including the analytical conditions for UPLC-MS, genetic manipulation of the genes *amiA* and *amiB*, sample preparation for UPLC-MS analysis, and antimicrobial bioassay, are described in Supplemental Information.

1. Bachmann, B. O., Van Lanen, S. G. & Baltz, R. H. Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **41**, 175–184 (2014).
2. Cimermancic, P. et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).
3. Zhang, H. R., Bhogigraphan, B. A., Armando, J. & Pfeifer, B. A. Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.* **28**, 125–151 (2011).
31. Itoh, J.
32. Peypoux, F., Bonmatin, J. M. & Wallach, J. Recent trends in the biochemistry of
22. Herzner, A. M.
17. de Boer, A. S. & Diderichsen, B. On the safety of
25. Yuksel, S. & Hansen, J. N. Transfer of nisin gene cluster from
26. Zobel, S., Kumpfmuller, J., Sussmuth, R. D. & Schweder, T.
23. Liu, W. & Hansen, J. N. Conversion of
24. Tsuge, K.
18. Leuschner, R. G. K.
28. Yamanaka, K.
11. de la Cruz, L. A., O’Donnell, N. A. & Costa, C. C. L. Current progress in the identification and investigation of bioactive metabolites from bacteria.
21. Chatterjee, S. & Pramanik, S. K. Non-ribosomal peptide synthetases - current status and future perspectives.
27. Sina, S. & Hershko, A. The death of a peptidylglycine α-hydroxylating enzyme: the Pfam33 family.
29. Itoh, J. et al. New classes of amicoumacins from Bacillus subtilis.
30. Wagner, J. K., Marquis, K. A. & Rudner, D. Z. SirA enforces dyplodiosis by inhibiting the replication initiator DNAAs during spore formation in Bacillus subtilis. Mol. Microbiol. 73, 963–974 (2009).
31. Solomon, J. M., Magnussen, R., Srivastava, A. & Grossman, A. D. Convergent sensing pathways mediate response to two extracellular competence factors in bacillus subtilis. Genes. Dev. 11, 605–613 (1997).
32. Lama, A., et al. Response of methicillin-resistant Staphylococcus aureus to amicoumacin A. PLoS One 7, e34037 (2012).
33. Polikanov, Y. S. et al. Amicoumacin A inhibits translation by stabilizing mRNA interaction with the ribosome. Mol. Cell 56, 531–540 (2014).
34. Li, Y. X. et al. Five new amicoumacins isolated from a marine-derived bacterium Bacillus subtilis. Mar. Drugs 10, 319–328 (2012).
35. Pinchuk, I. V. et al. Amicoumacin antibiotic production and genetic diversity of Bacillus subtilis strains isolated from different habitats. Res. Microbiol. 153, 269–276 (2002).
36. Nakano, M. M., Marahiel, M. A. & Zuber, P. Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in Bacillus subtilis. J. Bacteriol. 170, 5662–5668 (1988).
37. Gust, B. et al. PCR-targeted streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiperpene soil odor geosmin. Proc. Natl. Acad. Sci. USA 100, 1541–1546 (2003).
38. Biau, Y. et al. In vivo evidence for a prodrug activation mechanism during colibactin maturation. ChemBioChem 14, 1194–1197 (2013).
39. Reimer, D. & Bode, H. B. A natural prodigur activation mechanism in the biosynthesis of nonribosomal peptides. Nat. Prod. Rep. 31, 154–159 (2014).
40. Reimer, D. et al. A natural prodigur activation mechanism in nonribosomal peptide synthesis. Nat. Chem. Biol. 7, 888–890 (2011).
41. Fujii, K. et al. Further application of advanced Marfey’s method for determination of the absolute configuration of primary amino compound. Tetrahedron Lett. 39, 2579–2582 (1998).
42. Kallifidas, D. & Brady, S. F. Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. Methods Enzymol. 517, 225–239 (2012).
43. Kim, J. H. et al. Cloning large natural product gene clusters from the environment: Piceing environmental DNA gene clusters back together with TAR. Biopolymers 93, 833–844 (2010).
44. Blin, K. et al. Antismash 2.0-a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res. 41, W204–W212 (2013).
45. Rausch, C. et al. Specificity prediction of adenylaylations domain in nonribosomal peptide synthetases (NRPSs) using transductive support vector machines (TSVSM). Nucleic Acids Res. 33, 5799–5808 (2005).

Acknowledgments
This study was generously supported by a grant (DY125-15-T-02) from China Ocean Mineral Resources Research and Development Association, award SA-C0840/UK-C0016 from the King Abdullah University of Science and Technology to P.Y.Q., and funding from the NIH (R18-GM085770) to B.S.M.

Author contributions
Y.L., K.Y., R.K., B.S.M. and P.Q. designed the experiments, Y.L., Z.L., K.Y., R.K., B.S.M. and P.Q. analyzed the data, and Y.L., K.Y., R.K., B.S.M. and P.Q. designed the experiments, Y.L., Z.L. and W.Z. executed the experiments, Y.L., Z.L., K.Y., Y.X., W.Z., H.V., R.K., B.S.M. and P.Q. analyzed the data, and Y.L., B.S.M. and P.Q. wrote the manuscript with input from all authors.

Additional information
Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Li, Y. et al. Directed natural product biosynthesis gene cluster capture and expression in the model bacterium Bacillus subtilis. Sci. Rep. 5, 9383; DOI:10.1038/srep09383 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/