

# Alternative Sigma Factor Over-Expression Enables Heterologous Expression of a Type II Polyketide Biosynthetic Pathway in *Escherichia coli*

David Cole Stevens<sup>1</sup>, Kyle R. Conway<sup>1</sup>, Nelson Pearce<sup>1</sup>, Luis Roberto Villegas-Peñaranda<sup>1</sup>, Anthony G. Garza<sup>2</sup>, Christopher N. Boddy<sup>1,3\*</sup>

**1** Department of Chemistry, University of Ottawa, Ottawa, Ontario, Canada, **2** Department of Biology, Syracuse University, Syracuse, New York, United States of America, **3** Department of Biology, University of Ottawa, Ottawa, Ontario, Canada

## Abstract

**Background:** Heterologous expression of bacterial biosynthetic gene clusters is currently an indispensable tool for characterizing biosynthetic pathways. Development of an effective, general heterologous expression system that can be applied to bioprospecting from metagenomic DNA will enable the discovery of a wealth of new natural products.

**Methodology:** We have developed a new *Escherichia coli*-based heterologous expression system for polyketide biosynthetic gene clusters. We have demonstrated the over-expression of the alternative sigma factor  $\sigma^{54}$  directly and positively regulates heterologous expression of the oxytetracycline biosynthetic gene cluster in *E. coli*. Bioinformatics analysis indicates that  $\sigma^{54}$  promoters are present in nearly 70% of polyketide and non-ribosomal peptide biosynthetic pathways.

**Conclusions:** We have demonstrated a new mechanism for heterologous expression of the oxytetracycline polyketide biosynthetic pathway, where high-level pleiotropic sigma factors from the heterologous host directly and positively regulate transcription of the non-native biosynthetic gene cluster. Our bioinformatics analysis is consistent with the hypothesis that heterologous expression mediated by the alternative sigma factor  $\sigma^{54}$  may be a viable method for the production of additional polyketide products.

**Citation:** Stevens DC, Conway KR, Pearce N, Villegas-Peñaranda LR, Garza AG, et al. (2013) Alternative Sigma Factor Over-Expression Enables Heterologous Expression of a Type II Polyketide Biosynthetic Pathway in *Escherichia coli*. PLoS ONE 8(5): e64858. doi:10.1371/journal.pone.0064858

**Editor:** Vasu D. Appanna, Laurentian University, Canada

**Received:** March 29, 2011; **Accepted:** April 22, 2013; **Published:** May 28, 2013

**Copyright:** © 2013 Stevens et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by NSERC, Ontario Ministry of Research and Innovation, and the University of Ottawa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: cboddy@uottawa.ca

## Introduction

Bacterial polyketides possess an enormous range of chemical diversity and biological function. Many polyketides such as tetracycline [1], epothilone [2], and rapamycin [3] have been developed into key clinical pharmaceuticals in a broad range of therapeutic areas [4]. Sequencing of bacterial genomes, especially those of major polyketide producers such as *Actinomycetes* and  $\delta$ -*proteobacteria*, have shown that there are many more polyketide biosynthetic pathways than polyketides isolated from standard cultivation techniques [5–7]. These genetically encoded polyketide natural products from cultivatable and uncultivable bacteria represent one of the greatest remaining untapped reservoirs of new natural product diversity. Methods to effectively access this diversity will have a major impact on drug discovery [8].

To access this untapped diversity of polyketide products, a general method for heterologous expression of these pathways is needed. The selection of a heterologous host is contingent upon its ability to efficiently transcribe non-native pathways, translate the often GC rich transcripts, and possess all the starter and extender units necessary for polyketide production [9]. Hosts highly related to the native polyketide-producing organism often meet these

requirements. Hosts such as *Streptomyces coelicolor* [10], *Streptomyces lividans* [11], and *Myxococcus xanthus* [12] have proved successful for heterologous production of a number of different polyketide products. Screening of multiple hosts related to the producing organism is often necessary to identify a successful heterologous host [13,14]. In addition, various culture conditions also need to be screened to identify conditions that produce the desired compounds [15]. The lack of a general heterologous expression system has made heterologous expression incompatible with screening bacterial genomic DNA libraries for new polyketide products.

*Escherichia coli* has a number of advantages that make it an appealing host for a heterologous expression system. The ease of genetic manipulation and culturing as compared to other heterologous hosts makes *E. coli* far more useful for bioprospecting from genomic DNA libraries. Codon usage as well as starter and extender unit availability have proven not to be obstacles for heterologous production in *E. coli* [16–18]. The principle impediment to the use of *E. coli* as a heterologous host is its inability to effectively transcribe heterologous pathways. In all examples of heterologous expression of polyketides in *E. coli*, all of the biosynthetic genes have been placed under the control of the

T7 promoter [16,17]. The native promoters present in these heterologous pathways are not sufficient to ensure expression of all the genes in a given pathway under standard *E. coli* culture conditions.

To convert *E. coli* into a general heterologous expression system, a mechanism for ensuring transcription of all the biosynthetic genes in a foreign pathway is required. We hypothesized that a general transcriptional regulator of polyketide biosynthesis should be present in bacteria because horizontal transfer of biosynthetic pathways is a key mechanism for their proliferation among species [19]. Strong evidence for substantial horizontal transfer of pathways can be seen in the large number of diverse *Streptomyces* strains which produce the tetracycline family of antibiotics, the presence of yersiniabactin-like gene clusters in diverse *Actinomycetes* [6], and the presence of highly related pederin, onnamide A and psymberin gene clusters in unidentified bacterial strains from the beetle *Paederus fuscipes* and the sponges *Theonella swinhoei* and *Psammocinia* aff. *bulbosa* respectively [20]. Because polyketide production is positively regulated in most bacteria by stress [21], we further hypothesized that alternative sigma factors downstream of the stringent response should positively regulate transcription of biosynthetic genes.

Herein we show that this new mechanism for heterologous expression of the oxytetracycline polyketide biosynthetic pathway, where a high-level pleiotropic sigma factor from the heterologous host is used to positively regulate transcription of the non-native biosynthetic gene cluster, is highly effective. We demonstrate that one of the six alternative sigma factors in *E. coli* can selectively, directly, and positively regulate heterologous expression from the *Streptomyces rimosus* oxytetracycline biosynthetic gene cluster [1] in *E. coli*. This is the first successful heterologous production of an aromatic polyketide in *E. coli* [22], demonstrating the utility of this method.

## Results

### Standard *E. coli* culture conditions do not enable heterologous expression

To test our hypothesis that an alternative sigma factor downstream of the stringent response could positively regulate transcription of polyketide biosynthetic pathways in *E. coli*, we investigated the ability of *E. coli* to heterologously express the oxytetracycline biosynthetic pathway from *S. rimosus* [1]. Oxytetracycline is produced by a 32-kb type II polyketide synthase gene cluster consisting of 21 genes. Thirteen of these genes encode proteins required for the biosynthesis of oxytetracycline [23] and thus must be transcribed for successful heterologous production of oxytetracycline. The small intergenic spaces between open reading frames and the changes in direction of transcription in the oxytetracycline gene cluster suggest that the entire pathway is expressed as, at minimum, five putative operons with at least one biosynthetic gene on each transcript (Figure 1a).

To determine if *E. coli* could produce oxytetracycline under standard culture conditions, *E. coli* BAP1 [16] was transformed with the *S. rimosus* oxytetracycline pathway [12] and cultured in rich liquid media. BAP1 was used because it possesses a chromosomal insertion of the phosphopantetheinyl transferase from *Bacillus subtilis*, *sfp*, which ensures that polyketide synthase acyl carrier proteins (ACPs) are post-translationally modified [16]. LC-MS/MS analysis of the organic extracts from the culture media did not show any detectable oxytetracycline in the culture broth (see Figure S3). LC-MS/MS analysis was performed using multiple reaction monitoring (MRM) mode with a lower limit of detection of 100  $\mu\text{g/L}$ . These results indicate that *E. coli*, under

standard culture conditions, cannot heterologously produce oxytetracycline from the native *S. rimosus* gene cluster.

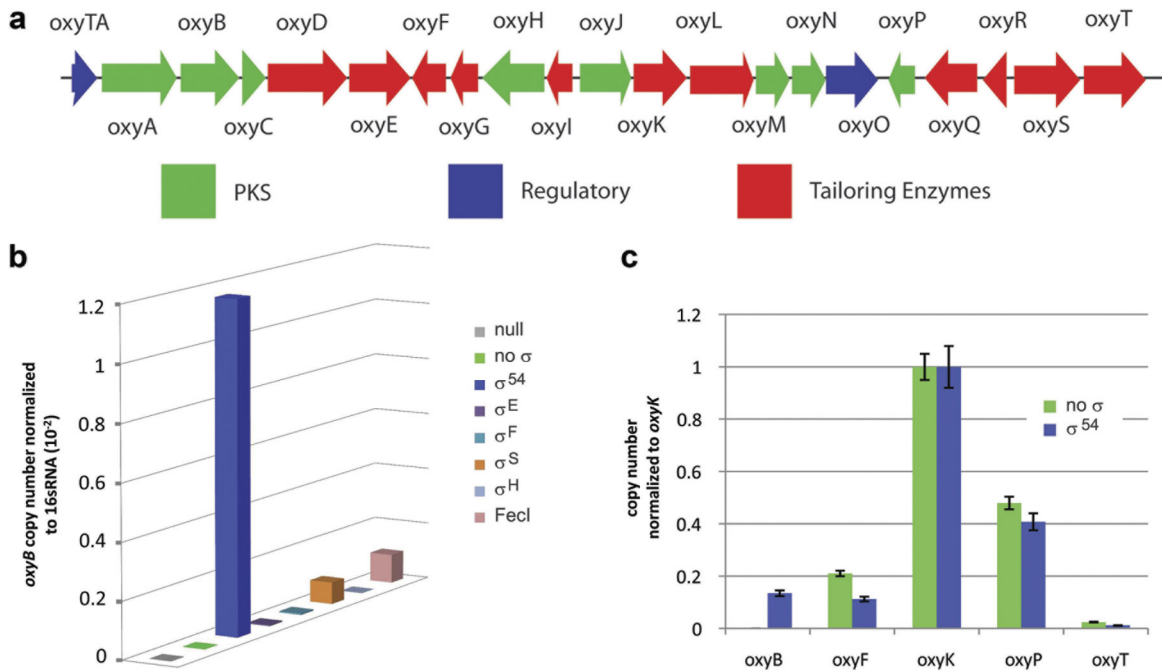
To determine if transcription of the genes in the oxytetracycline gene cluster was limiting production of oxytetracycline in *E. coli*, we quantified mRNA levels using quantitative PCR (qPCR) for five genes in the pathway. The five genes *oxyB*, *oxyF*, *oxyK*, *oxyP*, and *oxyT*, one per putative operon, were selected. *oxyB* encodes the  $\beta$ -ketosynthase, *oxyF* and *oxyT* encode methyltransferases and *oxyK* encodes an aromatase all of which are required for oxytetracycline biosynthesis [23]. *oxyP* encodes a malonyl-ACP transferase, which may not be required for biosynthesis of oxytetracycline [23]. Transcripts for four of the genes, *oxyF*, *oxyK*, *oxyP* and *oxyT* could be easily detected and quantified (Figure 1c). Transcript levels for *oxyB* were indistinguishable from negative control samples that lacked the oxytetracycline gene cluster, indicating that *oxyB* was not transcribed (Figure 1b, Figure S1, Table S1). As OxyB is required for generation of the polyketide backbone of oxytetracycline, the lack of transcription of this gene is sufficient to account for the inability of *E. coli* to produce oxytetracycline under standard culture conditions. These results, in conjunction with published work [16,17], confirm that transcription of heterologous polyketide pathways in *E. coli* limits the ability of the heterologous host to produce polyketide products when using the native promoters.

### Over-expression of alternative sigma factors $\sigma^{54}$ , $\sigma^S$ , and FecI positively regulate transcription of *oxyB* in *E. coli*

To test our hypothesis that an alternative sigma factor downstream of the stringent response could positively regulate transcription of polyketide biosynthetic pathways in *E. coli*, we investigated the impact of alternative sigma factor over-expression on transcription from the oxytetracycline gene cluster. In their native hosts, polyketide biosynthetic gene clusters are principally regulated by one or more pathway specific regulators, which can respond to a diversity of stimuli [21,24]. The most general observed positive regulator of polyketide biosynthesis is nutrient limitation [21], which initiates the stringent response in bacteria. During stringent response the levels of the pleiotropic regulators phosphorylated guanosine nucleotides, ppGpp and pppGpp, increase, leading to an increase in alternative sigma factor-mediated transcription [24]. As increases in (p)ppGpp have been shown to correlate with polyketide production in diverse bacterial species [25,26], we hypothesized that this positive regulation could be mediated by alternative sigma factors.

Sigma factors control the specificity of gene transcription by binding to RNA polymerase (RNAP) and recruiting it to promoter sequences upstream of the gene to be transcribed [27]. In addition to the household sigma factor  $\sigma^{70}$ , *E. coli* has six alternative sigma factors,  $\sigma^{54}$  which is involved in nitrogen assimilation,  $\sigma^H$  which controls heat shock promoters,  $\sigma^S$  which controls stationary phase promoters,  $\sigma^F$  which controls flagellum related genes,  $\sigma^E$  controls response to extracytoplasmic stress, and FecI which is involved in iron transport [27]. Over-expression of sigma factors has been shown to increase expression of genes in their regulon in many bacteria, including *E. coli* [28–30]. If an alternative sigma factor is a positive regulator of polyketide biosynthesis, over-expression of that sigma factor will increase transcription from genes in the oxytetracycline gene cluster.

To determine if  $\sigma^{54}$ ,  $\sigma^H$ ,  $\sigma^S$ ,  $\sigma^F$ ,  $\sigma^E$ , and FecI could positively regulate polyketide biosynthesis in *E. coli*, we quantified mRNA levels using quantitative PCR for *oxyB*. Six *E. coli* BAP1 strains possessing the oxytetracycline biosynthetic gene cluster and over-expressing either  $\sigma^{54}$ ,  $\sigma^H$ ,  $\sigma^S$ ,  $\sigma^F$ ,  $\sigma^E$ , or FecI were investigated. No *oxyB* transcripts could be detected when  $\sigma^H$ ,  $\sigma^F$ , and  $\sigma^E$  were



**Figure 1. SYBR-green based qPCR analysis shows that transcription limits heterologous production of oxytetracycline in *E. coli*.** (a) The 32 kb oxytetracycline biosynthetic gene cluster is shown. Five putative operons, *oxyABCDE*, *oxyIHGF*, *oxyJKLMNO*, *oxyRQP*, and *oxyST* are predicted for this gene cluster (b) qPCR analysis shows that over-expression of the alternative sigma factors  $\sigma^{54}$ ,  $\sigma^S$  and FecI enable detectable levels of the *oxyB* transcript to be produced. Over-expression of no sigma factor,  $\sigma^E$ ,  $\sigma^F$  and  $\sigma^H$  do not lead to detectable levels of the *oxyB* transcript. (c) qPCR analysis shows that over-expression of the alternative sigma factor  $\sigma^{54}$  lead to detectable levels of transcripts for all five putative operons in the oxytetracycline biosynthetic pathway. In the absence of  $\sigma^{54}$  over-expression, the *oxyB* transcript cannot be detected. See also Table S1. No *oxyB*, *oxyF*, *oxyK*, *oxyP*, *oxyT* transcripts were observed in the null strains which lacked the oxytetracycline gene cluster. (Figures S1, S2). doi:10.1371/journal.pone.0064858.g001

over-expressed. Over-expression of  $\sigma^S$  and FecI led to detectable but low levels of *oxyB* transcription. When  $\sigma^{54}$  was over-expressed substantial levels of *oxyB* transcripts could be detected (Figure 1b, Figure S2, Table S1). Quantification of mRNA levels for *oxyB*, *oxyF*, *oxyK*, *oxyP* and *oxyT* in *E. coli* BAP1 possessing the oxytetracycline biosynthetic gene cluster and over-expressing *E. coli*  $\sigma^{54}$  showed that we were able to detect all five transcripts (Figure 1c). These data demonstrate that all of the putative transcripts required for oxytetracycline are present when  $\sigma^{54}$  is over-expressed.

### Over-expression of $\sigma^{54}$ enables oxytetracycline production in *E. coli*

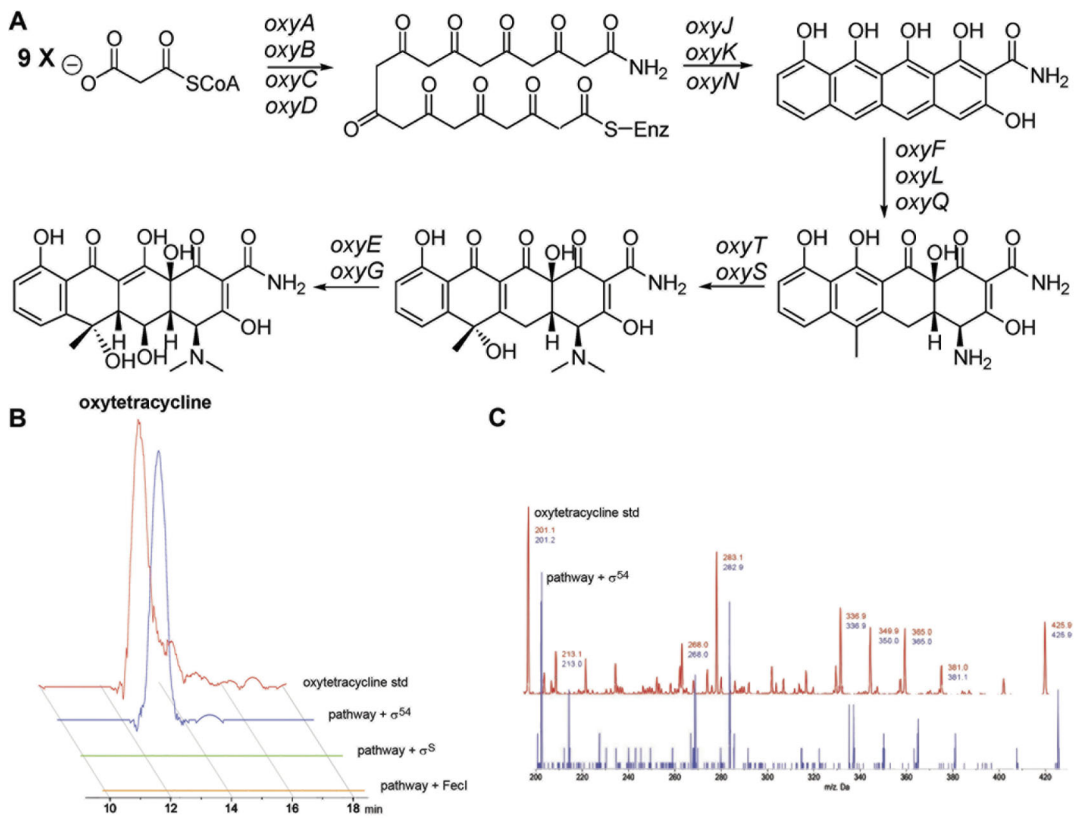
If transcription limits heterologous production of polyketides in *E. coli*, then over-expression of  $\sigma^{54}$  in the presence of the oxytetracycline gene cluster is predicted to lead to the production of oxytetracycline. To test this hypothesis we over-expressed  $\sigma^{54}$  in *E. coli* BAP1 containing the oxytetracycline gene cluster and analyzed the organic extracts with LC-MS/MS to determine if oxytetracycline was detectable. LC-MS/MS analysis using the MS<sup>2</sup> scan mode clearly identified the presence of oxytetracycline (Figure 2b). The LC retention time and MS<sup>2</sup> spectrum of our heterologously produced oxytetracycline was indistinguishable from authentic oxytetracycline standards (Figure 2c). Using MRM mode and a standard curve generated from authentic oxytetracycline, the titer of heterologously expressed oxytetracycline was determined to be  $2.0 \pm 0.1$  mg per L of culture broth. These results demonstrate that when  $\sigma^{54}$  is over-expressed in *E. coli* containing the oxytetracycline gene cluster, all the biosynthetic genes are transcribed and translated,

all the necessary proteins are functional *in vivo*, and the required starter and extender units are available, enabling oxytetracycline production.

Because over-expression of  $\sigma^S$  and FecI also produced detectable *oxyB* transcripts, it is possible that these alternative sigma factors may also enable heterologous production of oxytetracycline. We therefore investigated the ability of *E. coli* BAP1 cultures over-expressing  $\sigma^S$  and FecI to heterologously produce oxytetracycline. LC-MS/MS analysis showed that no detectable oxytetracycline was present in these culture broths (Figure 2b). As expected from our transcription data, over-expression of  $\sigma^H$ ,  $\sigma^F$ , and  $\sigma^E$  did not lead to heterologous production of detectable levels of oxytetracycline (Figure S3). These data demonstrate that only the alternative sigma factor,  $\sigma^{54}$ , can positively regulate transcription of the oxytetracycline biosynthetic pathway, leading to heterologous production of oxytetracycline in *E. coli*.

### Antibiotic activity of oxytetracycline does not limit heterologous expression

Oxytetracycline, like tetracycline, is a broad spectrum antibiotic that binds to the 30S subunit of the ribosome, inhibiting protein synthesis [31]. The minimum inhibitor concentration of *E. coli* strains to oxytetracycline can range from 0.5–16  $\mu\text{g}/\text{mL}$  with a typical value of 8  $\mu\text{g}/\text{mL}$  [32]. Because the oxytetracycline gene cluster used in this study lacked a resistance gene, we were concerned that the activity of oxytetracycline at our maximum titer of 2  $\mu\text{g}/\text{mL}$  may limit heterologous expression. To address this concern, we generated a tetracycline resistance vector pLRVP09 based on pBR322 and transformed this into a BAP1



**Figure 2. Over-expression of  $\sigma^{54}$  enables *E. coli* to heterologously produce oxytetracycline.** (a) The enzymatic pathway responsible for the biosynthesis of oxytetracycline. (b) ESI-LC-MS/MS analysis of an oxytetracycline standard and the organic extracts from *E. coli* cultures containing the oxytetracycline gene cluster and over-expressing  $\sigma^{54}$ ,  $\sigma^S$ , or Fecl. These traces show the ion extraction data from the Q3 scan of MS/MS experiments. Q1 was used to select the  $[MH]^+$  ion for oxytetracycline ( $m/z = 461$ ). Ion extractions were performed for the oxytetracycline fragment  $m/z = 283$  from the Q3 scan. Signals with a peak width of less than 0.1 s were regarded as noise and removed with the noise filter application. (c) The MS<sup>2</sup> spectrum of the  $m/z = 461$  peak for the oxytetracycline authentic standard and heterologously produced oxytetracycline. See also Figure S3. doi:10.1371/journal.pone.0064858.g002

strain with the oxytetracycline gene cluster and a  $\sigma^{54}$  expression vector. Heterologous production of oxytetracycline from this strain was identical to strains lacking the pBR322 derived TetC resistance determinant. This data indicates that the antibiotic activity of oxytetracycline does not limit the heterologous production of oxytetracycline to 2 mg/L.

### $\sigma^{54}$ over-expression does not lead to substantial accumulation of the oxytetracycline biosynthetic proteins

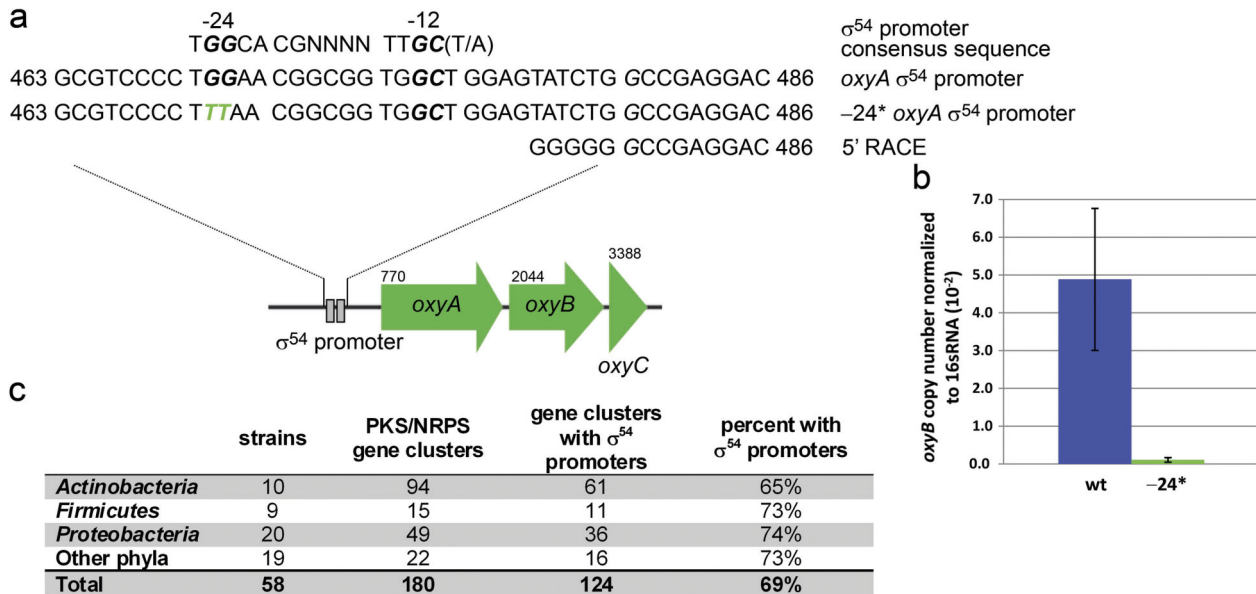
While most classes of bacterial natural products have been heterologously expressed in *E. coli*, type II polyketide biosynthetic pathways have not been [22]. Even when key type II biosynthetic genes, such as those encoding the ketosynthase (KS) or chain length factor (CLF), are placed under the control of known *E. coli* promoters, it has not been possible to produce type II polyketides, due predominantly to the production of insoluble KS-CLF heterodimer [22,33]. Because  $\sigma^{54}$  over-expression leads to production of oxytetracycline, some soluble, functional KS (OxyA) and CLF (OxyB) must be expressed. To evaluate if  $\sigma^{54}$  over-expression constituted an effective method to produce soluble KS-CLF heterodimer, we examined the soluble fraction of oxytetracycline producing strains by SDS-PAGE (Figure S4). While a band consistent with  $\sigma^{54}$  (54 kDa) could be easily detected, there was no evidence for high-level production of OxyA (45 kDa) or OxyB (44 kDa). Thus while  $\sigma^{54}$  over-expression must lead to some

functional soluble KS-CLF heterodimer, there is not substantial accumulation of these key proteins.

### $\sigma^{54}$ directly regulates heterologous production in *E. coli*

$\sigma^{54}$  can positively regulate transcription directly or indirectly. For direct regulation, the  $\sigma^{54}$ -RNAP complex binds to a  $\sigma^{54}$  consensus promoter upstream of the operon under  $\sigma^{54}$  control. The  $\sigma^{54}$  consensus promoter is unique among bacterial promoters with highly conserved residues -12 and -24 from the transcriptional start site (TGGCACG-N4-TTGC(T/A)) [34] and can reliably be identified from sequence data. If  $\sigma^{54}$  is directly regulating transcription of *oxyB*, a  $\sigma^{54}$  promoter consensus sequence must be present upstream of the operon containing *oxyB*. To determine if  $\sigma^{54}$  promoters were present in the oxytetracycline biosynthetic pathway the entire gene cluster was examined for  $\sigma^{54}$  promoter consensus sequences using the web-based tool PromScan (<http://molbiol-tools.ca/promscan/>) [35]. A high scoring consensus promoter was located upstream of the operon containing *oxyB* in a location that would allow direct transcription of the putative operon (Figure 3a, Table S2). The location and consensus sequence of the identified promoter provides strong evidence for direct regulation by  $\sigma^{54}$  of the heterologous production of oxytetracycline in *E. coli*.

To demonstrate that the predicted  $\sigma^{54}$  promoter is functional and involved in the transcriptional regulation of *oxyB*, the highly conserved GG -24 from the transcriptional start site was mutated



**Figure 3. Putative  $\sigma^{54}$  promoters are found in the majority of polyketide and non-ribosomal peptide biosynthetic gene clusters.** (a) The *oxyABCDEF* operon from the *S. rimosus* oxytetracycline biosynthetic pathways contains a putative  $\sigma^{54}$  promoter. The promoter shows high homology to the  $\sigma^{54}$  promoter consensus sequence especially at the key -12 and -24 positions. A mutation of the highly conserved GG residues at -24 (-24\*) and 5' RACE were used to confirm that this promoter is responsible for direct, positive transcriptional regulation of the *oxyB* gene by  $\sigma^{54}$  over-expression. (b) Transcription of the *oxyABCDEF* operon is directly controlled by the  $\sigma^{54}$  promoter upstream of *oxyA*. Mutation of the highly conserved GG residues -24 from the transcriptional start site to TT leads to a greater than 40 fold decrease in *oxyB* transcript levels as determined by qPCR. qPCR data was obtained from BAP1 transformed with pDCS11 and either pDCS61 or pDCS62. (c) Results of a bioinformatics analysis of 58 bacterial genomes containing 180 polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) biosynthetic gene clusters show that the majority of gene clusters contain putative  $\sigma^{54}$  promoters. It is particularly intriguing that *Actinobacteria* possess gene clusters with  $\sigma^{54}$  promoters since they lack the gene encoding  $\sigma^{54}$ . See also Tables S2, S3, S4. doi:10.1371/journal.pone.0064858.g003

to TT. The -24 region plays a major role in anchoring the  $\sigma^{54}$ -RNAP complex to the promoter and mutation of the conserved GG is known to decrease transcription of genes under  $\sigma^{54}$  control by ten to one hundred-fold [36,37]. qPCR analysis showed that *oxyB* transcript levels from the TT mutant  $\sigma^{54}$  promoter were greater than forty-fold lower as compared to the wild type (Figure 3b, Table S1). To identify promoter elements involved in transcription of the operon containing *oxyB*, we performed 5' Rapid Amplification of cDNA Ends (5' RACE). Our 5' RACE data identified a transcriptional start site consistent with the proposed  $\sigma^{54}$  promoter, confirming that this promoter is functional and that *oxyB* transcription is directly regulated by  $\sigma^{54}$ .

### $\sigma^{54}$ promoters are common in polyketide and non-ribosomal peptide biosynthetic pathways

If  $\sigma^{54}$  promoters are commonly found in polyketide and non-ribosomal peptide biosynthetic pathways,  $\sigma^{54}$  over-expression may be a general route to heterologous expression of polyketides and non-ribosomal peptides. To evaluate this possibility, we examined the genomes of 58 sequenced bacteria for polyketide and non-ribosomal peptide biosynthetic gene clusters and  $\sigma^{54}$  promoters. The 58 species included major secondary metabolite producers such as 10 *Actinobacteria*, 20 *Proteobacteria*, 9 *Firmicutes*, as well as 19 species from diverse phyla. This selection of species contains representative examples of most of the sequenced bacterial phylogenetic diversity.

Manual examination of all genes annotated as polyketide synthase, 3-oxoacyl acyl carrier protein synthase, and non-ribosomal peptide synthetase (COG3321, COG0304, and

COG1020 respectively) [38] and their adjacent genes, identified 180 polyketide and non-ribosomal peptide biosynthetic gene clusters (Figure 3c, Table S3).

Each bacterial genome was analyzed to identify all putative  $\sigma^{54}$  promoters. To identify  $\sigma^{54}$  promoters, a positional weighted matrix (PWM) describing the  $\sigma^{54}$  promoter consensus sequence [36] was aligned with all sites on the forward and reverse strands of the genome. The fit at each position was scored using the published algorithm [35] from the bioinformatics tool PromScan. Sequences with a fit scoring 75 or greater and on the same strand, less than 500 bases upstream of a start codon were identified as putative  $\sigma^{54}$  promoters. All experimentally validated  $\sigma^{54}$  promoters in the *E. coli* genome had scores ranging from the 80 s to high 90 s [30]. In GC rich organisms such as *M. xanthus* a number of previously identified  $\sigma^{54}$  promoters had scores in the low to mid 70 s [39,40]. Based on these data, a score of 75 is expected to provide the required low level of false negatives across a broad range of genomes.

To evaluate the level of false positives, a sequence randomized *E. coli* genome was analyzed [41]. 25.6% of genes in the native *E. coli* genome and 18.4% of the genes in the sequence randomized genome were predicted to have  $\sigma^{54}$  promoters. 18.4% thus provides a baseline for the level false positive predictions. This high false positive rate is typical of PWM based promoter prediction tools [42,43].

As false positives are not expected to be conserved across species, a comparative genomic approach that examines the predicted promoters across a wide range of bacterial genomes should decrease the false positive rate and increases the specificity [44]. Our data shows that >50% of the bacterial genomes possessing a *glmA* ortholog (COG0174), the archetypal gene

under direct  $\sigma^{54}$  transcriptional control, had one or more predicted  $\sigma^{54}$  promoters upstream of this gene. In comparison only 10% of the genomes examined had predicted  $\sigma^{54}$  promoters upstream of the well-characterized non- $\sigma^{54}$  regulated ribosomal proteins encoding genes (COG0093, COG0049) [45], transcription elongation factor encoding genes (COG0264, COG0231) [46,47], and the ATP synthase encoding genes (COG005, COG0224) [48]. Thus while individual promoter predictions have a high false positive rate, correlating promoter predictions across multiple species for individual orthologs substantially decreases the false positive predictions and increases the ability to accurately predict orthologs with  $\sigma^{54}$  promoters.

Of the 180 polyketide and non-ribosomal peptide biosynthetic gene clusters identified, 124 (69%) contained one or more  $\sigma^{54}$  promoters appropriately positioned to regulate transcription (Table S3, Table S4). 24 gene clusters possessed a single  $\sigma^{54}$  promoter, however the vast majority had two or more, with some pathways possessing up to two dozen  $\sigma^{54}$  promoters. Additionally 75 of the 124 (60%) clusters had a putative  $\sigma^{54}$  promoter appropriately placed to directly regulate either a polyketide synthase or a non-ribosomal peptide synthetase encoding ortholog. These results demonstrate that a majority (69%) of polyketide and non-ribosomal peptide biosynthetic gene clusters possess putative  $\sigma^{54}$  promoters and that these promoters are appropriately positioned to regulate transcription of at least one operon in these gene clusters. These results are consistent with the hypothesis that  $\sigma^{54}$  over-expression may be a general method for ensuring transcription of polyketide and non-ribosomal peptide biosynthetic gene clusters in heterologous hosts, such as *E. coli*.

## Discussion

Heterologous expression of polyketide and non-ribosomal peptide biosynthetic gene clusters will play a major role in the discovery of new natural products from metagenomic and environmental DNA samples. Currently no general heterologous expression method appropriate for screening DNA libraries exists. Herein we describe a new mechanism for heterologous expression of a polyketide biosynthetic pathway, where a high-level, pleiotropic alternative sigma factor from the heterologous host positively regulates transcription of the biosynthetic gene cluster. In contrast, known methods for heterologous expression rely on either replacing each native promoter with known, well-characterized promoter from the heterologous host, such as the T7 promoter in *E. coli* [16–18], or rely on the heterologous host to constitutively express each gene from the native promoters [10–15]. Our approach, which actively induces transcription of gene clusters by over-expression of alternative sigma factors, may provide a general solution to the heterologous expression problem that is compatible with screening DNA libraries.

Our results demonstrate that over-expression of the sigma factor,  $\sigma^{54}$ , enables efficient heterologous expression of oxytetracycline biosynthetic gene cluster in *E. coli*. Using qPCR, we have demonstrated that no transcript is detectable under standard culture conditions for the key oxytetracycline biosynthetic gene *oxyB*. LC-MS/MS analysis of these cultures showed no detectable oxytetracycline production in the absence of detectable *oxyB* transcripts. Over-expression of  $\sigma^{54}$  positively regulates transcription of *oxyB*, generating detectable levels of *oxyB* transcripts, enabling production of oxytetracycline. Bioinformatics analysis of the oxytetracycline biosynthetic gene cluster revealed a  $\sigma^{54}$  promoter sequence appropriately positioned to directly regulate *oxyB* transcription. Site directed mutagenesis and 5' RACE strongly support that this promoter is functional. Further

bioinformatic analysis of 180 polyketide and non-ribosomal peptide biosynthetic pathways shows that the majority of these pathways possess putative  $\sigma^{54}$  promoter sequences (69%), suggesting that  $\sigma^{54}$ -mediated heterologous expression may be a general phenomenon.

Substantial evidence exists to support the hypothesis that over-expression of alternative sigma factors, such as  $\sigma^{54}$ , can positively regulate diverse polyketide biosynthetic pathways. Nutrient limitation, which is one of the only known general regulators of polyketide biosynthesis [21], initiates the stringent response in bacteria. During stringent response the levels of the global bacterial regulator (p)ppGpp increases. Increases in (p)ppGpp have been shown to correlate with polyketide production in diverse bacterial species [25,26]. For example *relA*, the gene encoding the (p)ppGpp synthetase, in *S. coelicolor* is required to trigger polyketide biosynthesis<sup>26</sup>. (p)ppGpp disrupts the interaction between the RNA polymerase (RNAP) and the principle sigma factor,  $\sigma^{70}$ , enabling RNAP to interact with alternative sigma factors, which increases alternative sigma factor-mediated transcription [24]. Thus nutritional limitation and (p)ppGpp exert their regulatory effects by increasing alternative sigma factor-mediated transcription, suggesting that polyketide biosynthetic pathways may be under either direct or indirect alternative sigma factor transcriptional control.

We investigated the effect of over-expression of all the alternative sigma factors from *E. coli* on the transcription of *oxyB* and on the production of oxytetracycline. Only  $\sigma^{54}$  over-expression led to both detectable levels of the *oxyB* transcript and detectable levels of oxytetracycline production.  $\sigma^{54}$  is known to play a major role in response to nitrogen limitation in most bacteria [27] and sporulation in *delta-proteobacteria* [49]. Interestingly, both nitrogen limitation and sporulation correlate with polyketide production [21,25]. While this suggests that  $\sigma^{54}$  may play a role in regulation of polyketide biosynthesis in some bacteria, substantial work in native hosts is required to test this hypothesis.

A unique aspect of  $\sigma^{54}$ -mediated transcription is that it requires co-activation to initiate transcription [50]. Enhancer binding proteins (EBPs) allow  $\sigma^{54}$ -loaded RNA polymerase to form a transcriptionally active open complex. EBPs contain a DNA binding domain that interacts with a specific DNA sequence called the upstream activating sequence (UAS), directing the EBP to a specific  $\sigma^{54}$  promoter. The ATPase domain of the EBP then catalyzes ATP hydrolysis, leading to open complex formation and transcription. Without this co-activation step, transcription cannot occur. High levels of EBPs have been shown *in vivo* and *in vitro* to enable formation of the active open complex independent of EBP binding to the UAS [51,52].

As our data supports direct  $\sigma^{54}$ -mediated transcription of the operon containing *oxyB* in the heterologous host *E. coli*, co-activation must be occurring.  $\sigma^{54}$  over-expression has been shown to positively regulate many of the twelve EBPs native to *E. coli* [53], including *glnG*, *zraR*, and *fhfA* [30]. As no EBPs are present in the oxytetracycline gene cluster (no open reading frames in the oxytetracycline gene cluster possess the key AAA+ and helix-turn-helix domains found in EBPs [53]), a potential hypothesis for co-activation in our heterologous transcription is the UAS-independent co-activation by a positively regulated native *E. coli* EBP.

Over-expression of  $\sigma^S$  and FecI showed an increase in *oxyB* transcription but did not lead to detectable levels of oxytetracycline production. *oxyB* transcript levels were at least 10 fold lower for  $\sigma^S$  and FecI over-expression than when compared to  $\sigma^{54}$  over-expression. Presumably the low levels of the *oxyB* transcript were insufficient to generate enough OxyB to produce detectable levels

of oxytetracycline. A plausible explanation for the transcription of *oxyB* during  $\sigma^S$  and FecI over-expression is cross-talk between these alternative sigma factors and  $\sigma^{54}$ , leading to low levels of  $\sigma^{54}$ -mediated transcription. This is supported by the observation that  $\sigma^{54}$ ,  $\sigma^S$ , and FecI are known to be linked by the polyamine response [54,55].

The identification of a functional  $\sigma^{54}$  promoter in the *S. rimosus* oxytetracycline biosynthetic gene cluster and 236 putative  $\sigma^{54}$  promoters in 61 gene clusters from seven different actinobacterial genomes is highly unexpected (Table S3). *Actinobacteria* do not contain a gene encoding  $\sigma^{54}$ . Three hypotheses could account for this observation. These promoters are non-functional in *Actinobacteria* and are remnants of horizontal biosynthetic pathway transfer from non-*Actinobacteria* where these  $\sigma^{54}$  promoters are functional. Alternatively, a functional equivalent of  $\sigma^{54}$  could be present. Extensive research efforts into regulation of polyketide biosynthesis in *Streptomyces* has not uncovered a functional equivalent to  $\sigma^{54}$ , however the vast majority of these studies have been carried out in *S. coelicolor*, the only *Streptomyces* in our genome analysis to contain no  $\sigma^{54}$  promoters in any of its 10 polyketide and non-ribosomal peptide biosynthetic gene clusters (Table S3). A third hypothesis is that these promoters represent false positives from the bioinformatics-based  $\sigma^{54}$  promoter prediction. While the prediction of promoters for orthologs across a wide range of bacterial species, as is done in this study, can decrease the false positive rate associated with PWM-based analyses and improve the selectivity of genome wide predictions, experimentation is ultimately required to evaluate the functionality of individual promoters. Understanding the origins and roles of these putative  $\sigma^{54}$  promoter sequences thus remains a wide-open question.

We have developed a new *E. coli*-based heterologous expression system for oxytetracycline polyketide biosynthetic gene clusters. We have demonstrated the over-expression of the alternative sigma factor  $\sigma^{54}$  directly and positively regulates heterologous expression of the oxytetracycline biosynthetic gene cluster in *E. coli*. Bioinformatics analysis indicates that  $\sigma^{54}$  promoters are present in nearly 70% of polyketide and non-ribosomal peptide biosynthetic pathways suggesting that  $\sigma^{54}$ -mediated heterologous expression may be an effective, general approach. If sufficiently general, this approach may facilitate heterologous expression of new polyketides from metagenomic and environmental DNA samples.

This study also opens the door to characterizing and engineering polyketide biosynthetic pathways in *E. coli*. The vast majority of genetic experiments, such as deletion or complementation experiments, used to characterize these biosynthetic pathways [22,56–61] have been carried out in either native producing organisms or in *Streptomyces* heterologous hosts. By developing an *E. coli*-based heterologous expression system, these experiments can be carried out in the highly genetically malleable *E. coli*, simplifying these studies.

This study represents the first example of successfully expressing functional type II polyketide synthases in *E. coli* [22,33]. Type II polyketide synthases have been extremely challenging to produce in *E. coli*, even when placed under the control of known *E. coli* promoters. A major hurdle to production of type II PKS in *E. coli* has been the ability to generate soluble KS-CLF dimer, an essential part of the polyketide biosynthetic machinery. When expressed as standalone or fusion proteins, KS and CLF have always formed insoluble inclusion bodies [22,33]. The prevailing hypothesis for this result is an incompatibility between the rates of protein synthesis, subunit folding, and heterodimerization [33]. Our work shows that while it is possible to generate soluble functional KS and CLF from the oxytetracycline pathway, the

KS-CLF-heterodimer is expressed at very low levels, limiting this tool's utility as a recombinant protein expression system. Our results will however be of use for in vivo characterization of aromatic polyketide backbone construction and the ensuing tailoring chemistries.

Finally, it is not unreasonable to suggest that over-expression of other sigma factors, such as  $\sigma^H$ ,  $\sigma^S$  or FecI, may positively regulate transcription of other biosynthetic gene clusters, potentially expanding the scope of alternative sigma factor-mediated heterologous expression.

## Materials and Methods

### General

*Escherichia coli* XL1 Blue (Stratagene) and *E. coli* TOP10 (Invitrogen) cells were used for routine cloning and plasmid preparations. *E. coli* BAP1 was used in all heterologous production experiments. *E. coli* cells were grown in LB media (Fisher) supplemented with the appropriate antibiotics when necessary. Oxytetracycline standard and IPTG were purchased from Sigma-Aldrich. Antibiotics and media components were purchased from Fisher. All strains were chemically or electronically transformed using standard transformation protocols.

### Plasmid Construction

PrimeSTAR HS polymerase and the Thermocycler Mastercycler personal were used for all PCR reactions. Pfu Ultra II polymerase and the Thermocycler Mastercycler personal were used for mutagenesis reactions. Primers can be found in the expanded experimental procedures in the supplementary information (File S1). *E. coli* MG1655 genomic DNA was used as the template. Each sigma factor was cloned into pCR-Blunt and confirmed by sequencing. Genes were sub-cloned using NdeI and EcoRI restriction sites into pET28b or pKH22 (a pET21c derivative with an AvrII site engineered after the native EcoRI site) [62], except for *rpoE* and *rpoD*, where HindIII instead of EcoRI and NheI instead of NdeI were used to due presence of native restriction sites. To construct pDCS61 the *oxyTAIABCD* cassette was amplified from pMRH08 [12] using forward primer 5' – TAATACGACTCACTATAGGG – 3' and the reverse primer 5' – CAGTGAATTCTCATAGCTCCAGGCTG – 3'. The *oxyTAIABCD* cassette was then digested with EcoRI and inserted into pET28b. Site directed mutagenesis was then performed on pDCS61 using forward primer 5' – CCTGCGTCCCCTAAAACGGCGGTGGC – 3' and the reverse primer 5' – GCCACCGCCGTTTATAGGGGACGCAGG – 3' following the QuikChange mutagenesis protocol to construct pDCS62. The mutant was confirmed by sequencing. To construct pLRVP09, pBR322 was digested with ScaI and SspI to remove the  $\beta$ -lactamase marker and ligated closed. Plasmid names and descriptions can be found in the expanded experimental procedures in the supplementary information (File S1).

### qPCR Analysis

Strains examined were BAP1+pDCS11, BAP1+pMRH08, BAP1+pMRH08+either pDCS11, pDCS57, pDCS58, pDCS59, pNDP6, or pNDP7, and BAP1+pDCS11+either pDCS61 or pDCS62. All strains were grown in 25 mL LB in shake flasks at 37°C. At O.D.<sub>600</sub> = 0.4, cultures were induced with 1 mM IPTG and incubated at 20°C for 24 hours. Cells were harvested for RNA isolation. In concert with RNA isolation the broth from each culture was extracted and analysed for oxytetracycline production as described below. The SV Total RNA Isolation System was used to isolate RNA. cDNA was generated from RNA using AMV

Reverse Transcriptase. qPCR experiments were performed in triplicate for each strain with a 10-fold dilution series of the 16 s RNA acting as an internal standard. Primers were designed to provide 180–210 bp amplicons of the *rpoN*, *oxyB*, *oxyF*, *oxyK*, *oxyP*, and *oxyT* transcripts. qPCR primer sequences can be found in the expanded experimental procedures in the supplemental information (File S1). The qPCR reactions contained 1.0  $\mu$ M of each primer, 12.5  $\mu$ L Absolute SYBR Green QPCR mix, 0.5  $\mu$ L prepared cDNA, and 10.5  $\mu$ L dH<sub>2</sub>O to give a total reaction volume of 20  $\mu$ L. A Mx3000 qPCR Thermocycler was used with the following conditions: 1 cycle 95°C for 15 min, 40 cycles of 95°C for 15 s, 55°C for 30 s, and 72°C for 30 s, followed by denaturation for 1 cycle at 95°C for 1 min, 55°C for 30 s, and 95°C for 30 s. Standard curve R<sup>2</sup> values and amplification efficiency values ranged from 0.991–1.0 and 90.7%→100% respectively. The amplification efficiency was calculated using the formula  $A = 10^{(-1/\text{slope})}$  in which the slope was calculated by regression analysis obtained from the C<sub>t</sub> values versus calculating log number of cells in serial dilutions (See Table S1).

### Oxytetracycline Production, Isolation, Characterization

The negative control strains were prepared by electroporating *E. coli* BAP1 with either pDCS11, pDCS57, pDCS58, pDCS59, pNDP6, or pNDP7. These strains lacked the oxytetracycline pathway and thus cannot produce oxytetracycline. Test strains were prepared by electroporation of *E. coli* BAP1 with pMRH08, the oxytetracycline gene cluster, and either no additional plasmid, pDCS11, pDCS57, pDCS58, pDCS59, pNDP6, or pNDP7. Tetracycline resistant strains were generated by transforming electrocompetent BAP1/pMRH08 with pDCS11 and pLRVP09. All strains were grown in 25 mL LB medium with appropriate antibiotics in shake flasks at 37°C. At O.D.<sub>600</sub> = 0.4, cultures were induced with 1 mM IPTG and grown at 20°C for 48 hours. Cultures were treated with acetone (1 mL) and vigorously vortexed. Cell debris was removed by centrifugation and Amberlite XAD-16 resin (0.50 g) was added to the clarified media. The resin was incubated with the media overnight and collected by filtration. The organic extracts were eluted from the resin with 20 mL MeOH over 2 hours and concentrated *in vacuo*. Extracts were then resuspended in 100  $\mu$ L MeOH for ESI-LC-MS/MS analysis. All ESI-LC-MS/MS analysis was performed on an API2000 LC/MS/MS System equipped with a turbo-ion spray ESI probe interfaced with a Prominence UFLC. A reverse phase BDS Hypersil C18 column (100 mm×2.1 mm I.D., 3  $\mu$ m particle size,) was employed. Mobile phases throughout experimental were A: 5% MeCN:95% H<sub>2</sub>O with 0.05% formic acid and B: 95% MeCN:5% H<sub>2</sub>O with 0.05% formic acid. For initial determination of oxytetracycline production a gradient program (3 min 100% A, 40 min linear gradient to 100% B, 10 min 100% B) was developed with a flow rate of 0.250 mL/min into the mass spectrometer (Ion spray 5,500 V, mass range 250–550 m/z). For oxytetracycline MS<sup>2</sup> fragmentation scans a gradient program (3 min 100% A, 30 min linear gradient to 100% B, 2 min 100% B) was developed and the mass spectrometer settings were optimized for oxytetracycline (Ion spray 5,500 V, collision energy 50 eV, Q3 scan 400–500 m/z, MS<sup>2</sup> parent ion 461.3 m/z, MS<sup>2</sup> product ion range 198–462 m/z). For MRM quantification of oxytetracycline production, a gradient program identical to that of the MS<sup>2</sup> program was used with the following mass spectrometer settings: Ion spray 5,500 V, collision energy 33 eV, MRM Q1 461.3 m/z, MRM Q3 483.2 m/z. Quantification was performed in triplicate using a standard curve generated from authentic oxytetracycline.

### Analysis of protein over-expression

BAP1, BAP1/pDCS11, and BAP1/pDCS11/pMRH08 were each grown in 25 mL LB medium with appropriate antibiotics in shake flasks at 37°C. At O.D.<sub>600</sub> = 0.4, cultures were induced with 1 mM IPTG and grown at 20°C for 24 hours. Cells were harvested from 5 mL of cell culture by centrifugation at 4000 g and resuspended in 200  $\mu$ L of lysis buffer (100 mM sodium

Phosphate pH 8.0, 300 mM NaCl, 10% (*v/v*) glycerol, 1 mg/mL lysozyme, 1  $\mu$ g/mL pepstatin A, and 1  $\mu$ g/mL leupeptin). The cells were disrupted by sonication on ice, and cell debris was removed by centrifugation at 15000 g and 4°C. The resulting soluble fraction was analyzed by 4–20% gradient SDS-PAGE followed by staining with Coomassie Brilliant Blue.

### Gene cluster identification

To identify polyketide and non-ribosomal peptide biosynthetic gene clusters from bacterial genomes, all genes with cluster of orthologous groups (COG) annotations [35] of polyketide synthase (COG3321), 3-oxoacyl acyl carrier protein synthase (COG0304), and non-ribosomal peptide synthetase (COG1020) were identified from 58 bacterial genomes. A list of genomes used can be found in the expanded experimental procedures in the supplementary information (File S1). COG0304 annotated genes with gene names containing *fab* were discarded as they are involved in fatty acid biosynthesis. Operons containing genes commonly associated with secondary metabolite biosynthesis were included in the gene cluster. Transposons and housekeeping genes were considered to occur outside of the biosynthetic gene cluster. All identified gene clusters can be found in the supplementary information (Table S2).

### $\sigma^{54}$ promoter identification

To identify putative  $\sigma^{54}$  promoter sequences in bacterial genomes the PromScan Perl script [32] was modified to output all hits, including intragenic hits, with a score of 65 or higher. The complete script can be found in the expanded experimental procedures in the supplementary information (File S1). The script was run on Windows using Strawberry Perl (<http://strawberryperl.com/>). The inputs were bacterial genome DNA sequence files (FNA files from NCBI's Genome Database) and the  $\sigma^{54}$  promoter positional weighted matrix file, which is based on 186 known  $\sigma^{54}$  promoter sites and can be found in the expanded experimental procedures in the supplementary information [33].

The results were uploaded to an SQL database created using SQL Server 2008 Express (<http://www.microsoft.com/express/>). PTT files (from NCBI's Genome Database) containing annotation data were uploaded to the database. Using LINQ to SQL, hits generated from the modified PromScan algorithm were linked to genes located within 500 bp upstream and in the same coding direction as the gene start site. Promoter prediction data is available on the web at <http://www.sigma54.ca>.

### 5' RACE

Total RNA was isolated from BAP1 with pMRH08 and pDCS11 as described for qPCR analysis. cDNA was generated using AMV Reverse Transcriptase and the *oxyA* specific primer 5' – AGGCTCATCGTCATGCCGCA – 3', Poly(dC) tails were added using terminal deoxynucleotidyl transferase (Fermentas). The tailed cDNA was amplified by PCR by PrimeStar HS DNA polymerase (Takara) using the *oxyA* specific primer and the abridged anchor primer 5' – GGCCACGCGTCGACTAG-TACGGGIIIGGGIIGGGII – 3'. The diluted PCR reaction was used as the template for a second round of PCR amplification using the *oxyA* nested primer 5' – CCGCCTCGCCCGTCACAC



– 3′ and the abridged universal anchor primer 5′–GGCCACGCGTGCAGTAGTAC– 3′. The PCR products were gel purified and reamplified using the *oxyA* nested primer and the abridged universal anchor primers. The PCR products were then cloned into pCR-Blunt and sequenced.

## Supporting Information

**Figure S1 Transcript levels, determined by qPCR, for *E. coli* transformed with the oxytetracycline gene cluster (pMRH08) compared to a null strain lacking the oxytetracycline gene cluster.**

(TIF)

**Figure S2 Transcript levels, determined by qPCR, for *E. coli* transformed with the oxytetracycline gene cluster (pMRH08) and over-expressing  $\sigma^{54}$  compared to a null strain lacking the oxytetracycline gene cluster and over-expressing  $\sigma^{54}$ .**

(TIF)

**Figure S3 ESI-LC-MS/MS Q1 ion extraction chromatograms for the oxytetracycline  $[M+H]^+$   $m/z = 461$ . Standard:** an oxytetracycline standard. **Positive control:** Organic extracts from a culture of BAP1 with the oxytetracycline biosynthetic gene cluster over-expressing  $\sigma^{54}$ . **Negative control:** organic extracts from a culture of BAP1 with the oxytetracycline biosynthetic gene cluster.  $\sigma^E$ : organic extracts from a culture of BAP1 with the oxytetracycline biosynthetic gene cluster over-expressing  $\sigma^E$ .  $\sigma^F$ : organic extracts from a culture of BAP1 with the oxytetracycline biosynthetic gene cluster over-expressing  $\sigma^F$ .  $\sigma^H$ : organic extracts from a culture of BAP1 with the oxytetracycline biosynthetic gene cluster over-expressing  $\sigma^H$ . MS<sup>2</sup> traces further confirm that no oxytetracycline can be detected in the  $\sigma^E$ ,  $\sigma^F$ , and  $\sigma^H$  samples.

(TIF)

**Figure S4 SDS-PAGE of the soluble fraction from oxytetracycline producing strains.** Lane 1 is protein ladder, lane 2 is the soluble fraction from uninduced BAP1, lane 3 is the

soluble fraction 24 h post induction from BAP1/pDCS11, lane 4 is the soluble fraction 24 h post induction from BAP1/pMRH08/pDCS11. A band corresponding to  $\sigma^{54}$  (54 kDa) is seen in lanes 3 and 4. The OxyA-OxyB KS-CLF heterodimer (45 kDa and 44 kDa respectively) is not observed in lane 4.

(TIF)

**File S1 Expanded experimental procedures.**

(PDF)

**Table S1 Complete data set for qPCR analysis as shown in Figures 1 and 3.** Standard curve, efficiencies, R<sup>2</sup> and Ct values are given.

(PDF)

**Table S2 List of putative  $\sigma^{54}$  promoters identified from a bioinformatics analysis of the oxytetracycline gene cluster.**

(TIF)

**Table S3 List of PKS, NRPS and NRPS/PKS gene clusters with the associated  $\sigma^{54}$  promoters identified from our bioinformatics analysis.**

(PDF)

**Table S4 Tabulated number of PKS, NRPS and NRPS/PKS gene clusters with one of more  $\sigma^{54}$  promoters identified from our bioinformatics analysis at increasing cutoff stringencies.**

(TIF)

## Acknowledgments

The authors thank Philip Pelletier for helpful discussion.

## Author Contributions

Conceived and designed the experiments: DCS CNB. Performed the experiments: DCS KC NP LRVP CNB. Analyzed the data: DCS KC CNB. Contributed reagents/materials/analysis tools: AGG. Wrote the paper: DCS CNB.

## References

- Zhang W, Ames B, Tsai S, Tang Y (2006) Engineered Biosynthesis of a Novel Amidated Polyketide, Using the Malonamyl-Specific Initiation Module from the Oxytetracycline Polyketide Synthase. *Appl Environ Microbiol* 72: 2573–2580.
- Tang L, Shah S, Chung L, Carney J, Katz L, et al. (2000) Cloning and Heterologous Expression of the Epithilone Gene Cluster. *Science* 287: 640–642.
- Schwecke T, Aparicio JF, Molnár I, König A, Khaw LE, et al. (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc Natl Acad Sci USA* 92: 7839–7843.
- Newman DJ, Cragg GM (2007) Natural Products as Sources of New Drugs over the Last 25 Years. *J Nat Prod* 70: 461–477.
- Bode H, Müller R (2006) Analysis of myxobacterial secondary metabolism goes molecular. *J Ind Microbiol Biotechnol* 33: 577–588.
- Udvarý DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, et al. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* 104: 10376–10381.
- Van Lanen SG, Shen B (2006) Microbial genomics for the improvement of natural products discovery. *Curr Opin Microbiol* 9: 252–260.
- Li JWH, Vederas JC (2009) Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* 325: 161–165.
- Rodríguez E, Menzella HG, Gramajo H (2009) Heterologous Production of Polyketides in Bacteria. *Methods Enzymol* 459: 339–365.
- Kao CM, Katz L, Khosla C (1994) Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* 265: 509–512.
- Binz TM, Wenzel SC, Apotheke HJS, Bechthold A, Müller R (2008) Heterologous Expression And Genetic Engineering of the Phenalinolactone Biosynthetic Gene Cluster by Using Red/ET Recombinering. *ChemBioChem* 9: 447–454.
- Stevens DC, Henry MR, Murphy KA and Boddy CN (2010) Heterologous expression of the oxytetracycline biosynthetic pathway in *Mycococcus xanthus*. *Appl Environ Microbiol* 76: 2681–2684.
- Baltz RH (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J Ind Microbiol Biotechnol* 37: 759–772.
- Chen Y, Wendt-Pienkowski E, Shen B (2008) Identification and utility of FdmR1 as a *Streptomyces* antibiotic regulatory protein activator for fredericamycin production in *Streptomyces griseus* ATCC 49344 and heterologous hosts. *J Bacteriol* 190: 5587–5596.
- Gomez-Escribano JP, Bibb MJ (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotech* 4: 207–215.
- Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* 291: 1790–1792.
- Boddy CN, Hotta K, Tse ML, Watts RE, Khosla C (2004) Precursor-directed biosynthesis of epithilone in *Escherichia coli*. *J Am Chem Soc* 126: 7436–7437.
- Watanabe K, Hotta K, Praseuth AP, Koketsu K, Migita A, et al. (2006) Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat Chem Biol* 2: 423–428.
- Jørgensen H, Fjærviik E, Hakvåg S, Bruheim P, Bredholt H, et al. (2009) Candicidin Biosynthesis Gene Cluster Is Widely Distributed among *Streptomyces* spp. Isolated from the Sediments and the Neuston Layer of the Trondheim Fjord, Norway. *Applied Environ Microbiol* 75: 3296–3303.
- Fisch KM, Gurguil C, Heycke N, van der Sar SA, Anderson SA, et al. (2009) Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat Chem Biol* 5: 494–501.
- Bibb MJ (2005) Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol* 8: 208–215.
- Zhang W, Li Y, Tang Y (2008) Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*. *Proc Natl Acad Sci USA* 105: 20683–20688.
- Pickens LB, Tang Y (2010) Oxytetracycline biosynthesis. *J Biol Chem* 285: 27509–27515.

24. Potrykus K, Cashel M (2008) (p)ppGpp: Still Magical? *Annu Rev Microbiol* 62: 35–51.
25. Hesketh A, Sun J, Bibb MJ (2001) Induction of ppGpp synthesis in *Streptomyces coelicolor* A3(2) grown under conditions of nutritional sufficiency elicits *actII-ORF4* transcription and actinorhodin biosynthesis. *Mol Microbiol* 150: 1485–1493.
26. Knauber T, Doss SD, Gerth K, Perlova O, Müller R, et al. (2008) Mutation in the *rel* gene of *Sorangium cellulosum* affects morphological and physiological differentiation. *Mol Microbiol* 69: 254–266.
27. Wösten MM (1998) Eubacterial sigma factors. *FEMS Microbiol Rev* 22: 127–150.
28. Britton RA, Eichenberger P, Gonzalez-Pastor JE, Fawcett P, Monson R, et al. (2002) Genome-Wide Analysis of the Stationary-Phase Sigma Factor (Sigma-H) Regulon of *Bacillus subtilis*. *J Bacteriol* 184: 4881–4890.
29. Chen G, Schellhorn HE (2003) Controlled induction of the RpoS regulon in *Escherichia coli*, using an RpoS-expressing plasmid. *Can J Microbiol* 49: 733–740.
30. Zhao K, Liu M, Burgess RR (2010) Promoter and regulon analysis of nitrogen assimilation factor,  $\sigma^{54}$ , reveal alternative strategy for *E. coli* MG1655 flagellar biosynthesis. *Nucleic Acid Res* 38: 1273–1283.
31. Chopra I, Roberts M (2001) Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev* 65:232–260.
32. Burrows GE, Morton RJ, Fales WH (1993) Microdilution antimicrobial susceptibilities of selected gram-negative veterinary bacterial isolates. *J Vet Diagn Invest* 5:541–547.
33. Gao X, Wang P, Tang Y (2010) Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Appl Microbiol Biotechnol* 88:1233–1242.
34. Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD (2000) The Bacterial Enhancer-Dependent  $\sigma^{54}$  ( $\sigma^N$ ) Transcription Factor. *J Bacteriol* 182: 4129–4136.
35. Studholme DJ, Buck M, Nixon T (2000) Identification of potential N-dependent promoters in bacterial genomes. *Microbiology* 146: 3021–3023.
36. Barrios H, Valderrama B, Morett E (1999) Compilation and analysis of  $\sigma^{54}$ -dependent promoter sequences. *Nucleic Acids Res* 27: 4305–4313.
37. Gronewold TMA, Kaiser D (2007) Mutation of the Act Promoter in *Myxococcus xanthus*. *J Bacteriol* 189: 1836–1844.
38. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acid Res* 28: 33–36.
39. Ossa F, Diodati ME, Caberoy NB, Giglio KM, Edmonds M, et al. (2007) The *Myxococcus xanthus* Nla4 Protein Is Important for Expression of Stringent Response-Associated Genes, ppGpp Accumulation, and Fruiting Body Development. *J Bacteriol* 189: 8474–8483.
40. Giglio KM, Eisenstatt J, Garza AG (2010) Identification of Enhancer Binding Proteins Important for *Myxococcus xanthus* Development. *J Bacteriol* 192: 360–364.
41. Mulligan ME, McClure WR (1986) Analysis of the occurrence of promoter-sites in DNA. *Nucleic Acids Res* 14: 109–126.
42. Gordon JJ, Towsey MW, Hogan JM, Mathews SA, Timms P (2006) Improved prediction of bacterial transcription start sites. *Bioinformatics* 22:142–148.
43. McLeay RC, Leat CJ, Bailey TL (2011) Tissue-specific prediction of directly regulated genes. *Bioinformatics* 27: 2354–2360.
44. Qiu P (2003) Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. *Biochem Biophys Res Commun* 309: 495–501.
45. Nomura M, Gourse R, Baughman G (1984) Regulation of the Synthesis of Ribosomes and Ribosomal Components. *Ann Rev Biochem* 53: 75–117.
46. Aseev LV, Levandovskaya, AA Tchufistova LS, Saptsova NV, Boni IV (2008) A new regulatory circuit in ribosomal protein operons: S2-mediated control of the rpsB-tsif expression in vivo. *RNA* 14: 1882–1894.
47. Aoki H, Adams SL, Chung DG, Yaguchi M, Chuang SE, et al. (1991). Cloning, sequencing and overexpression of the gene for prokaryotic factor EF-P involved in peptide bond synthesis. *Nucleic Acids Res* 19: 6215–6220.
48. Kumar R, Shimizu K (2011) Transcriptional regulation of main metabolic pathways of *cyoA*, *cydB*, *fur*, and *fur* gene knockout *Escherichia coli* in C-limited and N-limited aerobic continuous cultures. *Microb Cell Fact* 10:3
49. Jakobsen JS, Jelsbak L, Jelsbak L, Welch RD, Cummings C, et al. (2004)  $\sigma^{54}$  Enhancer Binding Proteins and *Myxococcus xanthus* Fruiting Body Development. *J Bacteriol* 186: 4361–4368.
50. Schumacher J, Joly N, Rappas M, Zhang X, Buck M. (2006) Structures and organisation of AAA+ enhancer binding proteins in transcriptional activation. *J Struct Biol* 156: 190–199.
51. Dworkin J, Jovanovic G, Model P (1997) Role of Upstream Activation Sequences and Integration Host Factor in Transcriptional Activation by the Constitutively Active Prokaryotic Enhancer-binding Protein PspF. *J Mol Biol* 273: 377–388.
52. Janaszak A, Majczak W, Nadratowska B, Szalewska-Palasz A, Konopa G, et al. (2007) A  $\sigma^{54}$ -dependent promoter in the regulatory region of the *Escherichia coli* *rpoH* gene. *Microbiology* 153: 111–123.
53. Studholme DJ, Dixon R. (2003) Domain Architectures of  $\sigma^{54}$ -Dependent Transcriptional Activators. *J Bacteriol* 185: 1757–1767.
54. Terui Y, Higashi K, Taniguchi S, Shigemasa A, Nishimura K, et al. (2007) Enhancement of the synthesis of RpoN, Cra, and H-NS by polyamines at the level of translation in *Escherichia coli* cultured with glucose and glutamate. *J Bacteriol* 189: 2359–2368.
55. Terui Y, Higashi K, Tabai Y, Tomitori H, Yamamoto K, et al. (2009) Enhancement of the Synthesis of RpoE and StpA by Polyamines at the Level of Translation in *Escherichia coli* under Heat Shock Conditions. *J Bacteriol* 191: 5348–5357.
56. Lackner G, Schenk A, Xu Z, Reinhardt K, Yunt ZS, et al. (2007) Biosynthesis of Pentangular Polyphenols: Deductions from the Benastatin and Grisoorhodin Pathways. *J Am Chem Soc* 129: 9306–9312.
57. Buntin K, Irschik H, Weissman KJ, Luxenburger E, Blöcker H, et al. (2010) Biosynthesis of Thuggacins in Myxobacteria: Comparative Cluster Analysis Reveals Basis for Natural Product Structural Diversity. *Chem Biol* 17: 342–356.
58. Carlson JC, Fortman JL, Anzai Y, Li S, Burr DA, et al. (2010) Identification of the tirandamycin biosynthetic gene cluster from *Streptomyces* sp. 307–9. *ChemBioChem* 11: 564–572.
59. Zaleta-Rivera K, Charkoudian LK, Ridley CP, Khosla C (2010) Cloning, Sequencing, Heterologous Expression, and Mechanistic Analysis of A-74528 Biosynthesis. *J Am Chem Soc* 132: 9122–9128.
60. Andexer JN, Kendrew SG, Nur-e-Alam M, Lazos O, Foster TA, et al. (2011) Biosynthesis of the immunosuppressants FK506, FK520, and rapamycin involves a previously undescribed family of enzymes acting on chorismate. *Proc Natl Acad Sci USA* 108: 4776–4781.
61. Yurkovich ME, Tyrakis PA, Hong H, Sun Y, Samborsky M, et al. (2012) A Late-Stage Intermediate in Salinomycin Biosynthesis Is Revealed by Specific Mutation in the Biosynthetic Gene Cluster. *ChemBioChem* 13: 66–71.
62. Lundgren BR, Boddy CN (2007) Sialic acid and N-acyl sialic acid analog production by fermentation of metabolically and genetically engineered *Escherichia coli*. *Org Biomol Chem* 12: 1903–1909.