UNIVERSITY OF CALIFORNIA SAN DIEGO

Heterologous expression and genetic manipulation of natural product biosynthetic gene clusters from marine bacteria

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Marine Biology

by

Jamie R. Zhang

Committee in charge:

Professor Bradley S. Moore, Chair
Professor Rachel J. Dutton
Professor William H. Gerwick
Professor Susan S. Golden
Professor Chambers C. Hughes
Professor Kit Pogliano

2019

The Dissertation of Jamie R. Zhang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____
Chair

University of California San Diego

2019

DEDICATION

To Mr. Walter K. Erhardt, who taught me that education is about curiosity and persistence.

EPIGRAPH

Not all of us can do great things, but we can all do small things with great love.

Mother Teresa

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Bradley S. Moore for his support as the chair of my committee. He has been a great mentor and my strongest professional advocate for the last six years of my life. I cannot repay him but hope to have the opportunity to pay it forward in the future. I would also like to acknowledge Dr. Xiaoyu Tang for his immeasurable help and my parents and sister for their unconditional love.

Chapter 1, in part, has been submitted for publication of the material as it may appear in *Natural Product Reports*, Zhang, J. J.; Tang, X.; Moore, B. S., 2019. The dissertation author was the primary author of this material.

Chapter 2, in full, is a reprint of the material as it appears in *Applied Microbiology and Biotechnology*, Zhang, J. J.; Moore, B. S.; Tang, X., 2018. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in *mBio*, Zhang, J. J.; Tang, X.; Zhang, M.; Nguyen, D.; Moore, B. S., 2017. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, has been submitted for publication of the material as it may appear in *Nature Chemical Biology*, Zhang, J. J.; Tang, X.; Huan, T.; Ross, A. C.; Moore, B. S., 2019. The dissertation author was the primary investigator and author of this paper.

VITA

2013 Bachelor of Arts, Harvard College

2013-2016 National Science Foundation Graduate Research Fellow

2017-2019 National Institutes of Health F31 Predoctoral Fellow

2019 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Gunther, D., LeBlanc, G., Prasai, D., Zhang, J. R., Cliffel, D. E., Bolotin, K. I. & Jennings, G. K. Photosystem I on graphene as a highly transparent, photoactive electrode. *Langmuir* **29,** 4177-4180, doi:10.1021/la305020c (2013).

Tang, X., Li, J., Millan-Aguinaga, N., Zhang, J. J., O'Neill, E. C., Ugalde, J. A., Jensen, P. R., Mantovani, S. M. & Moore, B. S. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10,** 2841-2849, doi:10.1021/acschembio.5b00658 (2015).

Zhang, J. J. & Moore, B. S. Digging for biosynthetic dark matter. *Elife* **4,** e06453, doi:10.7554/eLife.06453 (2015).

Zhang, J. J., Tang, X., Zhang, M., Nguyen, D. & Moore, B. S. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. *MBio* **8**, doi:10.1128/mBio.01291-17 (2017).

Zhang, J. J., Moore, B. S. & Tang, X. Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters. *Appl. Microbiol. Biotechnol.* **102,** 8437-8446, doi:10.1007/s00253-018-9283-z (2018).

Zhang, J. J., Yamanaka, K., Tang, X. & Moore, B. S. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol.* **621**, 87-110 (2019).

Zhang, J. J., Tang, X. & Moore, B. S. Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.* doi:10.1039/C9NP00025A (2019).

Zhang, J. J., Tang, X., Huan, T., Ross, A. C. & Moore, B. S. Avant-garde assembly-line biosynthesis expands diversity of cyclic lipodepsipeptide products. *bioRxiv***,** 560987, doi:10.1101/560987 (2019).

ABSTRACT OF THE DISSERTATION

Heterologous expression and genetic manipulation of
natural product biosynthetic gene clusters from marine bacteria

by

Jamie R. Zhang

Doctor of Philosophy in Marine Biology

University of California San Diego, 2019

Professor Bradley S. Moore, Chair

Microorganisms are prolific producers of natural product small molecules, many of which have been developed into life-saving human drugs. Over the last century, naturally produced chemical compounds have been isolated and characterized from all kingdoms of life, but only recently have we begun to understand the genetic underpinnings of natural product biosynthesis in living cells, an endeavor which has been particularly fruitful in bacteria. Bacteria have relatively simpler genomes compared to higher organisms; furthermore, bacteria tend to cluster elements involved in production of specialized small molecules along their genomes in 'biosynthetic gene clusters,' or BGCs. BGCs encoding non-ribosomal peptide synthetase

(NRPS) or polyketide synthase (PKS) assembly line biosynthetic machineries have become easy to detect in sequenced bacterial genomes using automated bioinformatic tools, leading to the discovery of many more uncharacterized or 'cryptic' BGCs than isolated natural products identified from laboratory culture. Therefore, the work described in this dissertation focuses on the development of tools to characterize cryptic BGCs through heterologous expression and genetic manipulation. While initial efforts focused on terrestrial sources due to ease of access and historical precedence, pioneering work has led to the recognition that marine bacteria are also a rich source of bioactive compounds. Chapter 2 of this dissertation describes the development of a *Salinispora tropica* heterologous host, *S. tropica* CNB-4401, which represents the first marine actinobacterial host for natural product BGC expression. Alternatively, chapter 3 describes the development of a broad-host-range expression vector compatible with Gram-negative proteobacterial hosts, using the violacein BGC from the marine γ-proteobacterium *Pseudoalteromonas luteoviolacea* 2ta16 as a proof-of-principle. This work not only introduces a new genetic platform for cloning and heterologous expression of BGCs in Gram-negative hosts, it also challenges the assumption that host phylogeny is an accurate predictor of host compatibility. Finally, chapter 4 describes the thorough interrogation of two avant-garde hybrid NRPS/PKS assembly line biosynthetic pathways from marine α-protoebacteria responsible for producing a diverse group of immunosuppressive cyclic lipodepsipeptides called the thalassospiramides. This work leverages newly developed tools for cloning, expression, and precise genetic manipulation of large BGCs in a *Pseudomonas* heterologous host.

CHAPTER 1. Genetic platforms for heterologous expression of microbial natural products.

1.1 Abstract

Natural products are of paramount importance in human medicine. Not only are most antibacterial and anticancer drugs derived directly from or inspired by natural products, many other branches of medicine, such as immunology, neurology, and cardiology, have similarly benefited from natural product-based drugs. Typically, the genetic material required to synthesize a microbial specialized product is arranged in a multigene biosynthetic gene cluster (BGC), which codes for proteins associated with molecule construction, regulation, and transport. The ability to connect natural product compounds to BGCs and vice versa, along with ever-increasing knowledge of biosynthetic machineries, has spawned the field of genomics-guided natural product genome mining for the rational discovery of new chemical entities. One significant challenge in the field of natural product genome mining is how to rapidly link orphan biosynthetic genes to their associated chemical products. This review highlights state-of-the-art genetic platforms to identify, interrogate, and engineer BGCs from diverse microbial sources, which can be broken into three stages: (1) cloning and isolation of genomic loci, (2) heterologous expression in a host organism, and (3) genetic manipulation of cloned pathways. In the future, we envision natural product genome mining will be rapidly accelerated by *de novo* DNA synthesis and refactoring of whole biosynthetic pathways in combination with systematic heterologous expression methodologies.

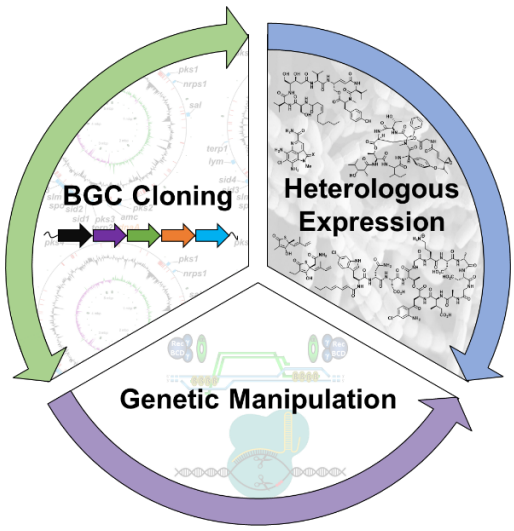


Figure 1.1 Graphical abstract of the chapter.

1.2 Introduction

Natural products, or specialized small molecules produced by living organisms, have long fascinated chemists due to their complex three-dimensional structures, which make them challenging to produce using synthetic organic chemistry.[1] Perhaps more importantly, structural complexity and multidimensionality confer many natural products with potent and specific biological activities, making them privileged scaffolds in the quest to develop new medicines.[2] Natural products have played an invaluable role in deepening our understanding of various cellular processes and even facilitated the discovery of important and evolutionarily conserved macromolecules, such as the protein kinase mTOR, or mechanistic target of rapamycin, which is blocked by the *Streptomyces* natural product rapamycin.[3] Rapamycin has become a life-saving immunosuppressant drug, and the discovery of mTOR spawned a vibrant field of study uncovering its central role in physiology, metabolism, aging, and common diseases such as cancer and epilepsy.[4]

The ability of all living organisms to biosynthesize endogenous, specialized small molecules is genetically encoded. Making connections between isolated small molecules and the genes responsible for their construction has been particularly productive in microorganisms, which generally cluster elements involved in natural product biosynthesis along their genomes in "biosynthetic gene clusters" (BGCs).

Clustering of specialized genetic elements carries an additional benefit: the ability to clone and transfer whole BGCs to heterologous host organisms for expression and characterization. Heterologous expression of BGCs for characterization or identification of new chemical entities is advantageous for several reasons. As more and more genome sequences from diverse microbial sources become available, heterologous expression circumvents the need to develop new genetic tools to interrogate pathways from each new genus or species of interest. Furthermore, it enables characterization of BGCs from microbes that have yet to be cultured such as those identified from obligate symbionts or environmental DNA (eDNA). Successful heterologous reconstitution of a

BGC allows rapid delineation of all essential specialized genes involved in the production of a microbial bio-chemical. Finally, genetic platforms for interrogation of BGCs that are developed can be optimized and universally applied, both for heterologous expression as well as rapid genetic manipulation of cloned pathways to perform biosynthetic investigations or BGC refactoring.

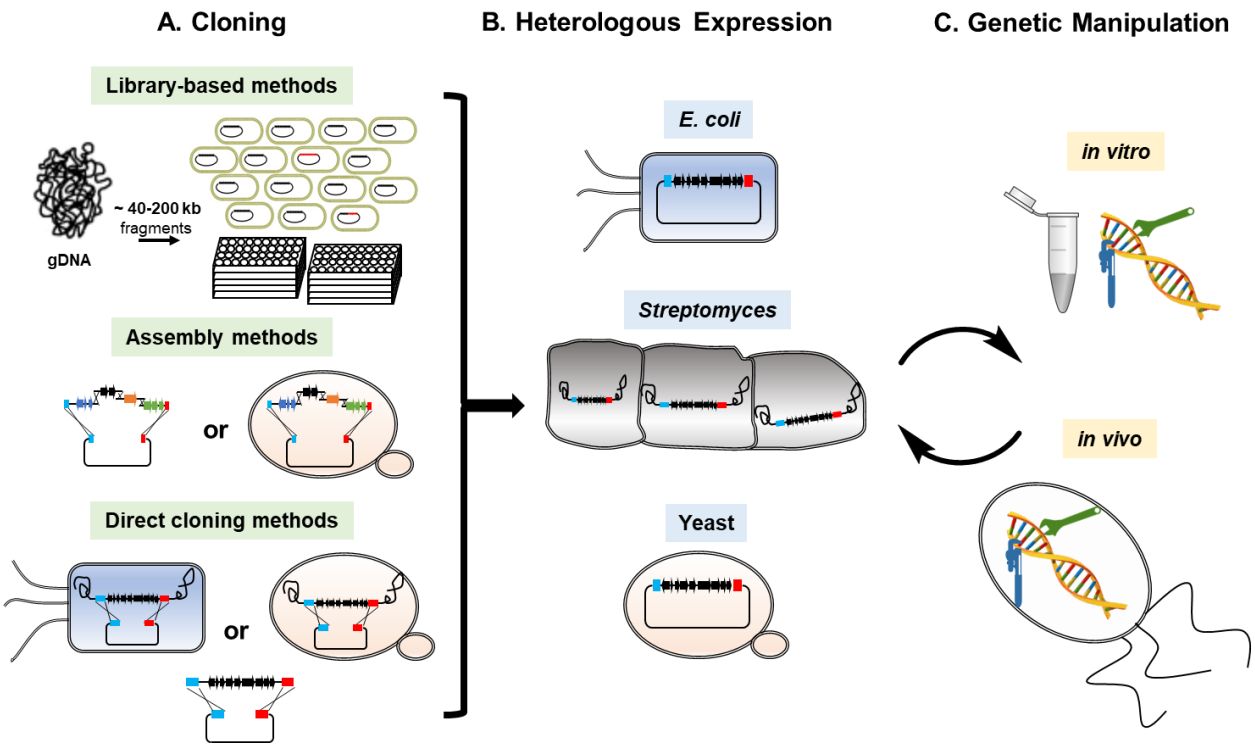


Figure 1.2 General workflow for (A) cloning, (B) heterologous expression, and (C) genetic manipulation of microbial BGCs.

In this article, we review genetic platforms that have been established for heterologous expression of microbial natural products. Integrated platforms include three stages: 1) cloning and isolation of selective genomic loci containing BGCs, 2) expression in a heterologous host organism, and 3) genetic manipulation of cloned pathways for interrogation or activation (**Figure 1.2**). Following this workflow, we highlight advancements in cloning methods, heterologous hosts, and genetic manipulation of large microbial BGCs used for successful heterologous reconstitution

of various natural products (**Figure 1.3**). Lastly, we highlight recent studies that utilize BGC refactoring, DNA synthesis, and *de novo* pathway design to access new chemical entities.

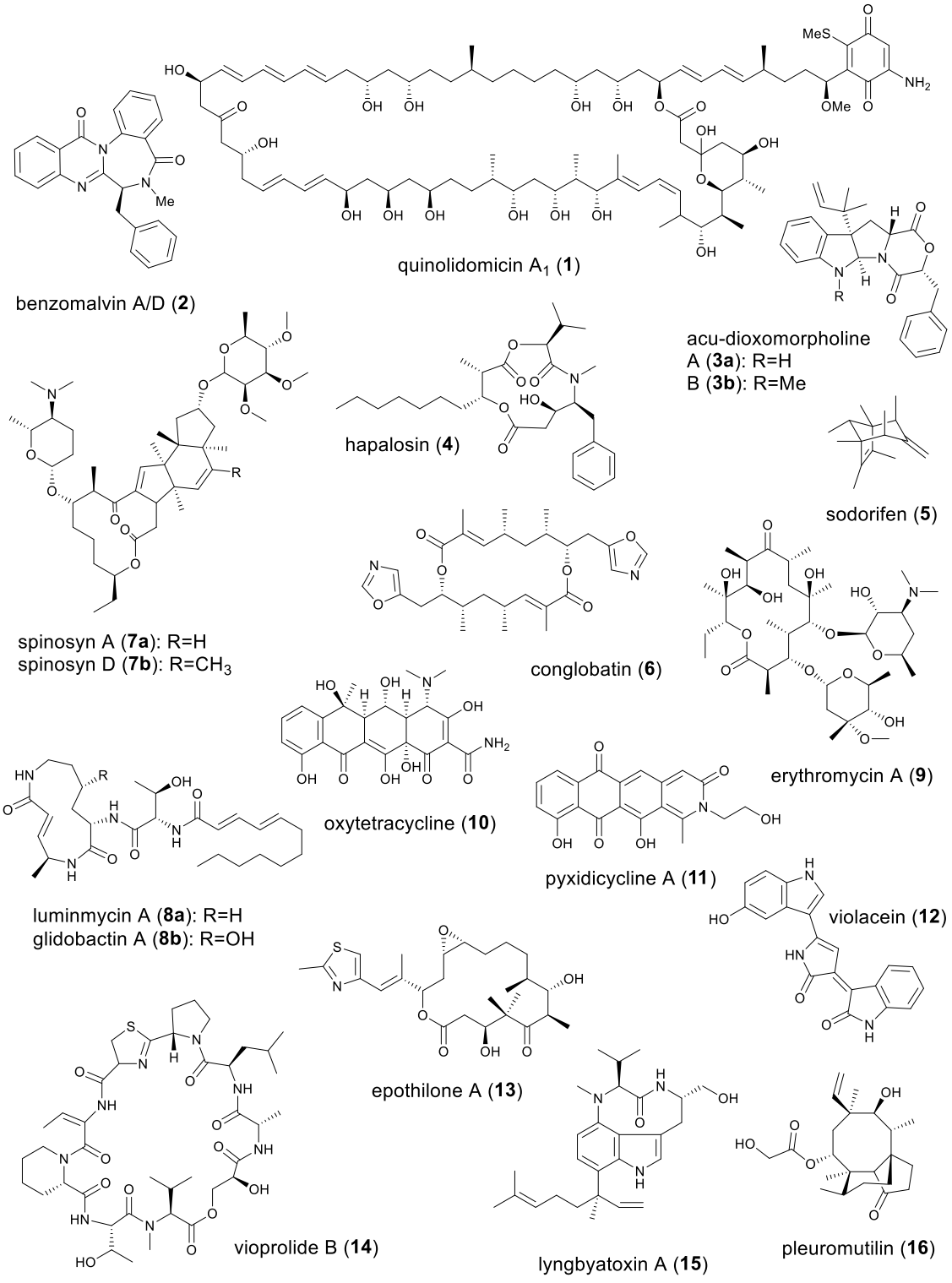


Figure 1.3 Chemical structures of natural products mentioned in this chapter.

1.3 Cloning of microbial BGCs

Due to the large size, repetitive nature, and high GC-content of many microbial BGCs, cloning has remained a challenging step in heterologous reconstitution of natural product pathways. While advancements in genome sequencing, bioinformatics, and molecular biology techniques have changed the landscape of BGC cloning over the years, many different types of cloning methods, including library-based methods, assembly methods, and direct cloning methods, have been developed and continue to be used successfully today. In this section, we highlight these techniques and their associated advantages and limitations (**Table 1.1**).

1.3.1 Library-based methods

Early efforts to clone microbial BGCs relied heavily on library-based methods, which involve the generation of a clone library of random genomic DNA (gDNA) fragments in *Escherichia coli.* Library generation is particularly useful when complete genome sequence information is lacking, or when it is advantageous to catalogue the complete genetic material from an organism or metagenomic sample. Furthermore, library generation breaks the chromosome into smaller chunks that can be more easily sequenced and assembled than whole chromosomes and has therefore been a vital component of numerous genome sequencing efforts. Many BGCs have been cloned from cultured organisms into cosmid[5-9] and fosmid[10-14] libraries, which hold inserts of approximately 40 kb in size and are packaged and delivered to *E. coli* by bacteriophages. While these libraries hold similar sized inserts, fosmids exist at low or single copy number in *E. coli* cloning hosts and are therefore considered more stable than cosmids, particularly for highly repetitive DNA. Brady and colleagues have pioneered the generation of cosmid libraries for discovery and characterization of BGCs from soil metagenomic DNA,[15-21] thus accessing pathways from organisms that have yet to be cultured. Cosmid libraries from environmental DNA (eDNA) can be maintained and continually re-screened for new categories of specialized biosynthetic genes.[16,19-21] Cloning of metagenomic DNA is particularly challenging due to the low

enrichment and potentially low quality of genetic material from an individual microbe within a heterogeneous population, but library cloning has proven to be a successful approach for accessing BGCs from eDNA. However, BGCs cloned using these methods are often split across multiple library clones and thus need to be stitched together and trimmed before subsequent use.[8,10,19,21] Analysis of 540 full BGCs deposited in the MIBiG database[22] as of August 2018 revealed that these pathways range in size from 204 bp (representing a small ribosomally encoded peptide) to 148,229 bp, with an average of approximately 36 kb. Although this is below the 40 kb threshold, it is likely that an average-sized BGC will be split across multiple fosmid or cosmid clones.

Library vectors that hold larger inserts include those based on bacterial artificial chromosomes (BACs) and P1 artificial chromosomes (PACs). Unlike cosmid and fosmid libraries, BAC and PAC libraries are prepared by direct transfer of DNA to *E. coli* via electroporation. High molecular weight DNA can be isolated using special techniques to promote uniform insert sizes above 100 kb, which are stably maintained within low copy number vector backbones. Several recent studies have utilized BAC library clones to characterize large BGCs (>55 kb) encoding assembly line biosynthetic pathways, which include modular polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) gene clusters.[23-27] The quinolidomicin A_1 (**1**) BGC, which encodes a PKS and spans over 200 kb in size, was very recently cloned in a BAC library and represents the largest BGC cloned and heterologously expressed to-date.[28] However, uniform, high molecular weight DNA can be difficult to isolate, and low copy number plasmids can also be more challenging to work with. Because of these technical difficulties, research groups are increasingly outsourcing large-insert library generation to companies. Several reports of large BGC cloning have leveraged PAC libraries generated by the Canadian company Bio S&T using vector pESAC13, which can be readily transferred to various *Streptomyces* heterologous hosts by conjugation.[29-33]

Recently, a group of researchers studying fungal secondary metabolism retrofitted a BAC vector backbone with an autonomously replicating sequence from *Aspergillus* (AMA1) to generate a self-replicating fungal artificial chromosome (FAC) that can be used for library generation.[34] FAC libraries generated using this vector can be directly transferred to the host *Aspergillus nidulans* for heterologous expression and mass spectrometry screening for identification of new chemical entities. Libraries were constructed and screened from the gDNA of *A. terrus*, *A. aculeatus*, and *A. wentii* and resulted in the identification of 15 unique mass signals produced by cryptic biosynthetic machinery, a few of which have been characterized in greater detail, including benzomalvin A/D (**2**) and acu-dioxomorpholines A (**3a**) and B (**3b**).[35-37]

Library-based cloning methods have contributed greatly to our understanding of microbial BGCs and continue to be used routinely and successfully. However, library cloning is an untargeted approach in which most library clones do not contain genomic regions of interest and thus must be extensively screened. With the proliferation of publicly available genome sequence information, more targeted approaches, such as assembly and direct cloning methods, are being utilized with greater frequency.

1.3.2 Assembly methods

Methods that rely on *in vitro* assembly of BGCs from smaller fragments represent an attractive alternative to library-based approaches, as small fragments of DNA can be generated quickly and cheaply using polymerase chain reaction (PCR) and are easy to work with. In 2009, Gibson and co-workers reported the *in vitro* assembly of very large DNA molecules leveraging the concerted action of three enzymes – a 5' exonuclease, a DNA polymerase, and a DNA ligase – in an isothermal, single-reaction method.[38] Gibson assembly kits can be purchased commercially and are used ubiquitously in molecular biology. Within the field of natural products research, Gibson assembly has been successfully used to reconstruct small BGCs, usually around 10 kb in length.[39-43] Direct pathway cloning, or DiPaC, which relies on Gibson assembly,

has been successfully used to clone small but also larger (20 to 55 kb) BGCs in a stepwise, combinatorial fashion for heterologous production of compounds such as hapalosin (**4**) and sodorifen (**5**).[44-46] Alternatively, a method called twin-primer assembly, or TPA, is designed for assembly of PCR-amplified fragments without the use of enzymes.[47] TPA relies of annealing of complementary single-stranded overhangs designed into PCR primer sequences and has been used to assemble a 31 kb plasmid at approximately 50% fidelity.[47]

Phage recombinases have also been used to assemble BGCs. Also referred to as integrases, these enzymes catalyze recombination across relatively short recognition sequences, often referred to as *attP* and *attB* sites, to generate new *attL* and *attR* sequence junctions. In this way, DNA fragments can be stitched together in a specific order and orientation into a self-replicating construct and selected for using selectable markers. Serine integrase recombinational assembly, or SIRA, utilizes purified bacteriophage integrases ΦC31 and Bxb1 as well as their recombination directionality factors (RDFs) to assemble multiple DNA fragments into functional plasmids.[48,49] Site-specific recombination-based tandem assembly, or SSRTA, is a similar approach that leverages other serine integrases, including ΦBT1, TG1, and ΦRv1 in addition to Bxb1.[50,51] However, when directly compared, Bxb1 displayed the highest *in vitro* assembly efficiency.[52] Homing endonucleases, which recognize long asymmetric sequences that are therefore rarely present in natural DNA, have also been used for the iterative assembly of standardized DNA parts through a cut-and-paste mechanism similar to standard restriction-digestion and ligation cloning in a method called iBrick.[53] Recently, programmable DNA-guided restriction enzymes have been developed leveraging the Argonaute enzyme from *Pyrococcus furiosus*, which enables the cleavage of virtually any DNA sequence and generation of defined sticky ends for facile DNA assembly.[54] These artificial restriction enzymes, or AREs, are easily programmable and will likely find broad utility in biological research.

In vivo DNA assembly methods have been widely used to reconstruct large microbial BGCs. DNA assembly leveraging high rates of *in vivo* homologous recombination in yeast has

been shown to be accurate for assembly of up to 10 fragments of DNA, which is particularly useful for large and high GC-content gene clusters that can only be amplified in 2-5 kb pieces.[55-59] Several named methods, including DNA assembler,[60-67] overlap extension PCR-yeast homologous recombination (ExRec),[68] and yeast recombinational cloning-enabled pathway transfer and expression tool (yTREX),[69] all rely on homologous recombination in yeast for assembly of multiple fragments of DNA into functional plasmids for heterologous expression of natural product BGCs. Although not as naturally recombinant as yeast, *E. coli* strains equipped with Red/ET recombineering machinery have also been used for *in vivo* assembly of BGCs in a method called artificial gene operon assembly system, or AGOS.[70]

While assembly methods generally require the least technical expertise and have the greatest potential for automation, they are often not feasible or cumbersome for large BGCs and present the greatest risk for introducing mutations into cloned pathways. Direct cloning methods, in which whole BGCs are directly targeted for cloning, represent the most elegant approach to obtaining BGCs for heterologous expression and are becoming increasingly user friendly.

1.3.3 Direct cloning methods

Yeast homologous recombination, in addition to catalyzing multi-part DNA assembly, has been extensively leveraged for direct cloning of whole BGCs with pre-defined boundaries. Transformation-associated recombination (TAR) cloning in *Saccharomyces cerevisiae* can be used for the selective isolation of any genomic fragment into a circular yeast artificial chromosome (YAC), which can propagate, segregate, and be selected for in yeast.[71] YAC clones arise from homologous recombination between gDNA fragments, which can be prepared by random shearing or enzymatic digestion, and a linearized TAR cloning vector containing two targeting hooks with sequences identical to those flanking the genomic loci of interest. Following successful cloning in yeast, constructs are shuttled to *E. coli* for detailed characterization and verification by PCR, sequencing, and/or restriction digestion. For the purposes of cloning BGCs for heterologous

expression, TAR cloning vectors can also include elements for transfer to and maintenance in a heterologous host organism. Moore and colleagues developed first-generation TAR cloning vectors, pCAP01[72] and pCAPB02,[73] which are yeast-*E. coli*-*Streptomyces* and yeast-*E. coli*-*Bacillus* shuttle vectors, respectively, and can be assembled into a cluster-specific capture vector by addition of two long homology arms of ~1 kb each. Many bacterial BGCs have been directly cloned and interrogated using these first-generation TAR vectors.[73-82] However, pCAP vector backbones possess yeast origins of replication, which are essential for cloning bacterial DNA but also result in high rates of plasmid recircularization via non-homologous end joining (NHEJ) to greatly increase the number of empty vectors that must be screened against. By adapting a previously established method utilizing counterselection to select against plasmid recircularization via NHEJ,[83] two second-generation vectors, pCAP03[84] and pCAP05,[55] were developed and used to clone and express bacterial BGCs.[85] These vectors also employ much shorter homology arms of 50 bp each, which simplifies the procedure for preparing cluster-specific capture vectors in addition to significantly improving the efficiency of the TAR cloning experiment by introducing a mechanism for counterselection.[84] Furthermore, pCAP05 is a broad-host-range expression vector for Gram-negative host organisms, thus expanding the host range of the pCAP vector series to include organisms such as *Pseudomonas* and *Agrobacterium*.[86]

TAR cloning is most efficient when homologous sequences are located as closely as possible to DNA ends,[87] perhaps because homologous recombination is a mechanism to repair DNA double-strand breaks (DSBs) that arise during mitosis. Thus, the success of TAR cloning is also greatly enhanced if restriction sites can be identified just beyond the boundaries of the BGC of interest and the targeting hooks are designed as closely as possible to these sites. Unfortunately, this may not be feasible for many BGCs due to a lack of available restriction sites associated with commercially available restriction enzymes that do not also cut within the BGC. Recently, Lee, Larionov, and Kouprina reported a method combining CRISPR/Cas9-mediated *in vitro* digestion of DNA with TAR cloning, which resulted in a dramatic increase in the fraction of

positive clones.[88] Although this method has not yet been reported for TAR cloning of a microbial BGC for heterologous expression, the use of CRISPR/Cas9 for *in vitro* digestion of gDNA is generally applicable for a number of direct cloning methods and has already been used in combination with RecE catalyzed linear-linear homologous recombination (LLHR) as well as Gibson assembly (detailed below).

Reconstitution of Rac prophage enzymes RecE and RecT in *E. coli* enables *in vivo* homologous recombination of two linear DNA fragments in a method called LLHR,[89] which is highly analogous to TAR cloning in yeast. While arguably not as robust, LLHR is more attractive than TAR because it uses *E. coli* as a cloning host, which makes it faster and also eliminates a step compared to the TAR cloning process. LLHR has been pioneered as a method for BGC cloning by Zhang, Stewart, Müller and colleagues and has been applied for investigation of BGCs from various microbial sources.[89-93] Exonuclease combined with RecET recombination, or ExoCET, is based on the principle that direct cloning efficiencies can be improved by pre-annealing of linear vectors and target DNA using the exonuclease T4 DNA polymerase before delivery to *E. coli* cells.[94,95] Furthermore, LLHR operates on the same principle as homologous recombination in yeast, wherein precise digestion of gDNA prior to cloning using either commercially available restriction enzymes or CRISPR/Cas9 programmed restriction digestion greatly enhances cloning efficiencies.[94]

Finally, various *in vitro* methods have been used for direct cloning of BGCs. Single-strand overlapping annealing, or SSOA, is an *in vitro* approach that, like ExoCET, also utilizes restriction digestion coupled with exonuclease treatment, but does not leverage any *in vivo* recombination machinery.[96] Direct cloning of large pathways from complex mixtures of gDNA using Gibson assembly has also been reported.[97-99] In work by Leadlay and colleagues on expression of the anticancer compound conglobatin (**6**), a precise 41 kb fragment of DNA was generated by gDNA digestion using restriction enzymes XhoI and EcoRI, and digested DNA was carefully purified by gel electrophoresis to remove fragments less than 20 kb in size.[97] Alternatively, CRISPR/Cas9

can be used for programmed DNA digestion in agarose gel plugs prior to Gibson assembly, as has been reported for Cas9 assisted targeting of chromosome segments, or CATCH, which has been successfully used for targeting up to 100 kb in a single step.[98,99] Thus, like *in vivo* methods, *in vitro* direct cloning approaches greatly benefit from careful preparation of DNA through pre-treatment and purification.

If natural restriction sites are not present at the boundaries of a BGC of interest, researchers have also relied on genetic manipulation to introduce unique sites for restriction enzymes or homing endonucleases so that BGCs can be precisely targeted for digestion and self-ligation; in various reports, this process has been referred to as plasmid recovery,[100] plasmid rescue,[101] or iCATCH.[102] Editing of chromosomal sequences upstream and downstream of BGCs has also been used to introduce ΦBT1 or Cre-lox recombination sites to excise genomic loci via *in vitro* or *in vivo* recombination.[103-106] However, an important drawback to relying on genetic manipulation is that it is a labor-intensive approach that not all natural product producing organisms are amenable to, which may be the reason why cloning and heterologous expression were pursued for BGC characterization in the first place.

While still not trivial, cloning of large BGCs has become increasingly accessible, and many viable approaches have been developed and used to isolate DNA from complex genetic backgrounds. Following successful cloning, BGCs must be transferred to a heterologous host that can stably maintain and express the exogenous DNA. Furthermore, successful heterologous reconstitution also requires that the host is equipped with all biosynthetic building blocks and non-clustered enzymatic machinery or accessory factors essential for natural product production. In the next section, we highlight various host organisms that have been developed, optimized, and used for heterologous expression of natural product BGCs.

Table 1.1 Summary of BGC cloning methods and their associated advantages and disadvantages

		Examples	Advantages	Disadvantages	Ref.
Library-based	Cosmid/ fosmid	-	Sequence-independent, whole genome cloning	Untargeted; BGCs may be split across multiple library clones	5-14
	eDNA	-	Access BGCs from uncultured microbes	Unknown producer and final product; laborious screening necessary	15-21
	BAC/PAC	LEXAS	Large insert stability	Technically challenging	23-33
	FAC	-	Library can be readily screened in *A. nidulans*	Requires extensive screening, some false positives	34-37
Assembly	*in vitro*	Gibson, DiPAC, TPA, SIRA, SSRTA, iBrick	Technically easier; rapid and potential to be automated	Impractical or cumbersome for large BGCs	38-54
	in vivo	DNA assembler, ExRec, yTREX, AGOS	Can assemble many DNA fragments (10+)	Difficult to troubleshoot; risk of mutation	55-70
Direct cloning	TAR	-	Robust direct cloning of whole BGCs	Can be technically challenging; must use yeast	72-82, 84-86
	LLHR	ExoCET	Utilizes *E. coli* as a cloning host	Technically challenging	88-95
	in vitro	SSOA, Gibson, CATCH, plasmid rescue	Streamlined *in vitro* approach	Requires careful preparation and/or manipulation of gDNA	96-106

eDNA, environmental DNA; BAC, bacterial artificial chromosome; PAC, P1 artificial chromosome; FAC, fungal artificial chromosome; TAR, transformation-associated recombination; LLHR, linear-linear homologous recombination; LEXAS, library expression analysis system; DiPAC, direct pathway cloning; TPA, twin-primer assembly; SIRA, serine integrase recombinational assembly; SSRTA, site-specific recombination-based tandem assembly; ExRec, overlap extension PCR-yeast homologous recombination; yTREX, yeast recombinational cloning-enabled pathway transfer and expression tool; AGOS, artificial gene operon assembly system; ExoCET, exonuclease combined with RecET recombination; SSOA, single-strand overlapping annealing; CATCH, Cas9 assisted targeting of chromosome segments.

1.4 Hosts for heterologous expression

There are several obvious features of a good heterologous host organism: it grows fast, is genetically manipulatable, and is easy to work with in the laboratory. However, there are many additional traits that organisms must carry, either naturally or through engineering, to enable heterologous reconstitution of natural product BGCs. Some genera of host organisms have been developed and used more extensively than others, particularly *Streptomyces*, as this genus has been a naturally rich source of antibiotics and other small molecule natural products over the years. There is an underlying assumption, pervasive throughout the natural products research community, that BGCs are best expressed in organisms most closely related to their original source. Although this assumption is theoretically sound, it has not been rigorously proven, in part because it becomes completely irrelevant once a suitable host organism has been identified. Although it remains essentially impossible to predict host compatibility, educated guesses can be made, particularly as we understand more about biosynthetic mechanism and regulation. For all BGCs, there will be an assortment of additional elements that need to be "borrowed" from the host organism, ranging from biosynthetic precursors such as fatty acids, amino acids, and acyl-CoAs, to enzymes that post translationally modify biosynthetic enzymes such as phosphopantetheinyl transferases (PPTases), to regulatory genes that control BGC expression. These host elements must be compatible with exogenous BGCs or independently supplied from alternative sources. It has been shown that BGCs often move by horizontal gene transfer; thus, they are frequently not conserved throughout a taxonomic group.[107,108] Conversely, this means that similar BGCs can naturally function in organisms that are not closely related phylogenetically. Regardless of whether a more closely related host is truly "superior", it is ideal to have many different types of heterologous hosts available, particularly as we identify new BGCs from diverse microbial sources. Many studies have shown this empirically through successful expression of BGCs in some hosts but not others.[15,58,109-112]

Heterologous hosts can be engineered for enhanced expression of microbial BGCs in a background devoid of competing or contaminating pathways. General strategies for host optimization include genome minimization through deletion of native BGCs, nucleases, and proteases, as well as introduction of chromosomal integration elements, PPTases and other post translational modifying enzymes, or genes involved in precursor biosynthesis. In this section we review microbial heterologous host systems that have been developed, optimized, and used to successfully reconstitute natural product BGCs. Various bacterial and fungal genera are discussed (**Table 1.2**).

1.4.1 Actinobacteria

To date, actinobacterial hosts of the genus *Streptomyces* have been most widely used for BGC expression, primarily because many bioactive natural products have been isolated from *Streptomyces* and therefore the greatest attention has been paid to studying their secondary metabolism and developing optimized *Streptomyces* hosts. *Streptomyces* species that have been used as heterologous hosts include *S. coelicolor*,[113] *S. avermitilis*,[114,115] *S. lividans*,[116] *S. albus*,[117,118] *S. venezuelea*,[119] *S. ambofaciens* and *S. roseosporus*,[120] *S. flavogriseus*,[109,121] *S. chattanoogensis*,[122] and *S. chartreusis*.[123] Additionally, fast-growing and moderately thermophilic strains of *Streptomyces* have been isolated and tested for heterologous expression, although these organisms have not been characterized at the strain level.[124] *Streptomyces* hosts have been optimized to various extents using several general strategies. Genome minimization through curing of self-replicating plasmids and deletion of non-essential genes, including native BGCs, has been widely applied. Deletion of endogenous BGCs has the two-fold effect of removing sinks for biosynthetic precursors and simultaneously simplifying the host's chemical profile.[113-115,117,122] Additionally, empirically identified mutations in *rpoB* and *rpsL* (which encode the RNA polymerase β-subunit and ribosomal protein S12, respectively) of *S. coelicolor* and *S. lividans* pleiotropically enhance secondary metabolite production and have been introduced.[113] Remodeling of global

regulatory circuits through introduction of new positive regulators or up- and down-regulation of native positive and negative regulators, respectively, has also been leveraged,[116,118] in addition to expression of codon-optimized efflux pumps to reduce toxicity and aid in natural product purification.[116] Please note that the *Streptomyces* hosts and references listed in this section are not comprehensive, as in depth review of optimized *Streptomyces* hosts is covered elsewhere in this issue.

Long and often repetitive natural product BGCs must be stably maintained in the host, either as an autonomously replicating sequence or integrated into the host genome. Replicative expression plasmids such as pOJ446, which is an *E. coli-Streptomyces* shuttle vector based on the *S. coelicolor* A3(2) plasmid SCP2*, have been widely used.[125-127] However, chromosomal integration of BGCs is more stable and, if irreversible, circumvents the need to use antibiotics for plasmid maintenance. Many BGCs have been integrated into *Streptomyces* genomes leveraging recombinases from actinophages, most notably ΦC31.[128-130] The ΦC31 *attB* attachment site is naturally present in the genome of many *Streptomyces* species and enables site-specific integration of large BGCs into the host chromosome. Other actinophages, including TG1, SV1, ΦBT1, R4, ΦHau, and ΦJoe, have been identified, characterized, and in some cases used for BGC integration.[131-133] Having access to multiple attachment sites is helpful for performing complementation experiments to confirm that gene deletion mutants do not carry polar effects. Furthermore, BGCs can be split and integrated into multiple genomic loci to simplify cloning and assist in BGC engineering.[134,135] The precise site of genomic integration could be important in a heterologous expression experiment, as it has recently been shown that gene expression can vary up to eight-fold depending on the position of chromosomal integration in *S. albus*.[136] Thus, having access to multiple attachment sites leveraging different actinophages can be a valuable asset for heterologous expression of BGCs in *Streptomyces*.

ΦC31 attachment sites are also naturally present in the genomes of several other Actinobacteria and have been leveraged for BGC expression in the rare actinomycete

Nonomuraea sp. strain ATCC 39727[6,137] and the nocardioform actinomycete *Amycolatopsis japonicum*.[138] Alternatively, the marine actinomycete *Salinispora tropica* CNB-440, which possesses several pseudo ΦC31 attachment sites, was engineered to introduce an authentic ΦC31 attachment site for integration of BGCs into the salinosporamide (*sal*) biosynthetic locus, simultaneously abolishing salinosporamide production to free up biosynthetic precursors.[139] The resulting strain, *S. tropica* CNB-4401, represents the first marine actinomycete heterologous host and is readily compatible with expression vectors containing the ΦC31 integrase. Finally, *Saccharopolyspora erythraea*, the original producer of the antibiotic erythromycin, was used as a host for heterologous expression of the spinosad BGC, encoding production of spinosyns A (**7a**) and D (**7b**), from *Saccharopolyspora spinosa* via chromosomal integration by double-crossover homologous recombination.[140]

1.4.2 Firmicutes

The rod-shaped, low G+C content Firmicute *Bacillus subtilis* has many desirable qualities of a heterologous host organism. It produces the cyclic lipopeptide surfactin and thus harbors the promiscuous PPTase Sfp, which has been used to activate carrier proteins from diverse microbial sources.[141] Thus, *B. subtilis* is naturally equipped with an important element for expression of assembly line BGCs. Furthermore, it has the capacity for natural genetic competence and homologous recombination, making it easy to transform and genetically manipulate, and it grows relatively fast, particularly in comparison to *Streptomyces*. Despite these attractive features, *Bacillus* has been used relatively infrequently for heterologous expression of BGCs, perhaps due to a natural reluctance to test non-*Bacillus* pathways in a *Bacillus* host organism. Consequently, there has not been the same level of investment in development and optimization of *Bacillus* heterologous hosts. *B. subtilis* and *B. amyloliquefaciens* have been used to express a number of BGCs, primarily those encoding peptide products originating from other *Bacilli*.[73,82,142-147] In most cases, gene clusters have been integrated into the *amyE* locus of *B. subtilis* hosts via double-

crossover homologous recombination. Interestingly, Li *et al.* attempted to express the surfactin BGC in *B. subtilis* ROM77, in which the native surfactin locus is disrupted, from a self-replicating plasmid but encountered insurmountable plasmid instability, prompting the authors to build an integratable heterologous expression plasmid instead.[73] Beyond *Bacillus*, *Lactococcus lactis* is another Firmicute that has been leveraged for heterologous expression of small bacteriocin BGCs responsible for production of lactococcin Z and plantaricyclin A, which are produced by other *Lactococci* and *Lactobacilli*, respectively.[148,149]

1.4.3 Proteobacteria

Although the majority of isolated natural products have been sourced from Gram-positive Actinobacteria, Gram-negative organisms such as Proteobacteria and Cyanobacteria are increasingly being recognized as capable and underexplored producers of bioactive small molecules.[150] Gram-negative Proteobacteria include many dangerous pathogens such as *Salmonella*, *Vibrio*, and *Yersinia*, which are difficult to treat using standard antibiotics. Conversely, Proteobacteria are also important commensals found within the microbiomes of many higher organisms, including humans. Thus, heterologous expression of proteobacterial BGCs represents an important tool for studying secondary metabolism of microorganisms that greatly influence human health and disease.

The disparate GC-content between microorganisms has often been cited as a reason for host-BGC incompatibility, although robust evidence supporting this claim is lacking. More likely, host factors that must be borrowed for heterologous production, whether regulatory or biosynthetic, may be missing or incompatible between BGC and host. Currently, it is not clear whether Gram-positive hosts can robustly express Gram-negative BGCs or vice versa, as this has not been rigorously tested or reported. Regardless, many Gram-negative bacteria are easier to work with than *Streptomyces*, and thus expansion of heterologous expression platforms to include more proteobacterial hosts represents a valuable endeavor, particularly for those species

that grow fast, are easy to manipulate, and are more naturally suited to natural product production than *E. coli*.

Of course, the most obvious proteobacterial expression host is the laboratory workhorse *E. coli*, which belongs to the class of γ-proteobacteria. *E. coli* is very amenable to laboratory manipulation and, under ideal conditions, has a doubling time of only 20 minutes. Unfortunately, commonly used strains of *E. coli* are not naturally equipped to support expression of many natural product BGCs for several reasons. In some cases, specific pitfalls have been addressed and overcome. The luminmycin A (**8a**) BGC from the γ-proteobacterium *Photorhabdus luminescens* and the glidobactin A (**8b**) BGC from the β-proteobacterium *Burkholderia* DSM7029, both encoding hybrid NRPS/PKS pathways, have been expressed in the probiotic strain *E. coli* Nissle 1917 under the control of a tetracycline inducible promoter.[151,152] Transcriptional regulation appears to be a common challenge that must be overcome for BGC expression in laboratory strains of *E. coli;*[153] for this reason, *E. coli* BL21(DE3), equipped with the T7 RNA polymerase (RNAP), has been used to express pathways retrofitted with the orthogonal and inducible T7 promoter. BGCs tested include those originating from other Proteobacteria,[58,74,154] but importantly also include NRPS and PKS pathways from *Streptomyces*.[155,156] T7 is used extensively, particularly for *E. coli* protein expression, but is limited for BGC expression as it can only reliably promote transcripts up to 20 kb in length.[153] In BL21(DE3), T7 is controlled by arabinose induction, which is amplified in an engineered strain, BT2, through genetic disruption of the endogenous arabinose catabolism pathway.[157]

Post-translational activation of biosynthetic enzymes represents another common challenge for heterologous expression of BGCs in *E. coli*. As the native PPTases of *E. coli* are not able to effectively activate acyl or peptidyl carrier proteins involved in natural product biosynthesis,[158] *E. coli* BAP1, which harbors a genomically integrated copy of the Sfp PPTase from *B. subtilis*,[157] and GB05-MtaA, which harbors the MtaA PPTase from *Stigmatella aurantiaca*,[89] were generated. Use of these strains has proven essential for heterologous

expression of several assembly line biosynthetic pathways.[44,159-163] BAP1 is derived from BL21(DE3) and, in addition to Sfp, harbors a re-engineered *prp* operon to enhance supply of the polyketide precursor propionyl-CoA.[157] Additional genetic manipulations have been made in BAP1 to further enhance precursor supply for production of the antibiotic erythromycin A (**9**), a glycosylated polyketide macrolide. These manipulations include deletion of *ygfH*, resulting in generation of *E. coli* TB3 to further support propionyl-CoA supply,[164] and introduction of *pccAB*, resulting in generation of *E. coli* BTRAP for conversion of propionyl-CoA to methylmalonyl-CoA.[157] Finally, *E. coli* LF01 was generated from TB3 via sequential deletion of genes from pathways capable of siphoning carbon from deoxysugar biosynthesis to improve erythromycin glycosylation.[165] Although these manipulations were made specifically for enhanced heterologous erythromycin production, they demonstrate the potential versatility of *E. coli* to be engineered for optimal heterologous expression of any natural product BGC given a comprehensive understanding of biosynthetic processes.

Despite challenges associated with heterologous expression of assembly line biosynthetic pathways, laboratory strains of *E. coli* are generally well suited for production of ribosomally synthesized and post-translationally modified peptides, or RiPPs. RiPPs represent an important and growing class of natural products;[166] RiPP BGCs are also relatively small, making them tractable for heterologous expression and engineering. Recently, van der Donk, Tavassoli, and colleagues reported the generation of a vast library of 10^6 lanthipeptide analogs in *E. coli,* which was screened for inhibitory activity against a key protein-protein interaction involved in HIV infection.[167]

Because most heterologous expression experiments in *E. coli* rely on self-replicating plasmids as opposed to chromosomal integration, strains of *E. coli* have also been engineered for enhanced plasmid stability, for example *E. coli* BT3 and BTRA, which are also derivatives of BAP1.[157] Tools for chromosomal integration of large DNA fragments in *E. coli* have also been developed, including a Cre-lox based system called recombinase-assisted genome engineering,

or RAGE.[168,169] In theory, RAGE can be applied to a range of organisms to enable chromosomal integration of large BGCs for heterologous expression.

Beyond *E. coli*, the γ-proteobacterium *Pseudomonas* has been used extensively for BGC heterologous expression. *Pseudomonas* grows fast, is genetically manipulatable, and species such as *P. putida* are generally recognized as safe (GRAS). Furthermore, *Pseudomonas* appears to be more naturally suited to BGC expression than *E. coli*, as most *Pseudomonas* used for heterologous expression to-date have not been purposefully modified.[40,58,110,111,170-176] *P. putida* KT2440 harbors only a single PPTase, but *in vivo* testing revealed that this PPTase was capable of effectively activating carrier proteins from *S. coelicolor* and several Myxobacteria.[177] *P. putida* KT2440 has been metabolically engineered for methylmalonyl-CoA biosynthesis to enable heterologous production of PKS pathways that require this extender unit.[178] Furthermore, *P. putida* KT2440 has been modified more extensively through a series of genomic deletions to enhance its ability to serve as a chassis for gene expression, resulting in the generation of *P. putida* EM383, which displays superior growth properties and improved genetic stability.[179] Despite this enhanced stability, expression of large BGCs in *P. putida* from self-replicating plasmids is still not as reliable as chromosomal integration, and therefore robust methods for site-specific integration of large BGCs need to be further developed and optimized for this host. It is worth reiterating that, because *P. putida* grows fast and can be transformed quickly and easily using electroporation, testing of genetic variants using this heterologous host can be performed in a much shorter time frame than for even the fastest growing *Streptomyces*. Furthermore, as *P. putida* has demonstrated ability to activate *S. coelicolor* carrier proteins, it may be worthwhile to test *Streptomyces* BGCs in *Pseudomonas* heterologous hosts.

Outside of γ-proteobacteria, δ-proteobacteria of the order myxobacteria have also been leveraged as heterologous hosts. Myxobacteria are incredibly interesting bacteria that exhibit social behaviors, moving and feeding in multicellular groups called swarms. Furthermore, analogous to *Streptomyces* and *Aspergillus*, myxobacteria undergo complex morphological and

developmental changes during their life cycle, forming spore-producing structures, or fruiting bodies, under conditions of starvation.[180] Myxobacteria have large genomes (9-10 Mbp) relative to other bacteria and are prolific producers of natural products. Unfortunately, many species of myxobacteria do not make ideal expression hosts, as are difficult to work with and require extensive microbiological experience to culture and manipulate. Nevertheless, a few model myxobacteria have been successfully leveraged for expression of BGCs primarily from other myxobacteria, although the oxytetracycline (**10**) BGC from *Streptomyces rimosus* was also successfully reconstituted in a myxobacterial host.[181] *Myxococcus xanthus* DK1622 has been the most widely used myxobacterial host, and a strain of *M. xanthus* has been generated in which the native myxochromide A BGC has been inactivated in an attempt to improve heterologous production.[173,182] Single- or double- crossover homologous recombination has been used for site-specific chromosomal integration of BGCs;[181,183] however, this can be difficult, especially for large constructs. Transposition has also been used in both *M. xanthus*[111,171] and *Corallococcus macrosporus* GT-2,[184] a moderately thermophilic, faster growing myxobacteria. While more efficient than homologous recombination, integration via transposition, which is stochastic, can change host transcriptomes, leading to varying levels of heterologous compound production and preventing reproducibility between experiments.[173,185] This also makes it impossible to compare results between experiments, for example if BGCs are genetically manipulated and re-transferred to the heterologous host. Thus, Mx9 and Mx8 integrases, identified from bacteriophages, have also been successfully applied for site-specific BGC integration.[100,112,186] It is noteworthy that the pyxidicycline A (**11**) BGC from *Pyxidicoccus fallax* An d48, which was expressed in both *M. xanthus* and *Stigmatella aurantiaca* DW4/3-1, displayed distinct, host-specific production profiles, where different analogs were favored in the hosts and native producer.[186] This has also been observed for other BGCs and may indicate key differences in precursor availability, biosynthesis, or regulation across host organisms.[139]

Finally, various α- and β-proteobacterial hosts have been used for heterologous expression, albeit much more rarely. An eDNA library prepared using a broad-host-range cosmid vector was screened in the β-proteobacterium *Ralstonia metallidurans*, resulting in the identification of terpene and type III PKS products that could not be produced in *E. coli*.[187] Subsequently, broad-host-range eDNA clones were screened in a more diverse range of proteobacterial hosts, including the α-proteobacteria *Agrobacterium tumefaciens* and *Caulobacter vibrioides* and the β-proteobacterium *Burkholderia graminis*, in addition to *R. metallidurans*, *P. putida*, and *E. coli*.[110] eDNA expressed across these six hosts showed minimal overlap in host compatibility.[110] A similar broad-host-range approach was taken for expression of the violacein (**12**) BGC from the γ-proteobacterium *Pseudoalteromonas luteoviolacea* 2ta16 and demonstrated that although the α-proteobacterium *A. tumefaciens* was the least phylogenetically related host organism tested, it produced the greatest yield of violacein due to differences in BGC regulation.[58] Finally, *Rhodobacter capsulatus* was successfully engineered for expression of various pigment BGCs,[172] and *Burkholderida* DSM 7029 has been used for heterologous production of epothilone A (**13**) and vioprolide B (**14**), which both originate from myxobacteria.[173,188]

1.4.4 Cyanobacteria

Cyanobacteria are photosynthetic bacteria, found in freshwater and marine ecosystems, that have historically been a rich source of bioactive natural products and toxins.[189] The cyanobacterial phylum is diverse, but unfortunately most Cyanobacteria grow slowly and are difficult to genetically manipulate. Despite these challenges, recent advances have been made toward developing genetic tools and cyanobacterial hosts for heterologous expression of intransigent BGCs. A broad-host-range vector was constructed and tested in an array of “model” cyanobacterial hosts, including *Synechococcus elongatus* PCC7942, *Synechocystis* sp. PCC6803, *Anabaena* sp. PCC7120, *Leptolyngbya* sp. BL0902, and *Nostoc punctiforme*

ATCC29133.[190] More recently, *S. elongatus* PCC7942 was used for heterologous expression of polybrominated diphenyl ethers, through integration of PBDE biosynthetic genes, identified in sponge-derived metagenomic cyanobacterial DNA sequences, via homologous recombination into neutral sites in the *S. elongatus* genome.[191] Gene expression was driven by a synthetic promoter-riboswitch, providing precise control over protein production. *S. elongatus* PCC7942 has also been optimized specifically for polyketide production via introduction of modules that aid in PKS extender unit production, regulation (via the T7 RNAP), and post-translational modification (via the Sfp PPTase).[192] *S.* sp. PCC6803 was successfully engineered for photosynthetic overproduction of mycosporine-like amino acids (MAA) through promoter refactoring and co-expression of the *Anabaena* sp. PCC7120 PPTase (APPT).[193] APPT has been characterized and displays broad substrate scope and good catalytic efficiency against a range of cyanobacterial and *Streptomyces* carrier proteins.[194] *Anabaena* sp. PCC7120 has also been further explored for heterologous expression of BGCs using replicative plasmids, resulting in successful heterologous production of lyngbyatoxin A (**15**) from *Moorea producens* and mycosporine-like amino acids from *Nostoc flagelliforme*.[195,196]

1.4.5 Fungi

Filamentous fungi, like filamentous bacteria, produce many bioactive natural products and harbor scores of uncharacterized NRPS, PKS, and miscellaneous BGCs. However, many important differences underly bacterial and fungal transcription, translation, and basic cell biology. Therefore, it seems reasonable that identifying viable fungal hosts would be very important for heterologous expression of fungal BGCs. Fungal hosts carry more biological complexity than bacteria but in most cases are easy to cultivate, genetically manipulatable, and much more tractable than higher eukaryotes.

The baker's yeast *Sacchromyces cerevisiae* has been successfully used for expression of many biosynthetic genes and gene clusters from filamentous fungi.[197-207] Like *E. coli*, *S.*

cerevisiae is genetically tractable, grows quickly, and has been a longstanding laboratory workhorse and model organism. Also like *E. coli*, it is not a naturally gifted natural product producer. For this reason, *S. cerevisiae* has been outfitted with the NpgA PPTase from *Aspergillus nidulans* specifically for expression of NRPS and PKS genes.[208] *S. cerevisiae* has also undergone extensive metabolic engineering to increase PKS precursor supply, specifically acetyl-CoA.[197,198] The majority of heterologous expression experiments in *S. cerevisiae* have involved only a single or few individual genes expressed from replicative vectors,[199-204] although entire pathways involving more than three or four genes have also been reconstituted in their entirety.[205-207] Most BGCs expressed in *S. cerevisiae* to date originate from other fungi, but it has also been used to reconstitute the bacterial NRPS BpsA from *Streptomyces lavendulae*.[209] Recently, HEx (Heterologous EXpression) was established as a high-throughput platform for expression of fungal BGCs in *S. cerevisiae*.[210] Using this platform, 41 orphan BGCs from diverse fungal species were interrogated, resulting in 22 detectable compounds.[210] Finally, new methods for BGC integration into yeast chromosomes are being developed, for example a method called delta integration CRISPR-Cas, or Di-CRISPR.[211]

Despite many successes, there are significant differences between intron processing in filamentous fungi and yeast.[212] Although intron sequences can be effectively removed by cloning from cDNA, this requires RNA isolation and also makes it difficult to clone and express multi-gene pathways. Thus, filamentous fungal hosts have also been developed. *Aspergillus nidulans* is a model organism that can be genetically manipulated and is highly recombinant; it has been used extensively for expression of BGCs from *Aspergilli* and other ascomycete fungi.[34-37,213-216] *A. nidulans* has been engineered through deletion of endogenous BGCs, and characterized promoters can be leveraged in this host.[217] *Aspergillus oryzae* has been successfully used for heterologous reconstitution of the antibiotic pleuromutilin (**16**), a diterpene produced by the basidiomycete *Clitopilus passeckerianus*.[218] Finally, *Fusarium heterosporum* ATCC 74349, an ascomycete that produces high levels of the natural product equisetin,[219] has been leveraged to

express BGCs from *Aspergillus flavus*, *Aspergillus terreus*, *Hapsidospora irregularis*, and an uncharacterized fern endophyte isolated from Papua New Guinea.[220-222] Furthermore, fusions of fungal PKS/NRPS genes were tested in this host strain to determine guidelines for engineering hybrid PKS/NRPS pathways.[223]

Table 1.2 Summary of heterologous hosts used for BGC expression.

	Phylum	Genus	Details	Ref.
Bacteria				
Gram-positive				
	Actino-bacteria			
		Streptomyces	Most widely used host, extensively optimized; site-specific BGC integration catalyzed by actinophage recombinases	109, 113-136
		Nonomuraea	All compatible with ΦC31 integrase;	6, 137
		Amycolatopsis	engineered strain *S. tropica* CNB-4401	138
		Salinispora	represents first marine actinomycete host	139
		Saccharo-polyspora	*S. erythraea*, native erythromycin producer, engineered for spinosad production	140
	Firmicutes			
		Bacillus	Harbors Sfp PPTase, easily transformed and integrated with BGCs but underutilized	73, 82, 142-147
		Lactococcus	Leveraged for expression of small bacteriocin BGCs	148-149
Gram-negative				
	Proteo-bacteria			
	γ	*Escherichia*	Strains include Nissle 1917, BL21(DE3), BAP1, GB05-MtaA; also optimized for precursor flux and genetic stability	44, 58, 74, 89, 151-167
		Pseudomonas	Talented host; potential for expression of myxobacterial and *Streptomyces* BGCs	40, 58, 110-111, 170-178
	δ	*Myxococcus*	*M. xanthus, S. aurantiaca* leveraged for expression of many myxobacterial BGCs using transposition and phage integration;	99, 111-112, 171, 173, 181-183, 185-186
		Stigmatella	can be difficult to work with and grow slowly	186
		Corallococcus	*C. macrosporus* is moderately thermophilic	184
	α	*Agrobacterium*	Leveraged for expression of violacein,	58, 110
		Caulobacter	zeaxanthin, and prodigiosin pigments,	110
		Rhodobacter	various eDNA library clones	172
	β	*Ralstonia*	Used for expression of carotenoid, type III	110, 187
		Burkholderia	PKS products from eDNA; also epothilones and vioprolides	110, 173, 188

Table 1.2 Summary of heterologous hosts used for BGC expression, continued.

	Phylum	Genus	Details	Ref.
Bacteria				
	Cyano-bacteria			
		Synechococcus	*S. elongatus* PCC7942, *S*. sp. PCC6803	190-193
		Anabaena	*Anabaena* sp. PCC7120 used for expression of cyanobacterial pathways	190, 195-196
Fungi				
		Saccharomyces	HEx platform developed for high-throughput functional evaluation of fungal BGCs	197-211
		Aspergillus *Fusarium*	More naturally suited for expression of BGCs from filamentous fungi	34-37, 213-218 220-223

Heterologous expression of microbial BGCs, enabled by the diverse range of hosts reviewed in this section, has greatly expanded our knowledge of natural product biosynthesis, bioactivity, and regulation. While it is worthwhile to develop new and diverse host organisms, current platforms are not designed to be broadly applicable, as dual cloning and expression vectors outfitted with host-specific elements make it difficult to test expression of cloned BGCs in different hosts, particularly across bacterial and fungal phyla. More universal systems would greatly streamline the ability to test BGC expression across multiple hosts. One potential approach involves introduction of modular expression elements (ideally for site-specific chromosomal integration) after cloning, enabling vectors to be easily retrofitted for expression testing across different hosts (**Figure 1.4**). Alternatively, diverse hosts can be engineered to carry the same BGC integration site, such that a single vector system is broadly compatible with many different expression hosts (**Figure 1.4**). This could be achieved, for example, leveraging a system like recombinase-assisted genome engineering (RAGE) to introduce the same attachment site in diverse hosts organisms.[169]

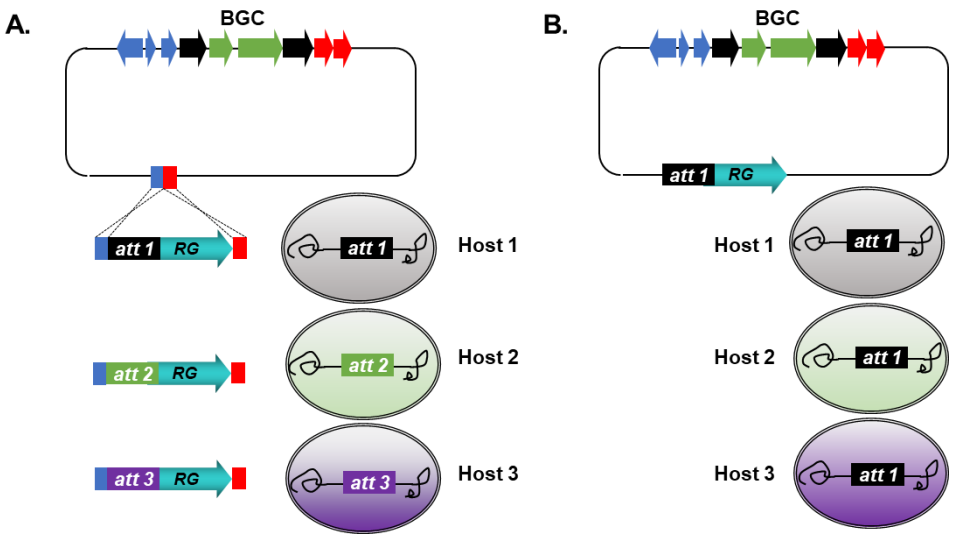


Figure 1.4 Streamlined design of universal BGC expression platforms. (A) Modular expression vector system, leveraging knock-in of modular expression elements for BGC integration across many hosts. (B) Single vector system, compatible across several engineered hosts. *RG*, resistance gene; *att*, attachment site.

1.5 Genetic manipulation of cloned BGCs

Cloned constructs can be quickly edited leveraging *E. coli* recombineering tools such as PCR targeting and λ-Red recombination.[224,225] Therefore, cloning of BGCs also enables rapid genetic manipulation of biosynthetic pathways through gene knock-in or knock-out experiments. Edited constructs can then be re-introduced to heterologous hosts to test for expression. In this manner, BGCs can be engineered for activation or enhanced production of chemical products,[226] generation of new analogs,[227] or interrogation of biosynthetic mechanism.[228,229] This process can be iterated to quickly connect biosynthetic genotype and chemotype. Standard *E. coli* recombineering leveraging antibiotic resistance cassettes makes most gene deletion and insertion experiments routine, although care should be taken to avoid polar effects in bacteria. Alternatively, yeast can also be used, leveraging natural recombination and various auxotrophic markers or resistance cassettes, for deletion or insertion.[72] This is particularly attractive for large, highly repetitive BGCs cloned into high copy number *E. coli* vectors, since recombineering strains of *E. coli* are inherently less genetically stable and could catalyze unwanted plasmid

recombination resulting in BGC rearrangements or deletions. However, yeast manipulations take longer and are only possible with vectors containing yeast origins of replication and selection markers.

Selection markers flanked by FRT (flippase recognition target) sequences can be removed using Flp (flippase) recombinase-mediated excision, leaving behind a small scar sequence. This adds an additional step, and clones must be carefully screened since there is no mechanism to select for excision. Recently, marker-less deletions were successfully made *in vitro* using CRISPR-Cas9 in a method termed ICE (*in vitro* CRISPR/Cas9 mediated editing).[230] While this report simplifies the process of generating marker-less deletion mutants, it remains significantly challenging to make subtle marker-less changes, for example point mutations, within large DNA constructs, as these types of manipulations would usually be performed using PCR mutagenesis. One potential application for improved marker-less mutation methods is for editing assembly line BGCs, as these generally involve large, multigene pathways. Recently, we developed a method for facile generation of point mutations in large cloned constructs, which paves the way for easy editing of large BGCs in the future.[231]

1.6 Conclusions and future perspectives

With advances in DNA sequencing technology and bioinformatics tools, researchers have identified huge untapped microbial biosynthetic potential *in silico*, and genomics-based approaches have become routine for natural product discovery. The diverse strategies summarized in this review are advancing our ability to leverage microbial genome sequence information for the purpose of discovering new metabolites and understanding biosynthetic logic.

Over the past few decades, hundreds of natural product BGCs have been cloned and expressed in various heterologous hosts, the vast majority associated with known natural products (estimate based on analysis of the MIBiG database).[22] Thus, the next frontier lies in eliciting expression of so-called “silent” or “cryptic” BGCs to identify their chemical products and

understand their biological activity. One promising approach has been the swapping of native promoters of “silent” BGCs with well-characterized constitutive or inducible promoters. This so-called “promoter refactoring’ strategy is particularly compatible with heterologous expression platforms, in which undetermined native transcriptional regulation systems can be replaced with well-characterized and controllable regulatory elements. While this approach has garnered significant interest and enthusiasm,[57,232-236] current cloning and editing methods make it impractical to perform large-scale BGC refactoring, requiring substantial investments in time, energy, and money. In the future, we envision that DNA synthesis will play an outsized role in obtaining BGCs for functional characterization. Although currently costly and impractical for large, repetitive, and high GC-content BGCs, a small number of studies have already used *de novo* DNA synthesis to access and refactor BGCs less than 10 kb in size.[237-239] With increasing demand for synthetic DNA driven by the field of synthetic biology,[240] technical advances will likely spur decreased cost and enhanced capabilities of large-scale *de novo* DNA synthesis.[241] Conceivably, this will shift the natural product genome mining paradigm from an empirical “clone-edit-test” workflow to a streamlined and hypothesis-driven “design-build-test” workflow (**Figure 1.5**). *De novo* DNA synthesis will enable experimental characterization of numerous BGCs identified from metagenomic sequence datasets, which represent the largest genomic reservoir for small molecule discovery and currently remains largely untapped.[242] Furthermore, establishing various synthetic transcriptional regulatory elements, including promoter sequences, ribosome binding sites, and terminator sequences, will expand the toolbox for BGC re-design and even enable generation of artificial biosynthetic pathways.[243-249] Finally, multi-omics techniques, including genomics, transcriptomics, proteomics, and metabolomics, can be integrated within heterologous expression platforms to diagnose production bottlenecks in a systematic manner.[250]

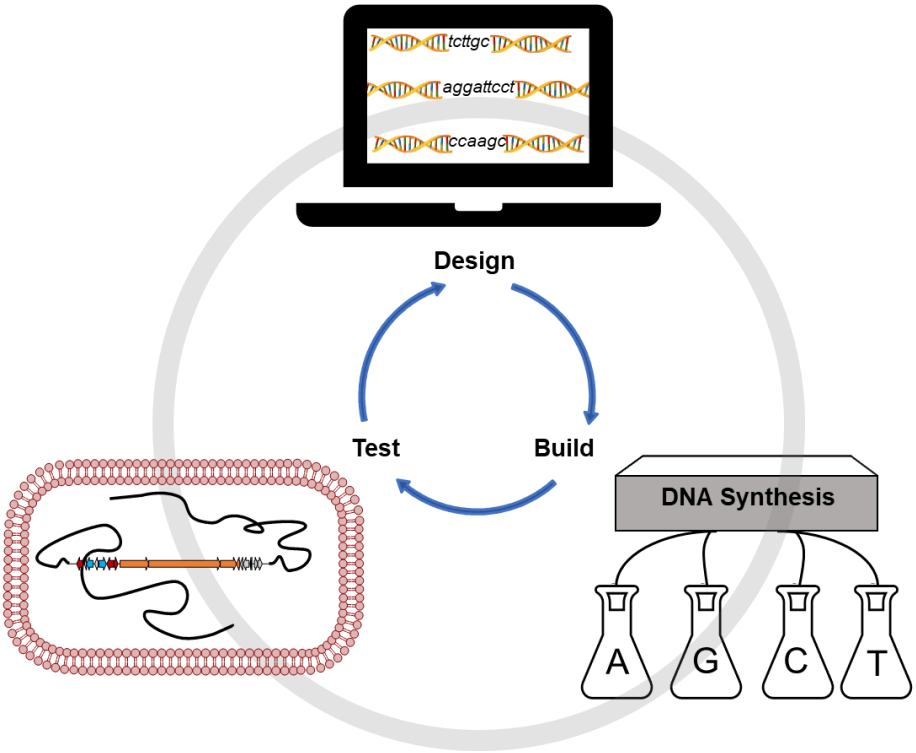


Figure 1.5 “Design, build, test” workflow for future natural product genome mining.

One hurdle BGC re-design cannot overcome is missing or incompatible “host factors” or biosynthetic elements essential for heterologous reconstitution of natural product production, especially when working with unconventional biosynthetic pathways where we may not know that a missing factor exists at all. As mentioned previously, current host selection and testing is limited and largely based on the assumption that the ideal heterologous host is most closely phylogenetically related to the native producing organism. However, this hypothesis has not been rigorously tested and there is limited empirical evidence to support this notion, particularly for “silent” or “cryptic” BGCs. Expanding the spectrum of heterologous hosts that can be tested for BGC expression in a streamlined manner will likely enhance our ability to characterize novel BGCs. Beyond phylogeny, other biological factors such as down-scale cultivability, growth speed, transformation efficiency, and genetic amenability should also be considered in terms of the development of new heterologous hosts with potential for high-throughput screening campaigns.

Finally, hosts extending into vastly distant biological domains, including archaea[251] and eukaryotes such as *Chlamydomonas reinhardtii*,[252] a green microalga, and *Nicotiana benthamiana*,[253] a wild relative of tobacco, will hopefully experience greater development and use as we continue to explore natural product biosynthesis across the tree of life.

Briefly, although this review focuses on heterologous expression-based genome mining strategies, we would also like to direct readers to a small number of studies that have pioneered the quest for mining natural products from native producers, including genetic and chemical induction-based approaches.[254-257] Furthermore, cell-free natural product biosynthesis, also referred to as enzymatic total synthesis or *in vitro* biosynthesis, has become a powerful approach to truly test our understanding of biosynthetic mechanism.[258-263] Cell-free methods also hold potential for obtaining scalable quantities of valuable bio-chemicals.[264,265]

Over the past few decades, we have witnessed a renaissance within the field of natural products research, reinvigorated by the development of genomics-guided identification of microbial BGCs. The cloning methods, heterologous hosts, and genetic editing tools cited in this review will continue to be developed and utilized for BGC characterization. We remain in the infancy of investigating natural product biosynthesis and biological activity, with many challenges still to be addressed and even greater discoveries yet to be made.

1.7 Acknowledgement

Chapter 1, in part, has been submitted for publication of the material as it may appear in *Natural Product Reports*, Zhang, J.J.; Tang, X.; Moore, B. S., 2019. The dissertation author was the primary author of this material.

1.8 References

1 Nicolaou, K. C. & Rigol, S. A brief history of antibiotics and select advances in their synthesis. *J. Antibiot. (Tokyo)* **71,** 153-184, doi:10.1038/ja.2017.62 (2018).

2 Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79,** 629-661, doi:10.1021/acs.jnatprod.5b01055 (2016).

3 Heitman, J., Movva, N. R. & Hall, M. N. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science* **253,** 905-909 (1991).

4 Sabatini, D. M. Twenty-five years of mTOR: Uncovering the link from nutrients to growth. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 11818-11825, doi:10.1073/pnas.1716173114 (2017).

5 Flinspach, K., Westrich, L., Kaysser, L., Siebenberg, S., Gomez-Escribano, J. P., Bibb, M., Gust, B. & Heide, L. Heterologous expression of the biosynthetic gene clusters of coumermycin A(1), clorobiocin and caprazamycins in genetically modified Streptomyces coelicolor strains. *Biopolymers* **93,** 823-832, doi:10.1002/bip.21493 (2010).

6 Foulston, L. C. & Bibb, M. J. Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13461-13466, doi:10.1073/pnas.1008285107 (2010).

7 Kaysser, L., Tang, X., Wemakor, E., Sedding, K., Hennig, S., Siebenberg, S. & Gust, B. Identification of a napsamycin biosynthesis gene cluster by genome mining. *Chembiochem* **12,** 477-487, doi:10.1002/cbic.201000460 (2011).

8 Yamanaka, K., Ryan, K. S., Gulder, T. A., Hughes, C. C. & Moore, B. S. Flavoenzyme-catalyzed atropo-selective N,C-bipyrrole homocoupling in marinopyrrole biosynthesis. *J. Am. Chem. Soc.* **134,** 12434-12437, doi:10.1021/ja305670f (2012).

9 Pait, I. G. U., Kitani, S., Roslan, F. W., Ulanova, D., Arai, M., Ikeda, H. & Nihira, T. Discovery of a new diol-containing polyketide by heterologous expression of a silent biosynthetic gene cluster from Streptomyces lavendulae FRI-5. *J. Ind. Microbiol. Biotechnol.* **45,** 77-87, doi:10.1007/s10295-017-1997-x (2018).

10 Kaysser, L., Bernhardt, P., Nam, S. J., Loesgen, S., Ruby, J. G., Skewes-Cox, P., Jensen, P. R., Fenical, W. & Moore, B. S. Merochlorins A-D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases. *J. Am. Chem. Soc.* **134,** 11988-11991, doi:10.1021/ja305665f (2012).

11 Mantovani, S. M. & Moore, B. S. Flavin-linked oxidase catalyzes pyrrolizine formation of dichloropyrrole-containing polyketide extender unit in chlorizidine A. *J. Am. Chem. Soc.* **135,** 18032-18035, doi:10.1021/ja409520v (2013).

12 Schorn, M., Zettler, J., Noel, J. P., Dorrestein, P. C., Moore, B. S. & Kaysser, L. Genetic basis for the biosynthesis of the pharmaceutically important class of epoxyketone proteasome inhibitors. *ACS Chem. Biol.* **9,** 301-309, doi:10.1021/cb400699p (2014).

13 Leipoldt, F., Santos-Aberturas, J., Stegmann, D. P., Wolf, F., Kulik, A., Lacret, R., Popadic, D., Keinhorster, D., Kirchner, N., Bekiesch, P., Gross, H., Truman, A. W. & Kaysser, L. Warhead biosynthesis and the origin of structural diversity in hydroxamate metalloproteinase inhibitors. *Nat Commun* **8,** 1965, doi:10.1038/s41467-017-01975-6 (2017).

14 Wolf, F., Leipoldt, F., Kulik, A., Wibberg, D., Kalinowski, J. & Kaysser, L. Characterization of the Actinonin Biosynthetic Gene Cluster. *Chembiochem* **19,** 1189-1195, doi:10.1002/cbic.201800116 (2018).

15 McMahon, M. D., Guan, C., Handelsman, J. & Thomas, M. G. Metagenomic analysis of Streptomyces lividans reveals host-dependent functional expression. *Appl. Environ. Microbiol.* **78,** 3622-3629, doi:10.1128/AEM.00044-12 (2012).

16 Owen, J. G., Charlop-Powers, Z., Smith, A. G., Ternei, M. A., Calle, P. Y., Reddy, B. V., Montiel, D. & Brady, S. F. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 4221-4226, doi:10.1073/pnas.1501124112 (2015).

17 Iqbal, H. A., Low-Beinart, L., Obiajulu, J. U. & Brady, S. F. Natural Product Discovery through Improved Functional Metagenomics in Streptomyces. *J. Am. Chem. Soc.* **138,** 9341-9344, doi:10.1021/jacs.6b02921 (2016).

18 Bitok, J. K., Lemetre, C., Ternei, M. A. & Brady, S. F. Identification of biosynthetic gene clusters from metagenomic libraries using PPTase complementation in a Streptomyces host. *FEMS Microbiol. Lett.* **364,** fnx155, doi:10.1093/femsle/fnx155 (2017).

19 Hover, B. M., Kim, S. H., Katz, M., Charlop-Powers, Z., Owen, J. G., Ternei, M. A., Maniko, J., Estrela, A. B., Molina, H., Park, S., Perlin, D. S. & Brady, S. F. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol* **3,** 415-422, doi:10.1038/s41564-018-0110-1 (2018).

20 Peek, J., Lilic, M., Montiel, D., Milshteyn, A., Woodworth, I., Biggins, J. B., Ternei, M. A., Calle, P. Y., Danziger, M., Warrier, T., Saito, K., Braffman, N., Fay, A., Glickman, M. S., Darst, S. A., Campbell, E. A. & Brady, S. F. Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat Commun* **9,** 4147, doi:10.1038/s41467-018-06587-2 (2018).

21 Feng, Z., Kim, J. H. & Brady, S. F. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J. Am. Chem. Soc.* **132,** 11902-11903, doi:10.1021/ja104550p (2010).

22 Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Dusterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kotter, P., Krug, D., Masschelein, J., Melnik, A. V., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nutzmann, H. W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona-Gomez, F., Bechthold, A., Bode, H. B., Borriss, R., Brady, S. F., Brakhage, A. A., Caffrey, P., Cheng, Y. Q., Clardy, J., Cox, R. J., De Mot, R., Donadio, S., Donia, M. S., van der Donk, W. A., Dorrestein, P. C., Doyle, S., Driessen, A. J., Ehling-Schulz, M., Entian, K. D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Hofte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N.

C., Kwon, H. J., Lautru, S., Lavigne, R., Lee, C. Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Mendez, C., Metsa-Ketela, M., Micklefield, J., Mitchell, D. A., Moore, B. S., Moreira, L. M., Muller, R., Neilan, B. A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M. S., Ostash, B., Payne, S. M., Pernodet, J. L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Sussmuth, R. D., Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J. M., Wohlleben, W., Wright, G. D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E. & Glockner, F. O. Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **11,** 625-631, doi:10.1038/nchembio.1890 (2015).

23 Tu, Q., Herrmann, J., Hu, S., Raju, R., Bian, X., Zhang, Y. & Muller, R. Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering. *Sci. Rep.* **6,** 21066, doi:10.1038/srep21066 (2016).

24 Xu, M., Wang, Y., Zhao, Z., Gao, G., Huang, S. X., Kang, Q., He, X., Lin, S., Pang, X., Deng, Z. & Tao, M. Functional Genome Mining for Metabolites Encoded by Large Gene Clusters through Heterologous Expression of a Whole-Genome Bacterial Artificial Chromosome Library in Streptomyces spp. *Appl. Environ. Microbiol.* **82,** 5795-5805, doi:10.1128/AEM.01383-16 (2016).

25 Deng, Q., Zhou, L., Luo, M., Deng, Z. & Zhao, C. Heterologous expression of Avermectins biosynthetic gene cluster by construction of a Bacterial Artificial Chromosome library of the producers. *Synth Syst Biotechnol* **2,** 59-64, doi:10.1016/j.synbio.2017.03.001 (2017).

26 Crusemann, M., Reher, R., Schamari, I., Brachmann, A. O., Ohbayashi, T., Kuschak, M., Malfacini, D., Seidinger, A., Pinto-Carbo, M., Richarz, R., Reuter, T., Kehraus, S., Hallab, A., Attwood, M., Schioth, H. B., Mergaert, P., Kikuchi, Y., Schaberle, T. F., Kostenis, E., Wenzel, D., Muller, C. E., Piel, J., Carlier, A., Eberl, L. & Konig, G. M. Heterologous Expression, Biosynthetic Studies, and Ecological Function of the Selective Gq-Signaling Inhibitor FR900359. *Angew. Chem. Int. Ed. Engl.* **57,** 836-840, doi:10.1002/anie.201707996 (2018).

27 Liu, H., Jiang, H., Haltli, B., Kulowski, K., Muszynska, E., Feng, X., Summers, M., Young, M., Graziani, E., Koehn, F., Carter, G. T. & He, M. Rapid cloning and heterologous expression of the meridamycin biosynthetic gene cluster using a versatile Escherichia coli-streptomyces artificial chromosome vector, pSBAC. *J. Nat. Prod.* **72,** 389-395, doi:10.1021/np8006149 (2009).

28 Hashimoto, T., Hashimoto, J., Kozone, I., Amagai, K., Kawahara, T., Takahashi, S., Ikeda, H. & Shin-Ya, K. Biosynthesis of Quinolidomicin, the Largest Known Macrolide of Terrestrial Origin: Identification and Heterologous Expression of a Biosynthetic Gene Cluster over 200 kb. *Org Lett* **20,** 7996-7999, doi:10.1021/acs.orglett.8b03570 (2018).

29 Jones, A. C., Gust, B., Kulik, A., Heide, L., Buttner, M. J. & Bibb, M. J. Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PLoS One* **8,** e69319, doi:10.1371/journal.pone.0069319 (2013).

30 Castro, J. F., Razmilic, V., Gomez-Escribano, J. P., Andrews, B., Asenjo, J. A. & Bibb, M. J. Identification and Heterologous Expression of the Chaxamycin Biosynthesis Gene

Cluster from Streptomyces leeuwenhoekii. *Appl. Environ. Microbiol.* **81,** 5820-5831, doi:10.1128/AEM.01039-15 (2015).

31 Qin, Z., Munnoch, J. T., Devine, R., Holmes, N. A., Seipke, R. F., Wilkinson, K. A., Wilkinson, B. & Hutchings, M. I. Formicamycins, antibacterial polyketides produced by Streptomyces formicae isolated from African Tetraponera plant-ants. *Chem. Sci.* **8,** 3218-3227, doi:10.1039/c6sc04265a (2017).

32 Tu, J., Li, S., Chen, J., Song, Y., Fu, S., Ju, J. & Li, Q. Characterization and heterologous expression of the neoabyssomicin/abyssomicin biosynthetic gene cluster from Streptomyces koyangensis SCSIO 5802. *Microb Cell Fact* **17,** 28, doi:10.1186/s12934-018-0875-1 (2018).

33 Kepplinger, B., Morton-Laing, S., Seistrup, K. H., Marrs, E. C. L., Hopkins, A. P., Perry, J. D., Strahl, H., Hall, M. J., Errington, J. & Allenby, N. E. E. Mode of Action and Heterologous Expression of the Natural Product Antibiotic Vancoresmycin. *ACS Chem. Biol.* **13,** 207-214, doi:10.1021/acschembio.7b00733 (2018).

34 Bok, J. W., Ye, R., Clevenger, K. D., Mead, D., Wagner, M., Krerowicz, A., Albright, J. C., Goering, A. W., Thomas, P. M., Kelleher, N. L., Keller, N. P. & Wu, C. C. Fungal artificial chromosomes for mining of the fungal secondary metabolome. *BMC Genomics* **16,** 343, doi:10.1186/s12864-015-1561-x (2015).

35 Clevenger, K. D., Bok, J. W., Ye, R., Miley, G. P., Verdan, M. H., Velk, T., Chen, C., Yang, K., Robey, M. T., Gao, P., Lamprecht, M., Thomas, P. M., Islam, M. N., Palmer, J. M., Wu, C. C., Keller, N. P. & Kelleher, N. L. A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nat. Chem. Biol.* **13,** 895-901, doi:10.1038/nchembio.2408 (2017).

36 Robey, M. T., Ye, R., Bok, J. W., Clevenger, K. D., Islam, M. N., Chen, C., Gupta, R., Swyers, M., Wu, E., Gao, P., Thomas, P. M., Wu, C. C., Keller, N. P. & Kelleher, N. L. Identification of the First Diketomorpholine Biosynthetic Pathway Using FAC-MS Technology. *ACS Chem. Biol.* **13,** 1142-1147, doi:10.1021/acschembio.8b00024 (2018).

37 Clevenger, K. D., Ye, R., Bok, J. W., Thomas, P. M., Islam, M. N., Miley, G. P., Robey, M. T., Chen, C., Yang, K., Swyers, M., Wu, E., Gao, P., Wu, C. C., Keller, N. P. & Kelleher, N. L. Interrogation of Benzomalvin Biosynthesis Using Fungal Artificial Chromosomes with Metabolomic Scoring (FAC-MS): Discovery of a Benzodiazepine Synthase Activity. *Biochemistry* **57,** 3237-3243, doi:10.1021/acs.biochem.8b00076 (2018).

38 Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6,** 343-345, doi:10.1038/nmeth.1318 (2009).

39 Linares-Otoya, L., Linares-Otoya, V., Armas-Mantilla, L., Blanco-Olano, C., Crusemann, M., Ganoza-Yupanqui, M. L., Campos-Florian, J., Konig, G. M. & Schaberle, T. F. Identification and heterologous expression of the kocurin biosynthetic gene cluster. *Microbiology* **163,** 1409-1414, doi:10.1099/mic.0.000538 (2017).

40 Vior, N. M., Lacret, R., Chandra, G., Dorai-Raj, S., Trick, M. & Truman, A. W. Discovery and Biosynthesis of the Antibiotic Bicyclomycin in Distantly Related Bacterial Classes. *Appl. Environ. Microbiol.* **84**, doi:10.1128/AEM.02828-17 (2018).

41 Mevaere, J., Goulard, C., Schneider, O., Sekurova, O. N., Ma, H., Zirah, S., Afonso, C., Rebuffat, S., Zotchev, S. B. & Li, Y. An orthogonal system for heterologous expression of actinobacterial lasso peptides in Streptomyces hosts. *Sci. Rep.* **8,** 8232, doi:10.1038/s41598-018-26620-0 (2018).

42 Cai, X., Challinor, V. L., Zhao, L., Reimer, D., Adihou, H., Grun, P., Kaiser, M. & Bode, H. B. Biosynthesis of the Antibiotic Nematophin and Its Elongated Derivatives in Entomopathogenic Bacteria. *Org Lett* **19,** 806-809, doi:10.1021/acs.orglett.6b03796 (2017).

43 Yurkovich, M. E., Jenkins, R., Sun, Y., Tosin, M. & Leadlay, P. F. The polyketide backbone of thiolactomycin is assembled by an unusual iterative polyketide synthase. *Chem. Commun. (Camb.)* **53,** 2182-2185, doi:10.1039/c6cc09934c (2017).

44 Greunke, C., Duell, E. R., D'Agostino, P. M., Glockle, A., Lamm, K. & Gulder, T. A. M. Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab Eng* **47,** 334-345, doi:10.1016/j.ymben.2018.03.010 (2018).

45 D'Agostino, P. M. & Gulder, T. A. M. Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth Biol* **7,** 1702-1708, doi:10.1021/acssynbio.8b00151 (2018).

46 Duell, E. R., D'Agostino, P. M., Shapiro, N., Woyke, T., Fuchs, T. M. & Gulder, T. A. M. Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in E. coli. *Microb Cell Fact* **18,** 32, doi:10.1186/s12934-019-1080-6 (2019).

47 Liang, J., Liu, Z., Low, X. Z., Ang, E. L. & Zhao, H. Twin-primer non-enzymatic DNA assembly: an efficient and accurate multi-part DNA assembly method. *Nucleic Acids Res.* **45,** e94, doi:10.1093/nar/gkx132 (2017).

48 Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., Keasling, J. D. & Rosser, S. J. Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.* **42,** e23, doi:10.1093/nar/gkt1101 (2014).

49 Olorunniji, F. J., Merrick, C., Rosser, S. J., Smith, M. C. M., Stark, W. M. & Colloms, S. D. Multipart DNA Assembly Using Site-Specific Recombinases from the Large Serine Integrase Family. *Methods Mol. Biol.* **1642,** 303-323, doi:10.1007/978-1-4939-7169-5_19 (2017).

50 Zhang, L., Ou, X., Zhao, G. & Ding, X. Highly efficient in vitro site-specific recombination system based on streptomyces phage phiBT1 integrase. *J. Bacteriol.* **190,** 6392-6397, doi:10.1128/JB.00777-08 (2008).

51 Zhang, L., Zhao, G. & Ding, X. Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. *Sci. Rep.* **1,** 141, doi:10.1038/srep00141 (2011).

52 Wang, X., Tang, B., Ye, Y., Mao, Y., Lei, X., Zhao, G. & Ding, X. Bxb1 integrase serves as a highly efficient DNA recombinase in rapid metabolite pathway assembly. *Acta Biochim Biophys Sin (Shanghai)* **49,** 44-50, doi:10.1093/abbs/gmw115 (2017).

53 Liu, J. K., Chen, W. H., Ren, S. X., Zhao, G. P. & Wang, J. iBrick: a new standard for iterative assembly of biological parts with homing endonucleases. *PLoS One* **9,** e110852, doi:10.1371/journal.pone.0110852 (2014).

54 Enghiad, B. & Zhao, H. Programmable DNA-Guided Artificial Restriction Enzymes. *ACS Synth Biol* **6,** 752-757, doi:10.1021/acssynbio.6b00324 (2017).

55 Kim, J. H., Feng, Z., Bauer, J. D., Kallifidas, D., Calle, P. Y. & Brady, S. F. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers* **93,** 833-844, doi:10.1002/bip.21450 (2010).

56 Bilyk, O., Sekurova, O. N., Zotchev, S. B. & Luzhetskyy, A. Cloning and Heterologous Expression of the Grecocycline Biosynthetic Gene Cluster. *PLoS One* **11,** e0158682, doi:10.1371/journal.pone.0158682 (2016).

57 Bauman, K. D., Li, J., Murata, K., Mantovani, S. M., Dahesh, S., Nizet, V., Luhavaya, H. & Moore, B. S. Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry. *Cell Chem Biol* **26,** 724-736, doi:10.1016/j.chembiol.2019.02.004 (2019).

58 Zhang, J. J., Tang, X., Zhang, M., Nguyen, D. & Moore, B. S. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. *MBio* **8,** e01291, doi:10.1128/mBio.01291-17 (2017).

59 Shi, Y., Jiang, Z., Li, X., Zuo, L., Lei, X., Yu, L., Wu, L., Jiang, J. & Hong, B. Biosynthesis of antibiotic chuangxinmycin from Actinoplanes tsinanensis. *Acta Pharm Sin B* **8,** 283-294, doi:10.1016/j.apsb.2017.07.005 (2018).

60 Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37,** e16, doi:10.1093/nar/gkn991 (2009).

61 Shao, Z., Luo, Y. & Zhao, H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* **7,** 1056-1059, doi:10.1039/c0mb00338g (2011).

62 Shao, Z., Luo, Y. & Zhao, H. DNA assembler method for construction of zeaxanthin-producing strains of Saccharomyces cerevisiae. *Methods Mol. Biol.* **898,** 251-262, doi:10.1007/978-1-61779-918-1_17 (2012).

63 Shao, Z. & Zhao, H. DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. *Methods Enzymol.* **517,** 203-224, doi:10.1016/B978-0-12-404634-4.00010-3 (2012).

64 Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y. & Zhao, H. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth Biol* **2,** 662-669, doi:10.1021/sb400058n (2013).

65 Shao, Z. & Zhao, H. Construction and engineering of large biochemical pathways via DNA assembler. *Methods Mol. Biol.* **1073,** 85-106, doi:10.1007/978-1-62703-625-2_9 (2013).

66 Luo, Y., Huang, H., Liang, J., Wang, M., Lu, L., Shao, Z., Cobb, R. E. & Zhao, H. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat Commun* **4,** 2894, doi:10.1038/ncomms3894 (2013).

67 Shao, Z. & Zhao, H. Manipulating natural product biosynthetic pathways via DNA assembler. *Curr. Protoc. Chem. Biol.* **6,** 65-100, doi:10.1002/9780470559277.ch130191 (2014).

68 Schimming, O., Fleischhacker, F., Nollmann, F. I. & Bode, H. B. Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. *Chembiochem* **15,** 1290-1294, doi:10.1002/cbic.201402065 (2014).

69 Domrose, A., Weihmann, R., Thies, S., Jaeger, K. E., Drepper, T. & Loeschcke, A. Rapid generation of recombinant Pseudomonas putida secondary metabolite producers using yTREX. *Synth Syst Biotechnol* **2,** 310-319, doi:10.1016/j.synbio.2017.11.001 (2017).

70 Basitta, P., Westrich, L., Rosch, M., Kulik, A., Gust, B. & Apel, A. K. AGOS: A Plug-and-Play Method for the Assembly of Artificial Gene Operons into Functional Biosynthetic Gene Clusters. *ACS Synth Biol* **6,** 817-825, doi:10.1021/acssynbio.6b00319 (2017).

71 Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. *Nat. Protoc.* **3,** 371-377, doi:10.1038/nprot.2008.5 (2008).

72 Yamanaka, K., Reynolds, K. A., Kersten, R. D., Ryan, K. S., Gonzalez, D. J., Nizet, V., Dorrestein, P. C. & Moore, B. S. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 1957-1962, doi:10.1073/pnas.1319584111 (2014).

73 Li, Y., Li, Z., Yamanaka, K., Xu, Y., Zhang, W., Vlamakis, H., Kolter, R., Moore, B. S. & Qian, P. Y. Directed natural product biosynthesis gene cluster capture and expression in the model bacterium Bacillus subtilis. *Sci. Rep.* **5,** 9383, doi:10.1038/srep09383 (2015).

74 Ross, A. C., Gulland, L. E., Dorrestein, P. C. & Moore, B. S. Targeted capture and heterologous expression of the Pseudoalteromonas alterochromide gene cluster in Escherichia coli represents a promising natural product exploratory platform. *ACS Synth Biol* **4,** 414-420, doi:10.1021/sb500280q (2015).

75 Bonet, B., Teufel, R., Crusemann, M., Ziemert, N. & Moore, B. S. Direct capture and heterologous expression of Salinispora natural product genes for the biosynthesis of enterocin. *J. Nat. Prod.* **78,** 539-542, doi:10.1021/np500664q (2015).

76 Cano-Prieto, C., Garcia-Salcedo, R., Sanchez-Hidalgo, M., Brana, A. F., Fiedler, H. P., Mendez, C., Salas, J. A. & Olano, C. Genome Mining of Streptomyces sp. Tu 6176: Characterization of the Nataxazole Biosynthesis Pathway. *Chembiochem* **16,** 1461-1473, doi:10.1002/cbic.201500153 (2015).

77 Ray, L., Yamanaka, K. & Moore, B. S. A Peptidyl-Transesterifying Type I Thioesterase in Salinamide Biosynthesis. *Angew. Chem. Int. Ed. Engl.* **55,** 364-367, doi:10.1002/anie.201508576 (2016).

78 Li, Z. R., Li, J., Gu, J. P., Lai, J. Y., Duggan, B. M., Zhang, W. P., Li, Z. L., Li, Y. X., Tong, R. B., Xu, Y., Lin, D. H., Moore, B. S. & Qian, P. Y. Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nat. Chem. Biol.* **12,** 773-775, doi:10.1038/nchembio.2157 (2016).

79 Larson, C. B., Crusemann, M. & Moore, B. S. PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete. *J. Nat. Prod.* **80,** 1200-1204, doi:10.1021/acs.jnatprod.6b01121 (2017).

80 Wu, N., Huang, H., Min, T. & Hu, H. TAR cloning and integrated overexpression of 6-demethylchlortetracycline biosynthetic gene cluster in Streptomyces aureofaciens. *Acta Biochim Biophys Sin (Shanghai)* **49,** 1129-1134, doi:10.1093/abbs/gmx110 (2017).

81 Garcia-Salcedo, R., Alvarez-Alvarez, R., Olano, C., Canedo, L., Brana, A. F., Mendez, C., de la Calle, F. & Salas, J. A. Characterization of the Jomthonic Acids Biosynthesis Pathway and Isolation of Novel Analogues in Streptomyces caniferus GUA-06-05-006A. *Mar. Drugs* **16,** 259, doi:10.3390/md16080259 (2018).

82 Hu, Y., Nan, F., Maina, S. W., Guo, J., Wu, S. & Xin, Z. Clone of plipastatin biosynthetic gene cluster by transformation-associated recombination technique and high efficient expression in model organism Bacillus subtilis. *J. Biotechnol.* **288,** 1-8, doi:10.1016/j.jbiotec.2018.10.006 (2018).

83 Noskov, V. N., Kouprina, N., Leem, S. H., Ouspenski, I., Barrett, J. C. & Larionov, V. A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics* **4,** 16, doi:10.1186/1471-2164-4-16 (2003).

84 Tang, X., Li, J., Millan-Aguinaga, N., Zhang, J. J., O'Neill, E. C., Ugalde, J. A., Jensen, P. R., Mantovani, S. M. & Moore, B. S. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10,** 2841-2849, doi:10.1021/acschembio.5b00658 (2015).

85 Alberti, F., Leng, D. J., Wilkening, I., Song, L., Tosin, M. & Corre, C. Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery. *Chem. Sci.* **10,** 453-463, doi:10.1039/c8sc03814g (2019).

86 Zhang, J. J., Yamanaka, K., Tang, X. & Moore, B. S. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol.* **621,** 87-110, doi:10.1016/bs.mie.2019.02.026 (2019).

87 Kouprina, N. & Larionov, V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat Rev Genet* **7,** 805-812, doi:10.1038/nrg1943 (2006).

88 Lee, N. C., Larionov, V. & Kouprina, N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res.* **43,** e55, doi:10.1093/nar/gkv112 (2015).

89 Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P. M., Plaza, A., Xia, L., Muller, R., Stewart, A. F. & Zhang, Y. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30,** 440-446, doi:10.1038/nbt.2183 (2012).

90 Xu, X., Zhou, H., Liu, Y., Liu, X., Fu, J., Li, A., Li, Y. Z., Shen, Y., Bian, X. & Zhang, Y. Heterologous Expression Guides Identification of the Biosynthetic Gene Cluster of Chuangxinmycin, an Indole Alkaloid Antibiotic. *J. Nat. Prod.* **81,** 1060-1064, doi:10.1021/acs.jnatprod.7b00835 (2018).

91 Huang, F., Tang, J., He, L., Ding, X., Huang, S., Zhang, Y., Sun, Y. & Xia, L. Heterologous expression and antitumor activity analysis of syringolin from Pseudomonas syringae pv. syringae B728a. *Microb Cell Fact* **17,** 31, doi:10.1186/s12934-018-0859-1 (2018).

92 Thongkongkaew, T., Ding, W., Bratovanov, E., Oueis, E., Garci, A. A. M. A., Zaburannyi, N., Harmrolfs, K., Zhang, Y., Scherlach, K., Muller, R. & Hertweck, C. Two Types of Threonine-Tagged Lipopeptides Synergize in Host Colonization by Pathogenic Burkholderia Species. *ACS Chem. Biol.* **13,** 1370-1379, doi:10.1021/acschembio.8b00221 (2018).

93 Tang, Y., Frewert, S., Harmrolfs, K., Herrmann, J., Karmann, L., Kazmaier, U., Xia, L., Zhang, Y. & Muller, R. Heterologous expression of an orphan NRPS gene cluster from Paenibacillus larvae in Escherichia coli revealed production of sevadicin. *J. Biotechnol.* **194,** 112-114, doi:10.1016/j.jbiotec.2014.12.008 (2015).

94 Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Muller, R., Fu, J., Stewart, A. F. & Zhang, Y. ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* **46,** e28, doi:10.1093/nar/gkx1249 (2018).

95 Song, C., Luan, J., Cui, Q., Duan, Q., Li, Z., Gao, Y., Li, R., Li, A., Shen, Y. M., Li, Y. Z., Stewart, A. F., Zhang, Y., Fu, J. & Wang, H. Enhanced heterologous spinosad production from a 79-kb synthetic multi-operon assembly. *ACS Synth Biol* **8,** 137-147, doi:10.1021/acssynbio.8b00402 (2018).

96 Wang, R. Y., Shi, Z. Y., Chen, J. C. & Chen, G. Q. Cloning large gene clusters from E. coli using in vitro single-strand overlapping annealing. *ACS Synth Biol* **1,** 291-295, doi:10.1021/sb300025d (2012).

97 Zhou, Y., Murphy, A. C., Samborskyy, M., Prediger, P., Dias, L. C. & Leadlay, P. F. Iterative Mechanism of Macrodiolide Formation in the Anticancer Compound Conglobatin. *Chem. Biol.* **22,** 745-754, doi:10.1016/j.chembiol.2015.05.010 (2015).

98 Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y. & Zhu, T. F. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun* **6,** 8101, doi:10.1038/ncomms9101 (2015).

99 Tang, W., Guo, Z., Cao, Z., Wang, M., Li, P., Meng, X., Zhao, X., Xie, Z., Wang, W., Zhou, A., Lou, C. & Chen, Y. d-Sedoheptulose-7-phosphate is a common precursor for the heptoses of septacidin and hygromycin B. *Proc. Natl. Acad. Sci. U. S. A.* **115,** 2818-2823, doi:10.1073/pnas.1711665115 (2018).

100 Osswald, C., Zaburannyi, N., Burgard, C., Hoffmann, T., Wenzel, S. C. & Muller, R. A highly unusual polyketide synthase directs dawenol polyene biosynthesis in Stigmatella aurantiaca. *J. Biotechnol.* **191,** 54-63, doi:10.1016/j.jbiotec.2014.07.447 (2014).

101 Nah, H. J., Woo, M. W., Choi, S. S. & Kim, E. S. Precise cloning and tandem integration of large polyketide biosynthetic gene cluster using Streptomyces artificial chromosome system. *Microb Cell Fact* **14,** 140, doi:10.1186/s12934-015-0325-2 (2015).

102 Wang, J., Lu, A., Liu, J., Huang, W., Wang, J., Cai, Z. & Zhao, G. iCatch: a new strategy for capturing large DNA fragments using homing endonucleases. *Acta Biochim Biophys Sin (Shanghai)* **51,** 97-103, doi:10.1093/abbs/gmy139 (2019).

103 Dai, R. X., Zhang, B., Zhao, G. P. & Ding, X. M. Site-specific recombination for cloning of large DNA fragments in vitro. *Eng Life Sci* **15,** 655-659, doi:10.1002/elsc.201400267 (2015).

104 Yu, Y., Tang, B., Dai, R., Zhang, B., Chen, L., Yang, H., Zhao, G. & Ding, X. Identification of the streptothricin and tunicamycin biosynthetic gene clusters by genome mining in Streptomyces sp. strain fd1-xmd. *Appl. Microbiol. Biotechnol.* **102,** 2621-2633, doi:10.1007/s00253-018-8748-4 (2018).

105 Du, D., Wang, L., Tian, Y., Liu, H., Tan, H. & Niu, G. Genome engineering and direct cloning of antibiotic gene clusters via phage varphiBT1 integrase-mediated site-specific recombination in Streptomyces. *Sci. Rep.* **5,** 8740, doi:10.1038/srep08740 (2015).

106 Nguyen, H. N., Ishidoh, K. I., Kinoshita, H. & Nihira, T. Targeted cloning of a large gene cluster from Lecanicillium genome by Cre/loxP based method. *J. Microbiol. Methods* **150,** 47-54, doi:10.1016/j.mimet.2018.05.017 (2018).

107 Ziemert, N., Lechner, A., Wietz, M., Millan-Aguinaga, N., Chavarria, K. L. & Jensen, P. R. Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1130-1139, doi:10.1073/pnas.1324161111 (2014).

108 Letzel, A. C., Li, J., Amos, G. C. A., Millan-Aguinaga, N., Ginigini, J., Abdelmohsen, U. R., Gaudencio, S. P., Ziemert, N., Moore, B. S. & Jensen, P. R. Genomic insights into specialized metabolism in the marine actinomycete Salinispora. *Environ. Microbiol.* **19,** 3660-3673, doi:10.1111/1462-2920.13867 (2017).

109 Martinez-Burgo, Y., Alvarez-Alvarez, R., Perez-Redondo, R. & Liras, P. Heterologous expression of Streptomyces clavuligerus ATCC 27064 cephamycin C gene cluster. *J. Biotechnol.* **186,** 21-29, doi:10.1016/j.jbiotec.2014.06.002 (2014).

110 Craig, J. W., Chang, F. Y., Kim, J. H., Obiajulu, S. C. & Brady, S. F. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl. Environ. Microbiol.* **76,** 1633-1641, doi:10.1128/AEM.02169-09 (2010).

111 Fu, J., Wenzel, S. C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A. F., Muller, R. & Zhang, Y. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* **36,** e113, doi:10.1093/nar/gkn499 (2008).

112 Wenzel, S. C., Hoffmann, H., Zhang, J., Debussche, L., Haag-Richter, S., Kurz, M., Nardi, F., Lukat, P., Kochems, I., Tietgen, H., Schummer, D., Nicolas, J. P., Calvet, L., Czepczor, V., Vrignaud, P., Muhlenweg, A., Pelzer, S., Muller, R. & Bronstrup, M. Production of the Bengamide Class of Marine Natural Products in Myxobacteria: Biosynthesis and Structure-Activity Relationships. *Angew. Chem. Int. Ed. Engl.* **54,** 15560-15564, doi:10.1002/anie.201508277 (2015).

113 Gomez-Escribano, J. P. & Bibb, M. J. Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* **4,** 207-215, doi:10.1111/j.1751-7915.2010.00219.x (2011).

114 Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E. & Ikeda, H. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2646-2651, doi:10.1073/pnas.0914833107 (2010).

115 Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozone, I., Izumikawa, M., Hashimoto, J., Takagi, M., Omura, S., Shin-ya, K., Cane, D. E. & Ikeda, H. Engineered Streptomyces avermitilis host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth Biol* **2,** 384-396, doi:10.1021/sb3001003 (2013).

116 Peng, Q., Gao, G., Lu, J., Long, Q., Chen, X., Zhang, F., Xu, M., Liu, K., Wang, Y., Deng, Z., Li, Z. & Tao, M. Engineered Streptomyces lividans Strains for Optimal Identification and Expression of Cryptic Biosynthetic Gene Clusters. *Front. Microbiol.* **9,** 3042, doi:10.3389/fmicb.2018.03042 (2018).

117 Myronovskyi, M., Rosenkranzer, B., Nadmid, S., Pujic, P., Normand, P. & Luzhetskyy, A. Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab Eng* **49,** 316-324, doi:10.1016/j.ymben.2018.09.004 (2018).

118 Kallifidas, D., Jiang, G., Ding, Y. & Luesch, H. Rational engineering of Streptomyces albus J1074 for the overexpression of secondary metabolite gene clusters. *Microb Cell Fact* **17,** 25, doi:10.1186/s12934-018-0874-2 (2018).

119 Salem, S. M., Kancharla, P., Florova, G., Gupta, S., Lu, W. & Reynolds, K. A. Elucidation of final steps of the marineosins biosynthetic pathway through identification and

characterization of the corresponding gene cluster. *J. Am. Chem. Soc.* **136,** 4565-4574, doi:10.1021/ja411544w (2014).

120 Alexander, D. C., Rock, J., He, X., Brian, P., Miao, V. & Baltz, R. H. Development of a genetic system for combinatorial biosynthesis of lipopeptides in Streptomyces fradiae and heterologous expression of the A54145 biosynthesis gene cluster. *Appl. Environ. Microbiol.* **76,** 6877-6887, doi:10.1128/AEM.01248-10 (2010).

121 Alvarez-Alvarez, R., Martinez-Burgo, Y., Perez-Redondo, R., Brana, A. F., Martin, J. F. & Liras, P. Expression of the endogenous and heterologous clavulanic acid cluster in Streptomyces flavogriseus: why a silent cluster is sleeping. *Appl. Microbiol. Biotechnol.* **97,** 9451-9463, doi:10.1007/s00253-013-5148-7 (2013).

122 Bu, Q. T., Yu, P., Wang, J., Li, Z. Y., Chen, X. A., Mao, X. M. & Li, Y. Q. Rational construction of genome-reduced and high-efficient industrial Streptomyces chassis based on multiple comparative genomic approaches. *Microb Cell Fact* **18,** 16, doi:10.1186/s12934-019-1055-7 (2019).

123 Zhang, B., Wang, K. B., Wang, W., Bi, S. F., Mei, Y. N., Deng, X. Z., Jiao, R. H., Tan, R. X. & Ge, H. M. Discovery, Biosynthesis, and Heterologous Production of Streptoseomycin, an Anti-Microaerophilic Bacteria Macrodilactone. *Org Lett* **20,** 2967-2971, doi:10.1021/acs.orglett.8b01006 (2018).

124 Chen, W. & Qin, Z. Development of a gene cloning system in a fast-growing and moderately thermophilic Streptomyces species and heterologous expression of Streptomyces antibiotic biosynthetic gene clusters. *BMC Microbiol.* **11,** 243, doi:10.1186/1471-2180-11-243 (2011).

125 Winter, J. M., Moffitt, M. C., Zazopoulos, E., McAlpine, J. B., Dorrestein, P. C. & Moore, B. S. Molecular basis for chloronium-mediated meroterpene cyclization: cloning, sequencing, and heterologous expression of the napyradiomycin biosynthetic gene cluster. *J. Biol. Chem.* **282,** 16362-16368, doi:10.1074/jbc.M611046200 (2007).

126 Zhang, G., Zhang, H., Li, S., Xiao, J., Zhang, G., Zhu, Y., Niu, S., Ju, J. & Zhang, C. Characterization of the amicetin biosynthesis gene cluster from Streptomyces vinaceusdrappus NRRL 2363 implicates two alternative strategies for amide bond formation. *Appl. Environ. Microbiol.* **78,** 2393-2401, doi:10.1128/AEM.07185-11 (2012).

127 Kudo, F., Tsunoda, T., Takashima, M. & Eguchi, T. Five-Membered Cyclitol Phosphate Formation by a myo-Inositol Phosphate Synthase Orthologue in the Biosynthesis of the Carbocyclic Nucleoside Antibiotic Aristeromycin. *Chembiochem* **17,** 2143-2148, doi:10.1002/cbic.201600348 (2016).

128 Thorpe, H. M. & Smith, M. C. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5505-5510 (1998).

129 Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, C. J. & Chater, K. F. Lambda red-mediated genetic manipulation of antibiotic-producing Streptomyces. *Adv. Appl. Microbiol.* **54,** 107-128, doi:10.1016/S0065-2164(04)54004-2 (2004).

130 Eustaquio, A. S., Gust, B., Galm, U., Li, S. M., Chater, K. F. & Heide, L. Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters. *Appl. Environ. Microbiol.* **71,** 2452-2459, doi:10.1128/AEM.71.5.2452-2459.2005 (2005).

131 Smith, M. C., Hendrix, R. W., Dedrick, R., Mitchell, K., Ko, C. C., Russell, D., Bell, E., Gregory, M., Bibb, M. J., Pethick, F., Jacobs-Sera, D., Herron, P., Buttner, M. J. & Hatfull, G. F. Evolutionary relationships among actinophages and a putative adaptation for growth in Streptomyces spp. *J. Bacteriol.* **195,** 4924-4935, doi:10.1128/JB.00618-13 (2013).

132 Fayed, B., Younger, E., Taylor, G. & Smith, M. C. A novel Streptomyces spp. integration vector derived from the S. venezuelae phage, SV1. *BMC Biotechnol.* **14,** 51, doi:10.1186/1472-6750-14-51 (2014).

133 Fogg, P. C., Haley, J. A., Stark, W. M. & Smith, M. C. Genome Integration and Excision by a New Streptomyces Bacteriophage, varphiJoe. *Appl. Environ. Microbiol.* **83,** e02767, doi:10.1128/AEM.02767-16 (2017).

134 Fayed, B., Ashford, D. A., Hashem, A. M., Amin, M. A., El Gazayerly, O. N., Gregory, M. A. & Smith, M. C. Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in Streptomyces spp. and combinatorial biosynthesis. *Appl. Environ. Microbiol.* **81,** 8402-8413, doi:10.1128/AEM.02403-15 (2015).

135 Li, L., Wei, K., Liu, X., Wu, Y., Zheng, G., Chen, S., Jiang, W. & Lu, Y. aMSGE: advanced multiplex site-specific genome engineering with orthogonal modular recombinases in actinomycetes. *Metab Eng* **52,** 153-167, doi:10.1016/j.ymben.2018.12.001 (2019).

136 Bilyk, B., Horbal, L. & Luzhetskyy, A. Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in Streptomyces albus J1074. *Microb Cell Fact* **16,** 5, doi:10.1186/s12934-016-0619-z (2017).

137 Sherwood, E. J., Hesketh, A. R. & Bibb, M. J. Cloning and analysis of the planosporicin lantibiotic biosynthetic gene cluster of Planomonospora alba. *J. Bacteriol.* **195,** 2309-2321, doi:10.1128/JB.02291-12 (2013).

138 Schwarz, P. N., Buchmann, A., Roller, L., Kulik, A., Gross, H., Wohlleben, W. & Stegmann, E. The Immunosuppressant Brasilicardin: Determination of the Biosynthetic Gene Cluster in the Heterologous Host Amycolatopsis japonicum. *Biotechnol J* **13,** 1700527, doi:10.1002/biot.201700527 (2018).

139 Zhang, J. J., Moore, B. S. & Tang, X. Engineering Salinispora tropica for heterologous expression of natural product biosynthetic gene clusters. *Appl. Microbiol. Biotechnol.* **102,** 8437-8446, doi:10.1007/s00253-018-9283-z (2018).

140 Huang, J., Yu, Z., Li, M. H., Wang, J. D., Bai, H., Zhou, J. & Zheng, Y. G. High Level of Spinosad Production in the Heterologous Host Saccharopolyspora erythraea. *Appl. Environ. Microbiol.* **82,** 5603-5611, doi:10.1128/AEM.00618-16 (2016).

141 Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **31,** 61-108, doi:10.1039/c3np70054b (2014).

142 Herzner, A. M., Dischinger, J., Szekat, C., Josten, M., Schmitz, S., Yakeleba, A., Reinartz, R., Jansen, A., Sahl, H. G., Piel, J. & Bierbaum, G. Expression of the lantibiotic mersacidin in Bacillus amyloliquefaciens FZB42. *PLoS One* **6,** e22389, doi:10.1371/journal.pone.0022389 (2011).

143 Choi, S. K., Park, S. Y., Kim, R., Kim, S. B., Lee, C. H., Kim, J. F. & Park, S. H. Identification of a polymyxin synthetase gene cluster of Paenibacillus polymyxa and heterologous expression of the gene in Bacillus subtilis. *J. Bacteriol.* **191,** 3350-3358, doi:10.1128/JB.01728-08 (2009).

144 Borisova, S. A., Circello, B. T., Zhang, J. K., van der Donk, W. A. & Metcalf, W. W. Biosynthesis of rhizocticins, antifungal phosphonate oligopeptides produced by Bacillus subtilis ATCC6633. *Chem. Biol.* **17,** 28-37, doi:10.1016/j.chembiol.2009.11.017 (2010).

145 Zobel, S., Kumpfmuller, J., Sussmuth, R. D. & Schweder, T. Bacillus subtilis as heterologous host for the secretory production of the non-ribosomal cyclodepsipeptide enniatin. *Appl. Microbiol. Biotechnol.* **99,** 681-691, doi:10.1007/s00253-014-6199-0 (2015).

146 Kumpfmuller, J., Methling, K., Fang, L., Pfeifer, B. A., Lalk, M. & Schweder, T. Production of the polyketide 6-deoxyerythronolide B in the heterologous host Bacillus subtilis. *Appl. Microbiol. Biotechnol.* **100,** 1209-1220, doi:10.1007/s00253-015-6990-6 (2016).

147 Liu, Q., Shen, Q., Bian, X., Chen, H., Fu, J., Wang, H., Lei, P., Guo, Z., Chen, W., Li, D. & Zhang, Y. Simple and rapid direct cloning and heterologous expression of natural product biosynthetic gene cluster in Bacillus subtilis via Red/ET recombineering. *Sci. Rep.* **6,** 34623, doi:10.1038/srep34623 (2016).

148 Daba, G. M., Ishibashi, N., Zendo, T. & Sonomoto, K. Functional analysis of the biosynthetic gene cluster required for immunity and secretion of a novel Lactococcus-specific bacteriocin, lactococcin Z. *J. Appl. Microbiol.* **123,** 1124-1132, doi:10.1111/jam.13564 (2017).

149 Borrero, J., Kelly, E., O'Connor, P. M., Kelleher, P., Scully, C., Cotter, P. D., Mahony, J. & van Sinderen, D. Plantaricyclin A, a Novel Circular Bacteriocin Produced by Lactobacillus plantarum NI326: Purification, Characterization, and Heterologous Production. *Appl. Environ. Microbiol.* **84,** e01801, doi:10.1128/AEM.01801-17 (2018).

150 Masschelein, J., Jenner, M. & Challis, G. L. Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights. *Nat. Prod. Rep.* **34,** 712-783, doi:10.1039/c7np00010c (2017).

151 Bian, X., Plaza, A., Zhang, Y. & Muller, R. Luminmycins A-C, cryptic natural products from Photorhabdus luminescens identified by heterologous expression in Escherichia coli. *J. Nat. Prod.* **75,** 1652-1655, doi:10.1021/np300444e (2012).

152 Bian, X., Huang, F., Wang, H., Klefisch, T., Muller, R. & Zhang, Y. Heterologous production of glidobactins/luminmycins in Escherichia coli Nissle containing the glidobactin biosynthetic gene cluster from Burkholderia DSM7029. *Chembiochem* **15,** 2221-2224, doi:10.1002/cbic.201402199 (2014).

153 Stevens, D. C., Hari, T. P. & Boddy, C. N. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat. Prod. Rep.* **30,** 1391-1411, doi:10.1039/c3np70060g (2013).

154 Agarwal, V., El Gamal, A. A., Yamanaka, K., Poth, D., Kersten, R. D., Schorn, M., Allen, E. E. & Moore, B. S. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol.* **10,** 640-647, doi:10.1038/nchembio.1564 (2014).

155 Watanabe, K., Hotta, K., Praseuth, A. P., Koketsu, K., Migita, A., Boddy, C. N., Wang, C. C., Oguri, H. & Oikawa, H. Total biosynthesis of antitumor nonribosomal peptides in Escherichia coli. *Nat. Chem. Biol.* **2,** 423-428, doi:10.1038/nchembio803 (2006).

156 Wang, Y. & Pfeifer, B. A. 6-deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in E. coli. *Metab Eng* **10,** 33-38, doi:10.1016/j.ymben.2007.09.002 (2008).

157 Zhang, H., Fang, L., Osburne, M. S. & Pfeifer, B. A. The Continuing Development of E. coli as a Heterologous Host for Complex Natural Product Biosynthesis. *Methods Mol. Biol.* **1401,** 121-134, doi:10.1007/978-1-4939-3375-4_8 (2016).

158 Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C. & Walsh, C. T. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3,** 923-936 (1996).

159 Zhang, W., Li, Y. & Tang, Y. Engineered biosynthesis of bacterial aromatic polyketides in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 20683-20688, doi:10.1073/pnas.0809084105 (2008).

160 Zhang, H., Wang, Y., Wu, J., Skalina, K. & Pfeifer, B. A. Complete biosynthesis of erythromycin A and designed analogs using E. coli as a heterologous host. *Chem. Biol.* **17,** 1232-1240, doi:10.1016/j.chembiol.2010.09.013 (2010).

161 Stevens, D. C., Conway, K. R., Pearce, N., Villegas-Penaranda, L. R., Garza, A. G. & Boddy, C. N. Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in Escherichia coli. *PLoS One* **8,** e64858, doi:10.1371/journal.pone.0064858 (2013).

162 Bian, X., Huang, F., Stewart, F. A., Xia, L., Zhang, Y. & Muller, R. Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in E. coli through Red/ET recombineering. *Chembiochem* **13,** 1946-1952, doi:10.1002/cbic.201200310 (2012).

163 Cai, X., Nowak, S., Wesche, F., Bischoff, I., Kaiser, M., Furst, R. & Bode, H. B. Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design. *Nat. Chem.* **9,** 379-386, doi:10.1038/nchem.2671 (2017).

164 Zhang, H., Boghigian, B. A. & Pfeifer, B. A. Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli. *Biotechnol. Bioeng.* **105,** 567-573, doi:10.1002/bit.22560 (2010).

165 Fang, L., Guell, M., Church, G. M. & Pfeifer, B. A. Heterologous erythromycin production across strain and plasmid construction. *Biotechnol. Prog.* **34,** 271-276, doi:10.1002/btpr.2567 (2018).

166 Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M. & van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30,** 108-160, doi:10.1039/c2np20085f (2013).

167 Yang, X., Lennard, K. R., He, C., Walker, M. C., Ball, A. T., Doigneaux, C., Tavassoli, A. & van der Donk, W. A. A lanthipeptide library used to identify a protein-protein interaction inhibitor. *Nat. Chem. Biol.* **14,** 375-380, doi:10.1038/s41589-018-0008-5 (2018).

168 Santos, C. N., Regitsky, D. D. & Yoshikuni, Y. Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat Commun* **4,** 2503, doi:10.1038/ncomms3503 (2013).

169 Santos, C. N. & Yoshikuni, Y. Engineering complex biological systems in bacteria through recombinase-assisted genome engineering. *Nat. Protoc.* **9,** 1320-1336, doi:10.1038/nprot.2014.084 (2014).

170 Wenzel, S. C., Gross, F., Zhang, Y., Fu, J., Stewart, A. F. & Muller, R. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. *Chem. Biol.* **12,** 349-356, doi:10.1016/j.chembiol.2004.12.012 (2005).

171 Chai, Y., Shan, S., Weissman, K. J., Hu, S., Zhang, Y. & Muller, R. Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis. *Chem. Biol.* **19,** 361-371, doi:10.1016/j.chembiol.2012.01.007 (2012).

172 Loeschcke, A., Markert, A., Wilhelm, S., Wirtz, A., Rosenau, F., Jaeger, K. E. & Drepper, T. TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synth Biol* **2,** 22-33, doi:10.1021/sb3000657 (2013).

173 Yan, F., Auerbach, D., Chai, Y., Keller, L., Tu, Q., Huttel, S., Glemser, A., Grab, H. A., Bach, T., Zhang, Y. & Muller, R. Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidinecarboxylic Acid Formation. *Angew. Chem. Int. Ed. Engl.* **57,** 8754-8759, doi:10.1002/anie.201802479 (2018).

174 Ramel, C., Tobler, M., Meyer, M., Bigler, L., Ebert, M. O., Schellenberg, B. & Dudler, R. Biosynthesis of the proteasome inhibitor syringolin A: the ureido group joining two amino

acids originates from bicarbonate. *BMC Biochem.* **10,** 26, doi:10.1186/1471-2091-10-26 (2009).

175 Dudnik, A., Bigler, L. & Dudler, R. Heterologous expression of a Photorhabdus luminescens syrbactin-like gene cluster results in production of the potent proteasome inhibitor glidobactin A. *Microbiol. Res.* **168,** 73-76, doi:10.1016/j.micres.2012.09.006 (2013).

176 Kretsch, A. M., Morgan, G. L., Tyrrell, J., Mevers, E., Vallet-Gely, I. & Li, B. Discovery of (Dihydro)pyrazine N-Oxides via Genome Mining in Pseudomonas. *Org Lett* **20,** 4791-4795, doi:10.1021/acs.orglett.8b01944 (2018).

177 Gross, F., Gottschalk, D. & Muller, R. Posttranslational modification of myxobacterial carrier protein domains in Pseudomonas sp. by an intrinsic phosphopantetheinyl transferase. *Appl. Microbiol. Biotechnol.* **68,** 66-74, doi:10.1007/s00253-004-1836-7 (2005).

178 Gross, F., Ring, M. W., Perlova, O., Fu, J., Schneider, S., Gerth, K., Kuhlmann, S., Stewart, A. F., Zhang, Y. & Muller, R. Metabolic engineering of Pseudomonas putida for methylmalonyl-CoA biosynthesis to enable complex heterologous secondary metabolite formation. *Chem. Biol.* **13,** 1253-1264, doi:10.1016/j.chembiol.2006.09.014 (2006).

179 Martinez-Garcia, E., Nikel, P. I., Aparicio, T. & de Lorenzo, V. Pseudomonas 2.0: genetic upgrading of P. putida KT2440 as an enhanced host for heterologous gene expression. *Microb Cell Fact* **13,** 159, doi:10.1186/s12934-014-0159-3 (2014).

180 Munoz-Dorado, J., Marcos-Torres, F. J., Garcia-Bravo, E., Moraleda-Munoz, A. & Perez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **7,** 781, doi:10.3389/fmicb.2016.00781 (2016).

181 Stevens, D. C., Henry, M. R., Murphy, K. A. & Boddy, C. N. Heterologous expression of the oxytetracycline biosynthetic pathway in Myxococcus xanthus. *Appl. Environ. Microbiol.* **76,** 2681-2683, doi:10.1128/AEM.02841-09 (2010).

182 Sucipto, H., Pogorevc, D., Luxenburger, E., Wenzel, S. C. & Muller, R. Heterologous production of myxobacterial alpha-pyrone antibiotics in Myxococcus xanthus. *Metab Eng* **44,** 160-170, doi:10.1016/j.ymben.2017.10.004 (2017).

183 Perlova, O., Fu, J., Kuhlmann, S., Krug, D., Stewart, A. F., Zhang, Y. & Muller, R. Reconstitution of the myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in Myxococcus xanthus. *Appl. Environ. Microbiol.* **72,** 7485-7494, doi:10.1128/AEM.01503-06 (2006).

184 Perlova, O., Gerth, K., Kuhlmann, S., Zhang, Y. & Muller, R. Novel expression hosts for complex secondary metabolite megasynthetases: Production of myxochromide in the thermopilic isolate Corallococcus macrosporus GT-2. *Microb Cell Fact* **8,** 1, doi:10.1186/1475-2859-8-1 (2009).

185 Zhu, L. P., Yue, X. J., Han, K., Li, Z. F., Zheng, L. S., Yi, X. N., Wang, H. L., Zhang, Y. M. & Li, Y. Z. Allopatric integrations selectively change host transcriptomes, leading to varied

expression efficiencies of exotic genes in Myxococcus xanthus. *Microb Cell Fact* **14,** 105, doi:10.1186/s12934-015-0294-5 (2015).

186 Panter, F., Krug, D., Baumann, S. & Muller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9,** 4898-4908, doi:10.1039/c8sc01325j (2018).

187 Craig, J. W., Chang, F. Y. & Brady, S. F. Natural products from environmental DNA hosted in Ralstonia metallidurans. *ACS Chem. Biol.* **4,** 23-28, doi:10.1021/cb8002754 (2009).

188 Bian, X., Tang, B., Yu, Y., Tu, Q., Gross, F., Wang, H., Li, A., Fu, J., Shen, Y., Li, Y. Z., Stewart, A. F., Zhao, G., Ding, X., Muller, R. & Zhang, Y. Heterologous Production and Yield Improvement of Epothilones in Burkholderiales Strain DSM 7029. *ACS Chem. Biol.* **12,** 1805-1812, doi:10.1021/acschembio.7b00097 (2017).

189 Nunnery, J. K., Mevers, E. & Gerwick, W. H. Biologically active secondary metabolites from marine cyanobacteria. *Curr. Opin. Biotechnol.* **21,** 787-793, doi:10.1016/j.copbio.2010.09.019 (2010).

190 Taton, A., Unglaub, F., Wright, N. E., Zeng, W. Y., Paz-Yepes, J., Brahamsha, B., Palenik, B., Peterson, T. C., Haerizadeh, F., Golden, S. S. & Golden, J. W. Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Res.* **42,** e136, doi:10.1093/nar/gku673 (2014).

191 Agarwal, V., Blanton, J. M., Podell, S., Taton, A., Schorn, M. A., Busch, J., Lin, Z., Schmidt, E. W., Jensen, P. R., Paul, V. J., Biggs, J. S., Golden, J. W., Allen, E. E. & Moore, B. S. Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol.* **13,** 537-543, doi:10.1038/nchembio.2330 (2017).

192 Roulet, J., Taton, A., Golden, J. W., Arabolaza, A., Burkart, M. D. & Gramajo, H. Development of a cyanobacterial heterologous polyketide production platform. *Metab Eng* **49,** 94-104, doi:10.1016/j.ymben.2018.07.013 (2018).

193 Yang, G., Cozad, M. A., Holland, D. A., Zhang, Y., Luesch, H. & Ding, Y. Photosynthetic Production of Sunscreen Shinorine Using an Engineered Cyanobacterium. *ACS Synth Biol* **7,** 664-671, doi:10.1021/acssynbio.7b00397 (2018).

194 Yang, G., Zhang, Y., Lee, N. K., Cozad, M. A., Kearney, S. E., Luesch, H. & Ding, Y. Cyanobacterial Sfp-type phosphopantetheinyl transferases functionalize carrier proteins of diverse biosynthetic pathways. *Sci. Rep.* **7,** 11888, doi:10.1038/s41598-017-12244-3 (2017).

195 Videau, P., Wells, K. N., Singh, A. J., Gerwick, W. H. & Philmus, B. Assessment of Anabaena sp. Strain PCC 7120 as a Heterologous Expression Host for Cyanobacterial Natural Products: Production of Lyngbyatoxin A. *ACS Synth Biol* **5,** 978-988, doi:10.1021/acssynbio.6b00038 (2016).

196 Shang, J. L., Zhang, Z. C., Yin, X. Y., Chen, M., Hao, F. H., Wang, K., Feng, J. L., Xu, H. F., Yin, Y. C., Tang, H. R. & Qiu, B. S. UV-B induced biosynthesis of a novel sunscreen compound in solar radiation and desiccation tolerant cyanobacteria. *Environ. Microbiol.* **20,** 200-213, doi:10.1111/1462-2920.13972 (2018).

197 Cardenas, J. & Da Silva, N. A. Metabolic engineering of Saccharomyces cerevisiae for the production of triacetic acid lactone. *Metab Eng* **25,** 194-203, doi:10.1016/j.ymben.2014.07.008 (2014).

198 Cardenas, J. & Da Silva, N. A. Engineering cofactor and transport mechanisms in Saccharomyces cerevisiae for enhanced acetyl-CoA and polyketide biosynthesis. *Metab Eng* **36,** 80-89, doi:10.1016/j.ymben.2016.02.009 (2016).

199 Zhou, H., Qiao, K., Gao, Z., Vederas, J. C. & Tang, Y. Insights into radicicol biosynthesis via heterologous synthesis of intermediates and analogs. *J. Biol. Chem.* **285,** 41412-41421, doi:10.1074/jbc.M110.183574 (2010).

200 Xu, Y., Zhou, T., Zhang, S., Espinosa-Artiles, P., Wang, L., Zhang, W., Lin, M., Gunatilaka, A. A., Zhan, J. & Molnar, I. Diversity-oriented combinatorial biosynthesis of benzenediol lactone scaffolds by subunit shuffling of fungal polyketide synthases. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 12354-12359, doi:10.1073/pnas.1406999111 (2014).

201 Zabala, A. O., Chooi, Y. H., Choi, M. S., Lin, H. C. & Tang, Y. Fungal polyketide synthase product chain-length control by partnering thiohydrolase. *ACS Chem. Biol.* **9,** 1576-1586, doi:10.1021/cb500284t (2014).

202 Wang, M., Beissner, M. & Zhao, H. Aryl-aldehyde formation in fungal polyketides: discovery and characterization of a distinct biosynthetic mechanism. *Chem. Biol.* **21,** 257-263, doi:10.1016/j.chembiol.2013.12.005 (2014).

203 Mao, X. M., Zhan, Z. J., Grayson, M. N., Tang, M. C., Xu, W., Li, Y. Q., Yin, W. B., Lin, H. C., Chooi, Y. H., Houk, K. N. & Tang, Y. Efficient Biosynthesis of Fungal Polyketides Containing the Dioxabicyclo-octane Ring System. *J. Am. Chem. Soc.* **137,** 11904-11907, doi:10.1021/jacs.5b07816 (2015).

204 Chooi, Y. H., Krill, C., Barrow, R. A., Chen, S., Trengove, R., Oliver, R. P. & Solomon, P. S. An in planta-expressed polyketide synthase produces (R)-mellein in the wheat pathogen Parastagonospora nodorum. *Appl. Environ. Microbiol.* **81,** 177-186, doi:10.1128/AEM.02745-14 (2015).

205 Li, Y., Chooi, Y. H., Sheng, Y., Valentine, J. S. & Tang, Y. Comparative characterization of fungal anthracenone and naphthacenedione biosynthetic pathways reveals an alpha-hydroxylation-dependent Claisen-like cyclization catalyzed by a dimanganese thioesterase. *J. Am. Chem. Soc.* **133,** 15773-15785, doi:10.1021/ja206906d (2011).

206 Xu, W., Chooi, Y. H., Choi, J. W., Li, S., Vederas, J. C., Da Silva, N. A. & Tang, Y. LovG: the thioesterase required for dihydromonacolin L release and lovastatin nonaketide synthase turnover in lovastatin biosynthesis. *Angew. Chem. Int. Ed. Engl.* **52,** 6472-6475, doi:10.1002/anie.201302406 (2013).

207 Bond, C. M. & Tang, Y. Engineering Saccharomyces cerevisiae for production of simvastatin. *Metab Eng* **51,** 1-8, doi:10.1016/j.ymben.2018.09.005 (2019).

208 Lee, K. K., Da Silva, N. A. & Kealey, J. T. Determination of the extent of phosphopantetheinylation of polyketide synthases expressed in Escherichia coli and

Saccharomyces cerevisiae. *Anal. Biochem.* **394,** 75-80, doi:10.1016/j.ab.2009.07.010 (2009).

209 Wehrs, M., Prahl, J. P., Moon, J., Li, Y., Tanjore, D., Keasling, J. D., Pray, T. & Mukhopadhyay, A. Production efficiency of the bacterial non-ribosomal peptide indigoidine relies on the respiratory metabolic state in S. cerevisiae. *Microb Cell Fact* **17,** 193, doi:10.1186/s12934-018-1045-1 (2018).

210 Harvey, C. J. B., Tang, M., Schlecht, U., Horecka, J., Fischer, C. R., Lin, H. C., Li, J., Naughton, B., Cherry, J., Miranda, M., Li, Y. F., Chu, A. M., Hennessy, J. R., Vandova, G. A., Inglis, D., Aiyar, R. S., Steinmetz, L. M., Davis, R. W., Medema, M. H., Sattely, E., Khosla, C., St Onge, R. P., Tang, Y. & Hillenmeyer, M. E. HEx: A heterologous expression platform for the discovery of fungal natural products. *Sci Adv* **4,** eaar5459, doi:10.1126/sciadv.aar5459 (2018).

211 Shi, S., Liang, Y., Ang, E. L. & Zhao, H. Delta Integration CRISPR-Cas (Di-CRISPR) in Saccharomyces cerevisiae. *Methods Mol. Biol.* **1927,** 73-91, doi:10.1007/978-1-4939-9142-6_6 (2019).

212 Anyaogu, D. C. & Mortensen, U. H. Heterologous production of fungal secondary metabolites in Aspergilli. *Front. Microbiol.* **6,** 77, doi:10.3389/fmicb.2015.00077 (2015).

213 van Dijk, J. W., Guo, C. J. & Wang, C. C. Engineering Fungal Nonribosomal Peptide Synthetase-like Enzymes by Heterologous Expression and Domain Swapping. *Org Lett* **18,** 6236-6239, doi:10.1021/acs.orglett.6b02821 (2016).

214 Lin, T. S., Chiang, Y. M. & Wang, C. C. Biosynthetic Pathway of the Reduced Polyketide Product Citreoviridin in Aspergillus terreus var. aureus Revealed by Heterologous Expression in Aspergillus nidulans. *Org Lett* **18,** 1366-1369, doi:10.1021/acs.orglett.6b00299 (2016).

215 Yin, W. B., Chooi, Y. H., Smith, A. R., Cacho, R. A., Hu, Y., White, T. C. & Tang, Y. Discovery of cryptic polyketide metabolites from dermatophytes using heterologous expression in Aspergillus nidulans. *ACS Synth Biol* **2,** 629-634, doi:10.1021/sb400048b (2013).

216 Liu, N., Hung, Y. S., Gao, S. S., Hang, L., Zou, Y., Chooi, Y. H. & Tang, Y. Identification and Heterologous Production of a Benzoyl-Primed Tricarboxylic Acid Polyketide Intermediate from the Zaragozic Acid A Biosynthetic Pathway. *Org Lett* **19,** 3560-3563, doi:10.1021/acs.orglett.7b01534 (2017).

217 Chiang, Y. M., Oakley, C. E., Ahuja, M., Entwistle, R., Schultz, A., Chang, S. L., Sung, C. T., Wang, C. C. & Oakley, B. R. An efficient system for heterologous expression of secondary metabolite genes in Aspergillus nidulans. *J. Am. Chem. Soc.* **135,** 7720-7731, doi:10.1021/ja401945a (2013).

218 Alberti, F., Khairudin, K., Venegas, E. R., Davies, J. A., Hayes, P. M., Willis, C. L., Bailey, A. M. & Foster, G. D. Heterologous expression reveals the biosynthesis of the antibiotic pleuromutilin and generates bioactive semi-synthetic derivatives. *Nat Commun* **8,** 1831, doi:10.1038/s41467-017-01659-1 (2017).

219 Sims, J. W., Fillmore, J. P., Warner, D. D. & Schmidt, E. W. Equisetin biosynthesis in Fusarium heterosporum. *Chem. Commun. (Camb.)* **2,** 186-188, doi:10.1039/b413523g (2005).

220 Kakule, T. B., Jadulco, R. C., Koch, M., Janso, J. E., Barrows, L. R. & Schmidt, E. W. Native promoter strategy for high-yielding synthesis and engineering of fungal secondary metabolites. *ACS Synth Biol* **4,** 625-633, doi:10.1021/sb500296p (2015).

221 Jadulco, R. C., Koch, M., Kakule, T. B., Schmidt, E. W., Orendt, A., He, H., Janso, J. E., Carter, G. T., Larson, E. C., Pond, C., Matainaho, T. K. & Barrows, L. R. Isolation of pyrrolocins A-C: cis- and trans-decalin tetramic acid antibiotics from an endophytic fungal-derived pathway. *J. Nat. Prod.* **77,** 2537-2544, doi:10.1021/np500617u (2014).

222 Kakule, T. B., Zhang, S., Zhan, J. & Schmidt, E. W. Biosynthesis of the tetramic acids Sch210971 and Sch210972. *Org Lett* **17,** 2295-2297, doi:10.1021/acs.orglett.5b00715 (2015).

223 Kakule, T. B., Lin, Z. & Schmidt, E. W. Combinatorialization of fungal polyketide synthase-peptide synthetase hybrid proteins. *J. Am. Chem. Soc.* **136,** 17882-17890, doi:10.1021/ja511087p (2014).

224 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U S A* **97,** 6640-6645, doi:10.1073/pnas.120163297 (2000).

225 Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1541-1546, doi:10.1073/pnas.0337542100 (2003).

226 Montiel, D., Kang, H. S., Chang, F. Y., Charlop-Powers, Z. & Brady, S. F. Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 8953-8958, doi:10.1073/pnas.1507606112 (2015).

227 Eyles, T. H., Vior, N. M. & Truman, A. W. Rapid and Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related Metabolites. *ACS Synth Biol* **7,** 1211-1218, doi:10.1021/acssynbio.8b00038 (2018).

228 Tang, X., Eitel, K., Kaysser, L., Kulik, A., Grond, S. & Gust, B. A two-step sulfation in antibiotic biosynthesis requires a type III polyketide synthase. *Nat. Chem. Biol.* **9,** 610-615, doi:10.1038/nchembio.1310 (2013).

229 Tang, X., Li, J. & Moore, B. S. Minimization of the Thiolactomycin Biosynthetic Pathway Reveals that the Cytochrome P450 Enzyme TlmF Is Required for Five-Membered Thiolactone Ring Formation. *Chembiochem* **18,** 1072-1076, doi:10.1002/cbic.201700090 (2017).

230 Liu, Y., Tao, W., Wen, S., Li, Z., Yang, A., Deng, Z. & Sun, Y. In Vitro CRISPR/Cas9 System for Efficient Targeted DNA Editing. *MBio* **6,** e01714-01715, doi:10.1128/mBio.01714-15 (2015).

231 Zhang, J. J., Tang, X., Huan, T., Ross, A. C. & Moore, B. S. Avant-garde assembly-line biosynthesis expands diversity of cyclic lipodepsipeptide products. *bioRxiv***,** 560987, doi:10.1101/560987 (2019).

232 Myronovskyi, M. & Luzhetskyy, A. Native and engineered promoters in natural product discovery. *Nat. Prod. Rep.* **33,** 1006-1019, doi:10.1039/c6np00002a (2016).

233 Baral, B., Akhgari, A. & Metsa-Ketela, M. Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synth Syst Biotechnol* **3,** 163-178, doi:10.1016/j.synbio.2018.09.001 (2018).

234 Ren, H., Biswas, S., Ho, S., van der Donk, W. A. & Zhao, H. Rapid Discovery of Glycocins through Pathway Refactoring in Escherichia coli. *ACS Chem. Biol.* **13,** 2966-2972, doi:10.1021/acschembio.8b00599 (2018).

235 Yan, F., Burgard, C., Popoff, A., Zaburannyi, N., Zipf, G., Maier, J., Bernauer, H. S., Wenzel, S. C. & Muller, R. Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production. *Chem. Sci.* **9,** 7510-7519, doi:10.1039/c8sc02046a (2018).

236 Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13,** 509-523, doi:10.1038/nrmicro3496 (2015).

237 Guo, C. J., Chang, F. Y., Wyche, T. P., Backus, K. M., Acker, T. M., Funabashi, M., Taketani, M., Donia, M. S., Nayfach, S., Pollard, K. S., Craik, C. S., Cravatt, B. F., Clardy, J., Voigt, C. A. & Fischbach, M. A. Discovery of Reactive Microbiota-Derived Metabolites that Inhibit Host Proteases. *Cell* **168,** 517-526, doi:10.1016/j.cell.2016.12.021 (2017).

238 Gemperlein, K., Hoffmann, M., Huo, L., Pilak, P., Petzke, L., Muller, R. & Wenzel, S. C. Synthetic biology approaches to establish a heterologous production system for coronatines. *Metab Eng* **44,** 213-222, doi:10.1016/j.ymben.2017.09.009 (2017).

239 Coates, R. C., Bowen, B. P., Oberortner, E., Thomashow, L., Hadjithomas, M., Zhao, Z., Ke, J., Silva, L., Louie, K., Wang, G., Robinson, D., Tarver, A., Hamilton, M., Lubbe, A., Feltcher, M., Dangl, J. L., Pati, A., Weller, D., Northen, T. R., Cheng, J. F., Mouncey, N. J., Deutsch, S. & Yoshikuni, Y. An integrated workflow for phenazine-modifying enzyme characterization. *J. Ind. Microbiol. Biotechnol.* **45,** 567-577, doi:10.1007/s10295-018-2025-5 (2018).

240 Wright, G. Perspective: Synthetic biology revives antibiotics. *Nature* **509,** S13, doi:10.1038/509S13a (2014).

241 Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods* **11,** 499-507, doi:10.1038/nmeth.2918 (2014).

242 Lok, C. Mining the microbial dark matter. *Nature* **522,** 270-273, doi:10.1038/522270a (2015).

243 Segall-Shapiro, T. H., Sontag, E. D. & Voigt, C. A. Engineered promoters enable constant gene expression at any copy number in bacteria. *Nat. Biotechnol.* **36,** 352-358, doi:10.1038/nbt.4111 (2018).

244 Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., Woodruff, L. B. A., Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., Gordon, D. B., Densmore, D. & Voigt, C. A. Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.* **32,** 1241-1249, doi:10.1038/nbt.3063 (2014).

245 Birmingham, W. R., Starbird, C. A., Panosian, T. D., Nannemann, D. P., Iverson, T. M. & Bachmann, B. O. Bioretrosynthetic construction of a didanosine biosynthetic pathway. *Nat. Chem. Biol.* **10,** 392-399, doi:10.1038/nchembio.1494 (2014).

246 Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349,** 1095-1100, doi:10.1126/science.aac9373 (2015).

247 Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., Zhang, Y., Grzybowski, A. T., Harth, S., Lin, W., Lee, H., Yu, C., Shin, J., Deng, K., Benites, V. T., Wang, G., Baidoo, E. E. K., Chen, Y., Dev, I., Petzold, C. J. & Keasling, J. D. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567,** 123-126, doi:10.1038/s41586-019-0978-9 (2019).

248 Osswald, C., Zipf, G., Schmidt, G., Maier, J., Bernauer, H. S., Muller, R. & Wenzel, S. C. Modular construction of a functional artificial epothilone polyketide pathway. *ACS Synth Biol* **3,** 759-772, doi:10.1021/sb300080t (2014).

249 Winn, M., Francis, D. & Micklefield, J. De novo Biosynthesis of "Non-Natural" Thaxtomin Phytotoxins. *Angew. Chem. Int. Ed. Engl.* **57,** 6830-6833, doi:10.1002/anie.201801525 (2018).

250 Tan, G. Y., Deng, K., Liu, X., Tao, H., Chang, Y., Chen, J., Chen, K., Sheng, Z., Deng, Z. & Liu, T. Heterologous Biosynthesis of Spinosad: An Omics-Guided Large Polyketide Synthase Gene Cluster Reconstitution in Streptomyces. *ACS Synth Biol* **6,** 995-1005, doi:10.1021/acssynbio.6b00330 (2017).

251 Charlesworth, J. C. & Burns, B. P. Untapped Resources: Biotechnological Potential of Peptides and Secondary Metabolites in Archaea. *Archaea* **2015,** 282035, doi:10.1155/2015/282035 (2015).

252 Doron, L., Segal, N. & Shapira, M. Transgene Expression in Microalgae-From Tools to Applications. *Front Plant Sci* **7,** 505, doi:10.3389/fpls.2016.00505 (2016).

253 Reed, J., Stephenson, M. J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., Goss, R. J. M., Goossens, A., O'Connell, M. A. & Osbourn, A. A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab Eng* **42,** 185-193, doi:10.1016/j.ymben.2017.06.012 (2017).

254 Wang, B., Guo, F., Dong, S. H. & Zhao, H. Activation of silent biosynthetic gene clusters using transcription factor decoys. *Nat. Chem. Biol.* **15,** 111-114, doi:10.1038/s41589-018-0187-0 (2019).

255 Zhang, M. M., Wong, F. T., Wang, Y., Luo, S., Lim, Y. H., Heng, E., Yeo, W. L., Cobb, R. E., Enghiad, B., Ang, E. L. & Zhao, H. CRISPR-Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. *Nat. Chem. Biol.* **13,** 607-609, doi:10.1038/nchembio.2341 (2017).

256 Seyedsayamdost, M. R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 7266-7271, doi:10.1073/pnas.1400019111 (2014).

257 Xu, F., Wu, Y., Zhang, C., Davis, K. M., Moon, K., Bushin, L. B. & Seyedsayamdost, M. R. A genetics-free method for high-throughput discovery of cryptic microbial metabolites. *Nat. Chem. Biol.* **15,** 161-168, doi:10.1038/s41589-018-0193-2 (2019).

258 Cheng, Q., Xiang, L., Izumikawa, M., Meluzzi, D. & Moore, B. S. Enzymatic total synthesis of enterocin polyketides. *Nat. Chem. Biol.* **3,** 557-558, doi:10.1038/nchembio.2007.22 (2007).

259 Kalaitzis, J. A., Cheng, Q., Thomas, P. M., Kelleher, N. L. & Moore, B. S. In vitro biosynthesis of unnatural enterocin and wailupemycin polyketides. *J. Nat. Prod.* **72,** 469-472, doi:10.1021/np800598t (2009).

260 Agarwal, V. & Moore, B. S. Enzymatic synthesis of polybrominated dioxins from the marine environment. *ACS Chem. Biol.* **9,** 1980-1984, doi:10.1021/cb5004338 (2014).

261 Teufel, R., Kaysser, L., Villaume, M. T., Diethelm, S., Carbullido, M. K., Baran, P. S. & Moore, B. S. One-pot enzymatic synthesis of merochlorin A and B. *Angew. Chem. Int. Ed. Engl.* **53,** 11019-11022, doi:10.1002/anie.201405694 (2014).

262 Goering, A. W., Li, J., McClure, R. A., Thomson, R. J., Jewett, M. C. & Kelleher, N. L. In Vitro Reconstruction of Nonribosomal Peptide Biosynthesis Directly from DNA Using Cell-Free Protein Synthesis. *Acs Synthetic Biology* **6,** 39-44, doi:10.1021/acssynbio.6b00160 (2017).

263 McKinnie, S. M. K., Miles, Z. D., Jordan, P. A., Awakawa, T., Pepper, H. P., Murray, L. A. M., George, J. H. & Moore, B. S. Total Enzyme Syntheses of Napyradiomycins A1 and B1. *J. Am. Chem. Soc.* **140,** 17840-17845, doi:10.1021/jacs.8b10134 (2018).

264 Luhavaya, H., Sigrist, R., Chekan, J. R., McKinnie, S. M. K. & Moore, B. S. Biosynthesis of l-4-Chlorokynurenine, an Antidepressant Prodrug and a Non-Proteinogenic Amino Acid Found in Lipopeptide Antibiotics. *Angew. Chem. Int. Ed. Engl.* **58,** 8394-8399, doi:10.1002/anie.201901571 (2019).

265 Chekan, J. R., McKinnie, S. M. K., Moore, M. L., Poplawski, S. G., Michael, T. P. & Moore, B. S. Scalable Biosynthesis of the Seaweed Neurochemical, Kainic Acid. *Angew. Chem. Int. Ed. Engl.* **58,** 8454-8457, doi:10.1002/anie.201902910 (2019).

CHAPTER 2. Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters.

2.1 Abstract

Chapter 2 describes the development of a *Salinispora tropica* heterologous host, *S. tropica* CNB-4401. This represents the first marine actinobacterial host for heterologous expression of large natural product BGCs. Using this host, we express the thiolactomycin BGC from *Salinispora pacifica*, resulting in the detection of four thiolactomycin analogs of varying abundance. As *S. tropica* CNB-4401 is retrofitted with a ΦC31 genomic attachment site, it is compatible with established vectors commonly used for *Streptomyces* heterologous expression and represents a valuable addition to the arsenal of actinomycete hosts.

2.2 Reprint of “Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters”

Applied Microbiology and Biotechnology (2018) 102:8437–8446
https://doi.org/10.1007/s00253-018-9283-z

APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY

Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters

Jia Jia Zhang[1] & Bradley S. Moore [1,2] & Xiaoyu Tang[1,3]

Received: 1 May 2018 / Revised: 25 July 2018 / Accepted: 29 July 2018 / Published online: 13 August 2018

Abstract

The marine actinomycete genus *Salinispora* is a remarkably prolific source of structurally diverse and biologically active secondary metabolites. Herein, we select the model organism *Salinispora tropica* CNB-440 for development as a heterologous host for the expression of biosynthetic gene clusters (BGCs) to complement well-established *Streptomyces* host strains. In order to create an integratable host with a clean background of secondary metabolism, we replaced three genes (*salA–C*) essential for salinosporamide biosynthesis with a cassette containing the *Streptomyces coelicolor* ΦC31 phage attachment site *attB* to generate the mutant *S. tropica* CNB-4401 via double-crossover recombination. This mutagenesis not only knocks-in the attachment site *attB* in the genome of *S. tropica* CNB-440 but also abolishes production of the salinosporamides, thereby simplifying the strain's chemical background. We validated this new heterologous host with the successful integration and expression of the thiolactomycin BGC that we recently identified in several *S. pacifica* strains. When compared to the extensively engineered superhost *S. coelicolor* M1152, the production of thiolactomycins from *S. tropica* CNB-4401 was approximately 3-fold higher. To the best of our knowledge, this is the first example of using a marine actinomycete as a heterologous host for natural product BGC expression. The established heterologous host may provide a useful platform to accelerate the discovery of novel natural products and engineer biosynthetic pathways.

Keywords Heterologous expression · *Salinispora* · Natural products · Genetic engineering

Introduction

Microbially produced natural products are of paramount importance in human medicine. Not only are the majority of antimicrobial and anticancer drugs derived directly or inspired by natural products, many other branches of medicine, such as immunology, neurology, and cardiology, have similarly benefited from them (Newman and Cragg 2016). Bacteria in the order *Actinomycetales*, commonly called actinomycetes, are the undisputed leading source for natural product discovery, accounting for approximately 75% of the microbial natural products used in human therapy (Berdy 2005). Among the actinomycetes, the genus *Streptomyces* has been the major source of bioactive compounds discovered to date and has thus become the model organism for the study of natural product biosynthesis, pathway engineering, and genome mining (Baltz 2016; Chater 2016; Horinouchi 2007). Recently, however, common soil-derived *Streptomyces* bacteria have fallen out of favor as a resource for drug discovery due in part to the common re-discovery of well-known compounds. This has led to a search for other Actinobacteria and the exploration of actinomycetes from poorly exploited environments such as the ocean (Jensen et al. 2005; Schorn et al. 2016).

The marine actinomycete genus *Salinispora* was initially isolated from marine sediments and subsequently described as the first obligate marine actinomycete genus (Jensen et al. 2015). Since then, *Salinispora* species have displayed the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00253-018-9283-z) contains supplementary material, which is available to authorized users.

✉ Bradley S. Moore
bsmoore@ucsd.edu

✉ Xiaoyu Tang
xtang@microbechembio.org

[1] Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA

[2] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA, USA

[3] Genomic Medicine, J. Craig Venter Institute, La Jolla, CA, USA

ability to produce a diverse range of secondary metabolites, including salinosporamide A, which is currently in clinical trials as an anticancer agent (Gulder and Moore 2010). To date, 25 classes of compounds have been discovered from this genus, with 16 representing new carbon skeletons, correlating to virtually all known biosynthetic classes (Jensen et al. 2015). Recent genome sequencing efforts have revealed that only a small fraction of the natural product biosynthetic potential of *Salinispora* species has been identified (Amos et al. 2017; Letzel et al. 2017; Ziemert et al. 2014). In a recent study, the genomes of 75 *Salinispora* strains were queried for nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) biosynthetic gene clusters (BGCs), revealing 124 distinct biosynthetic pathways predicted to encode 124 distinct natural products (Ziemert et al. 2014). This provides an unprecedented opportunity to explore the biosynthetic potential of *Salinispora* to discover new compounds.

Heterologous expression of whole BGCs has become a routine approach not only for the discovery of metabolites from ʙorphan^ or ʙcryptic^ gene clusters (Ren et al. 2017; Rutledge and Challis 2015), but also to probe biosynthetic logic involved in natural product biosynthesis (Gomez-Escribano and Bibb 2014; Ongley et al. 2013). This approach relies on two factors: first, the ability to clone BGCs into a suitable expression vector, and second, the availability of a compatible heterologous host. Recently, several whole-pathway direct cloning approaches, such as full-length RecE-mediated recombination in *E. coli* (Fu et al. 2012), transformation-associated recombination (TAR) in *Saccharomyces cerevisiae* (Kim et al. 2010; Yamanaka et al. 2014), and Cas9-assisted in vitro assembly method (Jiang et al. 2015), have been developed, and these have greatly accelerated the BGC isolation procedure. On the other hand, several *Streptomyces* strains have been established for heterologous expression of natural product biosynthesis genes (Gomez-Escribano and Bibb 2011; Komatsu et al. 2013, 2010; Zhang et al. 2017). However, empirically, we have found that some *Salinispora* gene clusters, such as those encoding the salinosporamide (Eustaquio et al. 2009), lymphostin (Miyanaga et al. 2011), and lomaiviticin pathways (Kersten et al. 2013), are active in *S. tropica* strains but inactive when introduced to a *Streptomyces* heterologous host such as *S. coelicolor* or *S. lividans*. The expression of a BGC is a complicated cellular process that is highly dependent on the genetics and biochemistry of the producing organism (Baltz 2010). We believe that the ability to easily test the expression of BGCs across various host organisms is a valuable asset to maximize the likelihood of achieving successful heterologous production of natural product small molecules (Zhang et al. 2017). Therefore, establishing a biosynthetically versatile organism such as *Salinispora* as a heterologous host may facilitate the expression of additional BGCs for future discovery and engineering efforts.

S. tropica CNB-440 was the first whole genome-sequenced marine actinomycete and has been used as a model organism to investigate secondary metabolites (Udwary et al. 2007). So far, eight classes of compounds have been discovered from this strain alone, including the beta-lactone proteasome inhibitor salinosporamide A (Eustaquio et al. 2009) and the polyketide cytotoxin lomaiviticin A (Kersten et al. 2013). Here, we report the engineering of *S. tropica* CNB-440 as a heterologous host by introducing the phage ΦC31 attachment site *attB* into the genome for integratable expression and simultaneously abolishing background salinosporamide production. The new *Salinispora* host is directly compatible with all previously established genetic toolkits leveraging the phage ΦC31 integrase for *Streptomyces* expression (Tang et al. 2015; Yamanaka et al. 2014). We validated this new heterologous platform with the successful expression of the thiolactomycin BGC recently identified in several *S. pacifica* strains but absent in *S. tropica* (Tang et al. 2015). To the best of our knowledge, this is the first showcase of the engineering and use of a marine actinomycete as a heterologous host for natural product pathway expression.

Materials and methods

Bacterial strains, plasmids, and culture conditions

Chemicals and microbiological and molecular biological agents were purchased from standard commercial sources. Strains and plasmids used in this study are summarized in Table 1. *Salinispora tropica* CNB-440 (ATCC BAA-916 and DSM 44818) and its respective derivatives were maintained and grown on A1 agar (10 g soluble starch, 4 g yeast extract, 2 g peptone, 16 g agar, 1 l natural or artificial seawater) and/or A1 liquid medium (10 g soluble starch, 4 g yeast extract, 2 g peptone, 1 l natural or artificial seawater). *Escherichia coli* strains were cultivated in LB medium (components purchased from BD Biosciences or Fisher Scientific) supplemented with the appropriate antibiotics. DNA isolation and manipulations were carried out according to standard methods for *E. coli* and *Salinispora*.

Construction of a gene disruption cassette containing the *attB* attachment site

A 500-bp fragment containing the ΦC31 phage attachment site *attB* was amplified from the gDNA of *Streptomyces coelicolor* M1152 (Gomez-Escribano and Bibb 2011) with primers attBreg-F and attBreg-R (Table S1). The gene disruption cassettes *acc(3)IV* ($Apra^R$) and $oriT_{RK2}$ were jointly amplified from pIJ773 (Gust et al. 2003) with the primers P1reg-F and P2reg-R (Table S1). These two PCR products were assembled by an overlap extension PCR using the primers

Table 1 Plasmids and bacterial strains used in this work

Plasmid/strains	Description	References
Plasmids		
BHXS1782	pCC1FOS fosmid that carries a 44-kb genomic region containing the entire salinosporamide BGC from *S. tropica* CNB-440	Eustaquio et al. 2009
pMXT13	pCAP03 derivative that carries a 26-kb genomic region containing the entire thiolactomycin BGC from *S. pacifica* CNS-863	Tang et al. 2015
pIJ773	A plasmid containing the cassette *aac(3)IV* ($Apra^R$) + *oriT*	Gust et al. 2003
pKD20	λ-RED (*gam*, *bet*, *exo*), *bla*, *araC*, *oriR101*, *rep101ts*	Datsenko and Wanner 2000
pUZ8002	*neo*, RP4	Paget et al. 1999
pUB307	Self-transmissible plasmid that mobilizes other plasmids in trans for DNA transfer into hosts: RP4, *neo*	Flett et al. 1997
pMXT01	*acc(3)IV*, P*ermE*, FRT, *bla*	This study
pMXT04	BHXS1782 derivative in which *salA-C* has been replaced by the knock-out cassette from pMXT01	This study
pMXT10	pMXT04 derivative in which *acc(3)IV* (ApraR) has been removed by FLP-FRT recombination	This study
pMXT11	pMXT10 derivative in which chloramphenicol resistance gene *cat* has been replace by the cassette containing *oriT* and *aac(3)IV* (ApraR) from pIJ773	This study
Strains		
S. tropica CNB-440	Wild-type, ATCC BAA-916/DSM 44818	Udwary et al. 2007
S. tropica CNB-4401	*S. tropica* CNB-440 derivative, Δ*sal* and ΦC31*attB*	This study
S. tropica CNB-4401/*tlm*	*S. tropica* CNB-4401 derivative, Δ*sal* and ΦC31*attB* with *tlm* integration	This study
E. coli Top10	Host strain for routine cloning: F-, Δ*(araA-leu)7697*, *(araD139)B/r*, Δ*(codBlacI)3*, ϕ80d*lacZ58(M15)*, *galK0*, *mcrA0*, *galU*$^-$, *recA1*, *endA1*, *nupG*$^-$, *rpsL-(strR)*, Δ*(mcrC-mrr)715*	Invitrogen
E. coli BW25113	Host strain for λ-RED recombination, K12 derivative: F-, Δ*(araDaraB)567*, Δ*lacZ4787*(::rrnB-3), λ-, *rph-1*, Δ*(rhaD-rhaB)568*, *hsdR514*	Datsenko and Wanner 2000
E. coli BT340	Host strain for FLP recombination: F-, Δ*(argF-lac)169*, ϕ*80dlacZ58*(M15), *glnV44*(AS), λ-, *rfbC1*, *gyrA96*(NalR), *recA1*, *endA1*, *spoT1*, *thi-1*, *hsdR17*, pCP20	Cherepanov and Wackernagel 1995
E. coli ET12567	DNA methylation deficient donor strain for conjugation: F2 *dam13*::Tn9, *dcm*, *6 hsdM*, *hsdR*, *recF143*, *zjj-202*::Tn10, *galK2*, *galT22 ara-14*, *laacY1*, *xyl-5*, *leuB6*, *thi-1*, *tonA31*, *rpsL 136*, *hisG4*, *tsx-78*, *mtl-1*, *glnV44*	Flett et al. 1997

attBreg-F and P2reg-R (Table S1). The final disruption cassette was further cloned into pCR™2.1-TOPO® vector (Fisher Scientific, Hampton, NH, USA) to give pMXT01 and was verified by restriction analysis and sequencing.

Gene replacement on the fosmid BHXS1782 containing the salinosporamide gene cluster

The disruption cassette containing the ΦC31 phage attachment site *attB*, *acc(3)IV* ($Apra^R$), and $oriT_{RK2}$ was jointly amplified from the 1961-bp NotI/SpeI fragment of pMXT01 with the primer pair of salA-C_KO-F/salA-C_KO-R (Table S1). The PCR products were used to replace the genes *salA-C* on the fosmid BHXS1782 (Eustaquio et al. 2009) in *E. coli* BW25113/pKD20 (Datsenko and Wanner 2000) using the PCR targeting system. The resulting fosmid pMXT04 was confirmed by restriction analysis. Excision of the apramycin resistance cassette was performed in *E. coli* BT340 (Cherepanov and Wackernagel 1995), taking advantage of the FLP recognition sites surrounding the resistance cassette. *E. coli* clones containing the fosmid pMXT10 were screened

Allelic exchange in *Salinispora tropica* CNB-440

The modified fosmid pMXT11 were introduced into *E. coli* ET12567/pUZ8002 (Paget et al. 1999) by electroporation and then transferred to *S. tropica* CNB-440 by conjugation. Apramycin resistant ($Apra^R$) exconjugants were identified, indicating a single-crossover allelic exchange in *S. tropica* CNB-440. Single colonies were picked and streaked on A1 agar plates without any antibiotics. Apramycin-sensitive col- onies were further screened by PCR to obtain the double- crossover allelic exchange mutant *S. tropica* CNB-4401.

Heterologous expression of the thiolactomycin gene cluster

The plasmid pMXT13 (Tang et al. 2015), containing the thiolactomycin BGC (*tlm*), was transferred into *E. coli* ET12567 and introduced into *S. tropica* CNB-4401 by triparental intergeneric conjugation with the help of *E. coli* ET12567/pUB307 (Flett et al. 1997). Two hundred microliters overnight culture of *E. coli* was inoculated into 10-ml fresh LB medium plus appropriate antibiotics and grown for 3–4 h at 37 °C until an OD600 of 0.4 was reached. The cells were washed twice with 10 ml of LB to remove antibiotics and resuspended in 0.5 ml of LB. Simultaneously, 10 μl (10^8) of *S. tropica* spores was added to 200-μl A1 broth and heat shocked at 50 °C for 10 min. The spore suspension was mixed with 100 μl *E. coli* ET12567/pMXT13 and 100 μl *E. coli* ET12567/pUB307. The mixture was plated on an A1 agar plate and incubated at 30 °C for 16–20 h. Each plate was then overlaid with 1 ml water containing 4 mg nalidixic acid (to selectively kill *E. coli*) and 2 mg kanamycin (for selection of mutants). Following incubation at 30 °C for 7–9 days, kanamycin-resistant clones were selected, confirmed by PCR, and designated as *S. tropica* CNB-4401/*tlm*.

Production, extraction, and detection of salinosporamide A and thiolactomycins

Five milliliters of A1 medium was inoculated with 10-μl (10^8) spore suspension of *S. pacifica* CNS-863, *S. tropica* CNB-440, *S. tropica* CNB-4401, or a derivative thereof. The cultures were incubated for 3–4 days at 30 °C at 220 rpm. One milliliter of preculture was inoculated into 50-ml A1 medium using wide bore pipet tips. The cultures were incubated for 9 days at 30 °C with 220-rpm shaking. *S. coelicolor* M1152/*tlm* was cultured as described previously (Tang et al. 2015). The culture supernatant was adjusted to pH 4 with acetic acid and subsequently extracted with an equal volume of ethyl acetate. The organic phase was evaporated, and extracts were dissolved in 0.5 mL methanol (MeOH) and filtered through Acrodisc MS PTFE Syringe filters (Pall Inc., Ann Arbor, MI, USA) prior to HPLC and LC-MS analysis. For HPLC analysis, 10 μl of the MeOH extract was injected into an Agilent 1200 series analytical HPLC system monitored by UV lamp using a Luna 100A-C18 column (5 μm, 250 × 4.6 mm; Phenomenex, US) as follows: 0–23 min, 10–60% B; 23–28 min, 60–100% B; 29–34 min, 100% B; 34–35 min, 100–10% B; 36–40 min, 10% B (solvent A, water/trifluoroacetic acid (999:1); solvent B, acetonitrile/trifluoroacetic acid (999:1)). For LC-MS analysis, 5 μl of each MeOH extract was injected onto a Phenomenex Luna C18 reverse-phase HPLC column (5 μm, 250 mm × 4.6 mm; Phenomenex, USA) and analyzed with an Agilent 6530 Accurate-Mass LC-MS coupled to an Agilent 1260 LC system by gradient elution (A, acetonitrile with 0.1% formic acid (FA); B, H_2O with 0.1% FA: 35–70% A over 23 min, 70–100% A from 23 to 28 min, and 100% A from 28 to 33 min; 0.7 ml/min). Q-TOF MS settings during the LC gradient were as follows: acquisition—mass range *m/z* 100–1700, MS scan rate 1 s^{-1}, MS/MS scan rate 2 s^{-1}, fixed collision energy 20 eV; source—gas temperature 300 °C, gas flow 11 l/min; nebulizer 45 psig, ion polarity positive; scan source parameters—VCap 3000, Fragmentor 100, Skimmer 65, OctopoleRFPeak 750. The MS was autotuned using Agilent tuning solution in positive mode before each measurement. LC(DAD) data were analyzed with ChemStation software (Agilent), and MS data were analyzed with MassHunter software (Agilent).

Results

Adaptation of the ϕC31 int/*attP-attB* integration system in *S. tropica* CNB-440

ϕC31 integrase, a site-specific bacteriophage serine recombinase, catalyzes the insertion of a DNA fragment containing an *attP* site into a chromosomal *attB* site via recombination (Smith et al. 2010). As *attB* sites are widely distributed in a broad range of *Streptomyces* species (Combes et al. 2002; Smith et al. 2004), ϕC31 int/*attP* has been heavily employed in constructing vectors such as pSET152 (Flett et al. 1997), pCAP01 (Yamanaka et al. 2014), and pCAP03 (Tang et al. 2015) that can be integrated into the chromosome of various *Streptomyces* species for stable maintenance. In contrast to terrestrial *Streptomyces*, marine *Salinispora* species lack the *attB* attachment site in their genomes (Lechner et al. 2011). Previously, we identified three pseudo-*attB* sites in the genome of *S. tropica* CNB-440 that were used to integrate a ϕC31 phage-based vector carrying a single gene (Lechner et al. 2011). However, despite multiple attempts, we could not obtain exconjugants upon transfer of ϕC31 phage-based vectors carrying modest or large size BGCs (> 30 kb) into *S. tropica* CNB-440. This observation is consistent with an earlier study, which reported that insertion into the *attB* site is 300 times more efficient than into a pseudo-*attB* site (Combes et al. 2002).

We thus decided to introduce an authentic *attB* site into the chromosome of *S. tropica* CNB-440 by RP4-mediated intergeneric conjugation. We first amplified a 500-bp DNA fragment containing the *attB* site from the *g*DNA of *Streptomyces coelicolor* M1152. The PCR product was assembled with a cassette amplified from plasmid pIJ773 (Gust et al. 2003) to obtain an *attB*-containing disruptive cassette with an *attB* site, *aac(3)IV* ($Apra^R$), $oriT_{RK2}$, and two terminal FLP recognition target (FRT) sequences (Figure S1). This cassette was then cloned into pCR™2.1-TOPO® vector to form pMXT01 (Figure S1). pMXT01 was digested with NotI/SpeI, and the

1961-bp fragment was purified by agarose gel electrophoresis to eliminate intact plasmid and used as a template for cassette amplification.

Deletion of endogenous BGCs has been used as a general strategy to (1) increase the metabolic flux of precursor availability and productivity of heterologously expressed BGCs and (2) simplify the identification and characterization of heterologous products by activity screening and metabolite profiling (Gomez-Escribano and Bibb 2011; Komatsu et al. 2010). Thus, the fosmid BHXS1782 (Eustaquio et al. 2009) with a 44-kb insert containing the salinosporamide BGC (*sal*) was chosen to carry the *attB*-containing disruptive cassette into *S. tropica* CNB-440 to generate a double-crossover mutant. We first replaced assembly line biosynthesis genes *salA* to *salC* on the fosmid BHXS1782 with the *attB*-containing disruptive cassette by Red/ET-mediated recombination, generating the fosmid pMXT04 (Fig. 1). Subsequently, the region of *aac(3)IV* ($Apra^R$) plus $oriT_{RK2}$ was eliminated in vivo using FLP-mediated recombination to form pMXT10 (Fig. 1). Additionally, the cassette containing *aac(3)IV* ($Apra^R$) plus $oriT_{RK2}$ was used to replace the chloramphenicol resistance gene *cat*, as *aac(3)IV* is a better selection marker for generating single-crossover mutants in *Salinispora* and $oriT_{RK2}$ is essential for *E. coli*-mediated conjugation. The resulting fosmid, pMXT11, was introduced into *S. tropica* CNB-440 by conjugation (Figure S2). Single-crossover mutants were selected for their apramycin resistance. Intergeneric conjugation between *E. coli* ET12567/pUZ8002/pMXT11 and *S. tropica* CNB-440 produced 10–20 colonies with apramycin resistance per conjugation using 10 µl (10^8) *S. tropica* spores. Although the efficiency is much lower than for conjugation with *Streptomyces* spores (10^4–10^5 using 10^8 spores), this might be due to the high concentration of sea salt in the A1 agar plate, which is essential for the growth of *S. tropica*.

In order to obtain a double-crossover mutant, single-crossover mutants were streaked for single colonies on A1 agar plates without antibiotics. Fifty colonies were tested for apramycin sensitivity, and strains no longer resistant to apramycin were obtained for further verification. Total DNA was isolated from selected colonies, and a pair of primers flanking *salA-C* was designed for PCR confirmation. As designed, amplification of a double-crossover deletion mutant gave a 626-bp PCR product (Fig. 2). Furthermore, sequencing of the PCR product confirmed the presence of the *attB* site in the mutant strain, which was designated *S. tropica* CNB-4401. In the case of wild-type strain *S. tropica* CNB-440, no PCR products were obtained, as the targeted region was too large (9830 bp) (Fig. 2).

We also compared the production of salinosporamide A, the major product of the *sal* BGC, from *S. tropica* CNB-440 and *S. tropica* CNB-4401. Because *salA-C* code for the salinosporamide enzymatic assembly line, *S. tropica* CNB-4401 should lose all ability to produce salinosporamide A. After cultivation, extracts were subjected to high-resolution LC-ESI-MS analysis. In contrast to the wild-type strain *S. tropica* CNB-440, production of salinosporamide A (Fig. 2) and all other analogues was completely abolished in the double-crossover deletion mutant *S. tropica* CNB-4401. This result further confirmed the successful replacement of *salA-C* by the *attB*-containing disruptive cassette in the strain *S. tropica* CNB-4401. Additionally, the HPLC chromatogram of the ethyl acetate extract of the culture supernatant from *S. tropica* CNB-4401 contains a small number of peaks and a low baseline under various UV detection wavelengths (Figure S3). This would simplify downstream processes in detection and purification of heterologously produced metabolites.

Heterologous expression of an exogenous biosynthetic gene cluster in *S. tropica* CNB-4401

In order to assess the performance of *S. tropica* CNB-4401 as a heterologous host in expressing large BGCs, we selected the recently characterized BGC for the fatty acid synthase inhibitor thiolactomycin (*tlm*) that is only present in strains of *S. pacifica*. Other high-profile BGCs from *Salinispora*, such as the salinosporamide, lymphostin, and lomaiviticin BGCs, are endogenous to *S. tropica* CNB-440 and therefore were not selected for host validation. The *tlm* BGC encodes an unconventional polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) hybrid assembly line, which represents an important class of natural products (Tang et al. 2015). Furthermore, the *tlm* genes were previously cloned into the integration vector pCAP03 to generate pMXT13 and successfully expressed in the *Streptomyces coelicolor* M1152 ʙsuperhost,^ albeit at lower levels than the wild-type producer (Tang et al. 2015; Gomez-Escribano and Bibb 2011). Thus, the *tlm* BGC was directly introduced to the *S. tropica* CNB-4401 host via integration into the newly introduced genomic ϕC31 *attB* site. Kanamycin-resistant clones were selected and designated *S. tropica* CNB-4401/*tlm*. Intergeneric conjugation among *E. coli* ET12567/pUB307 (helper), *E. coli* ET12567/pMXT13, and *S. tropica* CNB-4401 yielded around 5 to 10 Kan^R exconjugants (experiment in sextuplicate) per conjugation using 10 µl (10^8) *S. tropica* CNB-4401 spores (Figure S4).

Although the integration efficiency is again much lower than that for *Streptomyces* spores, our data are comparable with the allelic exchange rate for making single-crossover mutants of *S. tropica* CNB-440 (10–20 mutants by biparental conjugation). This data thus supports the hypothesis that conjugation efficiency may be hampered by high salt concentrations necessary for the growth of some marine bacteria. No exconjugants were obtained from identical attempts to integrate the *tlm* BGC into the genome of *S. tropica* CNB-440 (Figure S4, experiment in sextuplicate). Therefore, our results strongly indicate that *S. tropica* CNB-4401 is capable of being readily integrated with large exogenous DNA fragments using

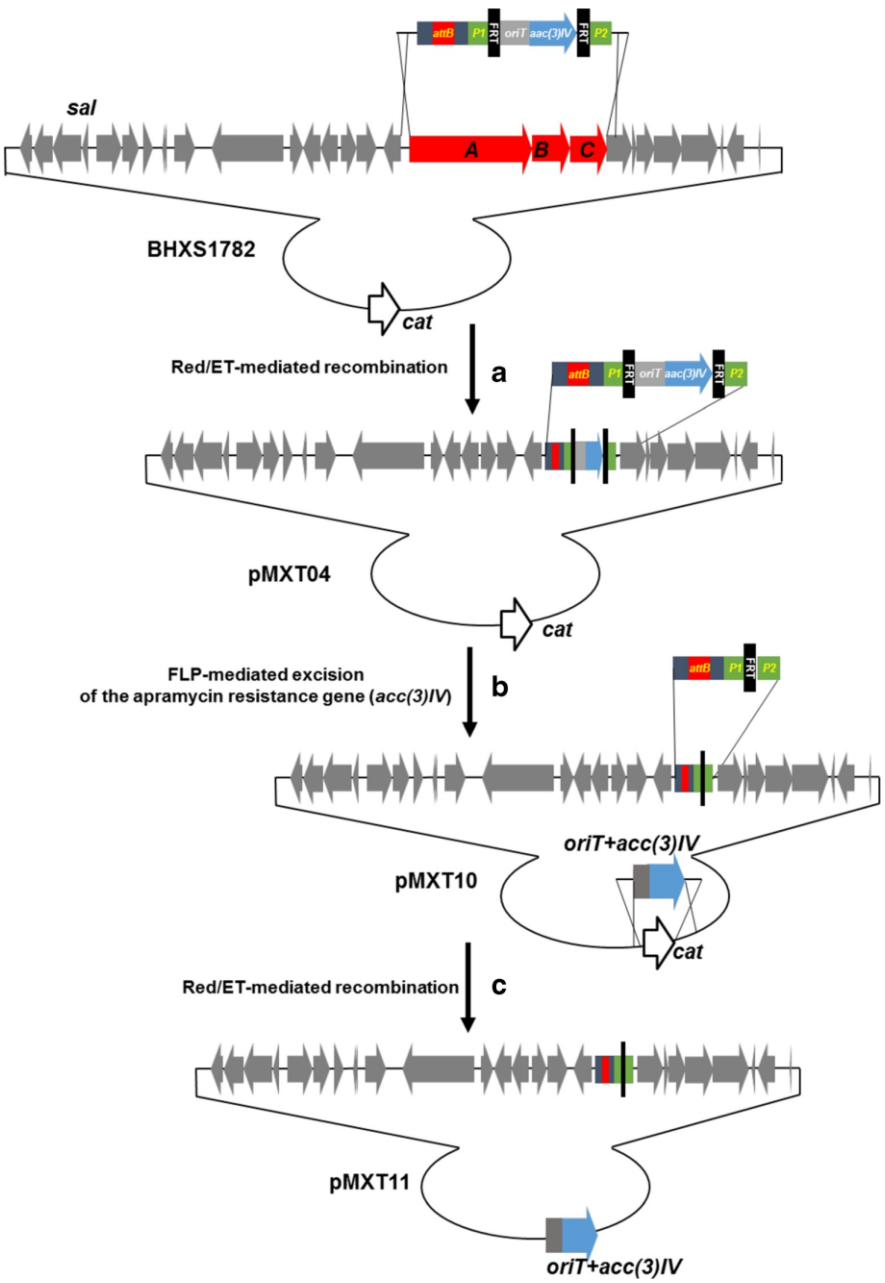


Fig. 1 Schematic diagram depicting the replacement of *salA-C* by Red/ET-mediated recombination with the *attB*-containing cassette (a); subsequent removal of the resistance marker by FLP-mediated excision (b); and replacement of the fosmid backbone chloramphenicol resistance gene *cat* with *acc(3)IV* ($Apra^R$) and *oriT* of RK2 to generate fosmid pMXT11 to simultaneously knock-in the *attB* attachment site and knock-out *salA-C* in *S. tropica* CNB-440 (c)

the ϕC31 int/*attP-attB* integration system, while *S. tropica* CNB-440 is not. This strategy may be widely applicable in engineering other bacteria as integratable heterologous hosts that are directly compatible with established ϕC31-containing vectors (Martinez et al. 2004).

With the successful integration of the *tlm* biosynthetic pathway in *S. tropica* CNB-4401, we cultured, extracted, and analyzed the supernatant extracts of the *S. tropica* CNB-4401/*tlm* heterologous host by HPLC and LC-MS. The production of thiolactomycin (1) and three analogues (10-methyl thiolactomycin (2), 11-methyl thiolactomycin (3), and thiotetromycin (4)) was identified from the extracts of *S. tropica* CNB-4401/*tlm* by comparison with thiolactomycins produced by the native producer *S. pacifica* CNS-863 as well as the heterologous producer *S. coelicolor* M1152/*tlm*. All four thiolactomycin analogues were previously fully characterized by 1D and 2D NMR experiments (Fig. 3). Corresponding parent mass peaks for (1) with *m/z* 211.0789 $[M+H]^+$ (calcd for

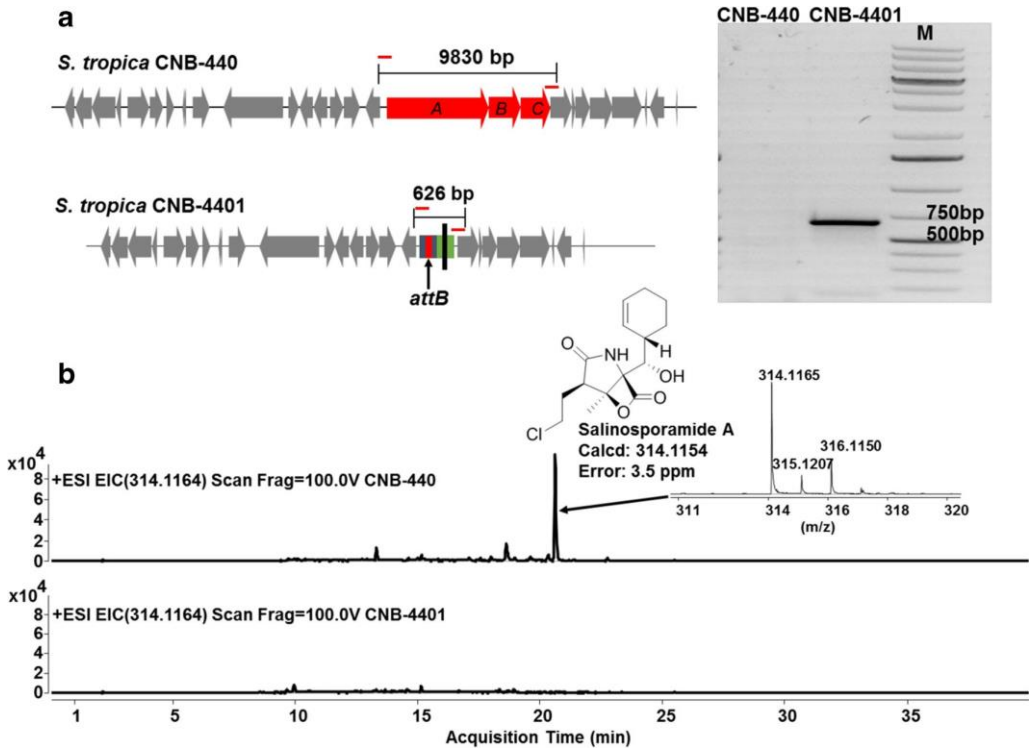


Fig. 2 Verification of *S. tropica* CNB-4401 by PCR amplification (a) and loss of production of salinosporamide A observed by LC-MS (b)

$C_{11}H_{15}O_2S$, 211.0793), (2) with *m/z* 225.0944 $[M+H]^+$ (calcd for $C_{12}H_{17}O_2S$, 225.0949), (3) with *m/z* 225.0942 $[M+H]^+$ (calcd for $C_{13}H_{19}O_2S$, 225.0949), and (4) with *m/z* 239.1097 $[M+H]^+$ (calcd for $C_{12}H_{17}O_2S$, 239.1106) were detected by high-resolution LC-MS (Figure S5). Thus, this is the first report of using the marine actinomycete *Salinispora* as an integratable host for natural product BGC heterologous expression.

In contrast to *S. coelicolor* M1152/*tlm*, the production of thiolactomycin (1) from *S. tropica* CNB-4401/*tlm* and *S. pacifica* CNS-863 was approximately 3-fold higher (Fig. 3), while production of (3) and (4) were only slightly lower. Overall, the production pattern of the four thiolactomycin analogues in *S. tropica* CNB-4401/*tlm* was very similar to that of the native producer, *S. pacifica* CNS-863, and subtly different from that of *S. coelicolor* M1152 (Fig. 3). This difference in production levels as well as PKS extender unit incorporation suggests that heterologous host context with respect to genetics, biochemistry, and metabolism influences heterologous thiolactomycin production and may also influence heterologous production of products associated with other BGCs.

Discussion

In this study, we selected *S. tropica* CNB-440 as the starting point for developing the first marine actinomycete heterologous host. In order to establish an integratable expression host with a cleaner background for convenient metabolic profiling, we introduced a disruptive cassette containing the well-established ΦC31 phage attachment site *attB* to replace three essential salinosporamide biosynthetic genes *salA–C* in *S. tropica* CNB-440. This newly established mutant strain *S. tropica* CNB-4401, therefore, possesses an authentic *attB* attachment site to integrate large exogenous BGCs through ϕC31 integrase activity for heterologous expression. We validated this new heterologous host with the successful integration and expression of the thiolactomycin BGC from several *S. pacifica* strains. To the best of our knowledge, this is the first example of using a marine actinomycete as a heterologous host for BGC expression.

The overall yield of all thiolactomycins produced heterologously by *S. tropica* CNB-4401 is very similar to the yield from the native producer *S. pacifica* CNS-863 and approximately 3 times higher than the yield from *S. coelicolor* M1152 (Fig. 3). This is despite the fact that the *S. coelicolor* M1152 ʙsuperhost^ has been subjected to about a dozen rounds of genetic manipulation to delete endogenous pathways and introduce a point mutation to *rpoB* that has been empirically shown to enhance antibiotic production in *Streptomyces* (Gomez-Escribano and Bibb 2011).

The different ratios of compounds 1–4 observed between *S. tropica* CNB-4401/*tlm* (or *S. pacifica* CNS-863) and *S.*

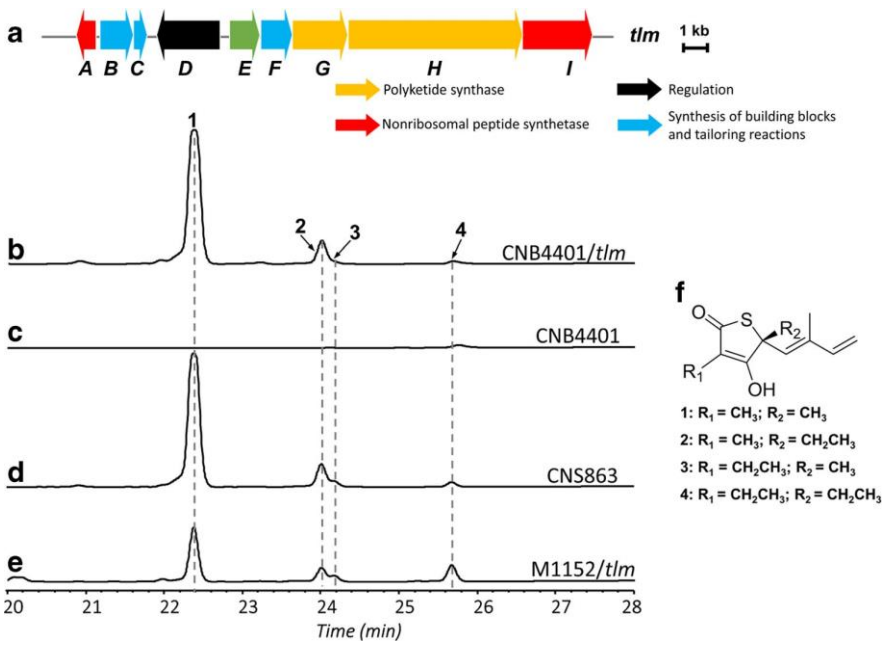


Fig. 3 Detection of thiolactomycin and its analogues in various producing strains by HPLC. a Schematic illustrating the thiolactomycin (*tlm*) biosynthetic gene cluster. HPLC chromatograms of ethyl acetate extracts from *S. tropica* CNB-4401/*tlm* (b), *S. tropica* CNB-4401 (c), *S. pacifica* CNS-863 (d), and *S. coelicolor* M1152/*tlm* (e). Detection at 239 nm. f Structures of thiolactomycin (1), 10-methylthiolactomycin (2), 11-methyl thiolactomycin (3), and thiotetromycin (4)

coelicolor M1152/*tlm* may be due to differences in methylmalonyl-CoA and ethylmalonyl-CoA precursor supply in these strains. However, previous gene deletion and heterologous expression experiments in *S. coelicolor* M1152 suggest that the pool of ethylmalonyl-CoA precursors used by the TlmH PKS to generate compounds 2–4 is primarily supplied by the pathway-specific crotonyl-CoA carboxylase (TlmB) and 3-hydroxybutyryl-CoA dehydrogenase (TlmC) (Tang et al. 2017). Thus, there may be additional mechanisms beyond precursor supply that influence TlmH precursor utilization, which results in a different profile of thiolactomycin production levels in *S. coelicolor* M1152 compared to *S. tropica* CNB-4401 and *S. pacifica* CNS-863. This supports our previous finding that host context can play an important role in heterologous compound production and provide biosynthetic insights (Zhang et al. 2017).

S. tropica expresses the unusual crotonyl-CoA carboxylase SalG that not only catalyzes the construction of ethylmalonyl-CoA, but also the unique PKS precursor chloroethylmalonyl-CoA (Eustaquio et al. 2009) and the rare propylmalonyl-CoA (Liu et al. 2009). We employed LC-MS to interrogate thiolactomycin analogues that may have incorporated these alternate substrates, but no new thiolactomycin analogues were identified, perhaps due to substrate selectivity of the *tlm* PKS assembly line. Alternatively, insertion of a large DNA fragment into the engineered attachment site, which is upstream of the crotonyl-CoA carboxylase SalG, may cause polar effects that disrupt expression of *salG*. Heterologous expression of large BGCs remains challenging, particularly for pathways that require rare precursors and/or utilize uncertain regulatory, biosynthetic, or resistance factors. Our previous work to develop improved direct cloning and heterologous expression vectors demonstrated that having access to genetically diverse heterologous host organisms will enhance the likelihood of achieving successful heterologous expression for biosynthetic studies and identification of new compounds (Zhang et al. 2017). As *S. tropica* CNB-4401 is a phylogenetically distinct actinomycete host that can now be easily leveraged using existing *Streptomyces* expression vectors for heterologous expression experiments, we believe that our study provides the natural product community an alternative and facile platform for discovering and engineering BGCs.

Acknowledgments We kindly thank W. Fenical and P. R. Jensen (Scripps Institution of Oceanography, UCSD) for providing the *S. tropica* strain, and P. R. Jensen and N. Ziemert (Scripps Institution of Oceanography, UCSD) for valuable discussion.

Author contributions X.T. and B.S.M. designed the research. X.T., J.J.Z., and B.S.M. performed the experiments and analyzed the data. X.T., J.J.Z., and B.S.M. wrote the manuscript.

Funding This study was supported by a graduate fellowship from the National Science Foundation to J.J.Z. and National Institutes of Health grants F31-AI129299 to J.J.Z. and R01-GM085770 and R01-AI117712 to B.S.M.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Ethical statement This article does not contain any studies with human participants or animals performed by any of the authors.

References

Amos GCA, Awakawa T, Tuttle RN, Letzel AC, Kim MC, Kudo Y, Fenical W, Moore BS, Jensen PR (2017) Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. Proc Natl Acad Sci U S A 114(52):E11121–E11130

Baltz RH (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol 37(8):759–772

Baltz RH (2016) Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other *actinomycetes*. J Ind Microbiol Biotechnol 43(2–3):343–370

Berdy J (2005) Bioactive microbial metabolites. J Antibiot (Tokyo) 58(1): 1–26

Chater KF (2016) Recent advances in understanding *Streptomyces*. F1000Res 5:2795

Cherepanov PP, Wackernagel W (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158(1):9–14

Combes P, Till R, Bee S, Smith MC (2002) The *Streptomyces* genome contains multiple *pseudo-attB* sites for the (phi)C31-encoded site-specific recombination system. J Bacteriol 184(20):5746–5752

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97(12):6640–6645

Eustaquio AS, McGlinchey RP, Liu Y, Hazzard C, Beer LL, Florova G, Alhamadsheh MM, Lechner A, Kale AJ, Kobayashi Y, Reynolds KA, Moore BS (2009) Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. Proc Natl Acad Sci U S A 106(30):12295–12300

Flett F, Mersinias V, Smith CP (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. FEMS Microbiol Lett 155(2): 223–229

Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Muller R, Stewart AF, Zhang Y (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol 30(5):440–446

Gomez-Escribano JP, Bibb MJ (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4(2):207–215

Gomez-Escribano JP, Bibb MJ (2014) Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J Ind Microbiol Biotechnol 41(2):425–431

Gulder TA, Moore BS (2010) Salinosporamide natural products: potent 20S proteasome inhibitors as promising cancer chemotherapeutics. Angew Chem Int Ed Engl 49(49):9346–9367

Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci U S A 100(4):1541–1546

Horinouchi S (2007) Mining and polishing of the treasure trove in the bacterial genus *Streptomyces*. Biosci Biotechnol Biochem 71(2): 283–299

Jensen PR, Gontang E, Mafnas C, Mincer TJ, Fenical W (2005) Culturable marine actinomycete diversity from tropical Pacific Ocean sediments. Environ Microbiol 7(7):1039–1048

Jensen PR, Moore BS, Fenical W (2015) The marine actinomycete genus *Salinispora*: a model organism for secondary metabolite discovery. Nat Prod Rep 32(5):738–751

Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF (2015) Cas9-assisted targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. Nat Commun 6:8101

Kersten RD, Lane AL, Nett M, Richter TK, Duggan BM, Dorrestein PC, Moore BS (2013) Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. Chembiochem 14(8):955–962

Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF (2010) Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers 93(9):833–844

Komatsu M, Komatsu K, Koiwai H, Yamada Y, Kozone I, Izumikawa M, Hashimoto J, Takagi M, Omura S, Shin-ya K, Cane DE, Ikeda H (2013) Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol 2(7):384–396

Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A 107(6):2646–2651

Lechner A, Eustaquio AS, Gulder TA, Hafner M, Moore BS (2011) Selective overproduction of the proteasome inhibitor salinosporamide A via precursor pathway regulation. Chem Biol 18(12):1527–1536

Letzel AC, Li J, Amos GCA, Millan-Aguinaga N, Ginigini J, Abdelmohsen UR, Gaudencio SP, Ziemert N, Moore BS, Jensen PR (2017) Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. Environ Microbiol 19(9):3660–3673

Liu Y, Hazzard C, Eustaquio AS, Reynolds KA, Moore BS (2009) Biosynthesis of salinosporamides from alpha,beta-unsaturated fatty acids: implications for extending polyketide synthase diversity. J Am Chem Soc 131(30):10376–10377

Martinez A, Kolvek SJ, Yip CL, Hopke J, Brown KA, MacNeil IA, Osburne MS (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. Appl Environ Microbiol 70(4):2452–2463

Miyanaga A, Janso JE, McDonald L, He M, Liu H, Barbieri L, Eustaquio AS, Fielding EN, Carter GT, Jensen PR, Feng X, Leighton M, Koehn FE, Moore BS (2011) Discovery and assembly-line biosynthesis of the lymphostin pyrroloquinoline alkaloid family of mTOR inhibitors in *Salinispora* bacteria. J Am Chem Soc 133(34):13311–13313

Newman DJ, Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79(3):629–661

Ongley SE, Bian X, Neilan BA, Muller R (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. Nat Prod Rep 30(8):1121–1138

Paget MS, Chamberlin L, Atrih A, Foster SJ, Buttner MJ (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). J Bacteriol 181(1):204–211

Ren H, Wang B, Zhao H (2017) Breaking the silence: new strategies for discovering novel natural products. Curr Opin Biotechnol 48:21–27

Rutledge PJ, Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol 13(8):509–523

Schorn MA, Alanjary MM, Aguinaldo K, Korobeynikov A, Podell S, Patin N, Lincecum T, Jensen PR, Ziemert N, Moore BS (2016) Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. Microbiology 162(12):2075–2086

Smith MC, Brown WR, McEwan AR, Rowley PA (2010) Site-specific recombination by phiC31 integrase and other large serine recombinases. Biochem Soc Trans 38(2):388–394

Smith MC, Till R, Brady K, Soultanas P, Thorpe H, Smith MC (2004) Synapsis and DNA cleavage in phiC31 integrase-mediated site-specific recombination. Nucleic Acids Res 32(8):2607–2617

Tang X, Li J, Millan-Aguinaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS (2015) Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem Biol 10(12):2841–2849

Tang X, Li J, Moore BS (2017) Minimization of the thiolactomycin biosynthetic pathway reveals that the cytochrome P450 enzyme TlmF is required for five-membered thiolactone ring formation. Chembiochem 18(12):1072–1076

Udwary DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, Fenical W, Jensen PR, Moore BS (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. Proc Natl Acad Sci U S A 104(25):10376–10381

Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci U S A 111(5):1957–1962

Zhang X, Lu C, Bai L (2017) Conversion of the high-yield salinomycin producer *Streptomyces albus* BK3-25 into a surrogate host for polyketide production. Sci China Life Sci 60(9):1000–1009

Zhang JJ, Tang X, Zhang M, Nguyen D, Moore BS (2017) Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in gram-negative bacteria. MBio 8(5):e01291–e01217

Ziemert N, Lechner A, Wietz M, Millan-Aguinaga N, Chavarria KL, Jensen PR (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. Proc Natl Acad Sci U S A 111(12):E1130–E1139

Electronic Supplementary Material for:

Applied Microbiology and Biotechnology

Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters

Jia Jia Zhang[1], Bradley S. Moore*[1,2], Xiaoyu Tang*[1,3]

[1]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA

[2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA, USA.

[3]Genomic Medicine, J. Craig Venter Institute, La Jolla, California, USA

* Corresponding Authors:

Xiaoyu Tang, Tel: +1(858)200-1888. Fax: +1(858)534-1318, and E-mail: xtang@microbechembio.org

Bradley S. Moore, Tel.: +1(858)822-6650. Fax: +1(858)534-1318, and E-mail: bsmoore@ucsd.edu

Table S1. Primers used in this work

Name	Sequence
attBreg-F	CTGGGTGTCGCCGTTGGT
attBreg-R	GGTCGACGGATCCCCGGAATAGATGGTCGATTTCGCCAT
P1reg-F	ATGGCGAAATCGACCATCTATTCCGGGGATCCGTCGACC
P2reg-R	TGTAGGCTGGAGCTGCTTC
salA-C_KO-F	TCACCAGTCGGTGCGCCAGAACCGTCGGCGCGCGAACGGATTAAACCTGGGTC
salA-C_KO-R	ATGTCTGGTGAGCAAGCTGAGCACGGTGTCCACGCGATCATGTAGGCTGGAGC
SalA-C-test650-F	ACCAGTCGGTGCGCCAGAAC
SalA-C-test650-R	GAGTTATCGAGATTTTCAGGAGCTAAG
Cat-knockout-F1	AACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAAAAACTTCGAAGTTC
Cat-knockout-R1	TTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAACTTATGAGCTCAGC

Figure S1. Schematic diagram depicting the assembly of a disruptive cassette containing the *attB* attachment site.

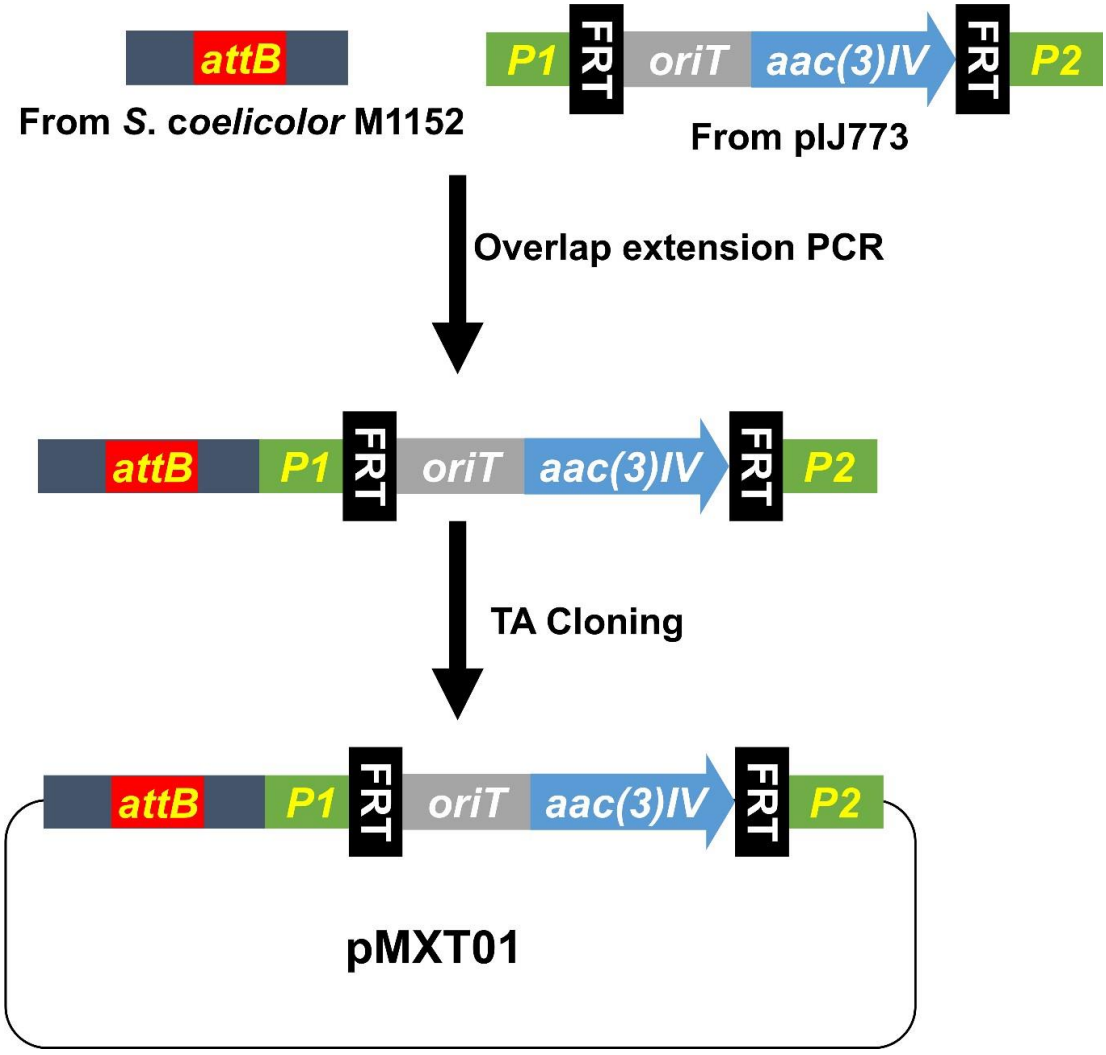


Figure S2. Schematic diagram depicting gene replacement in *S. tropica* CNB-440 to generate *S. tropica* CNB 4401.

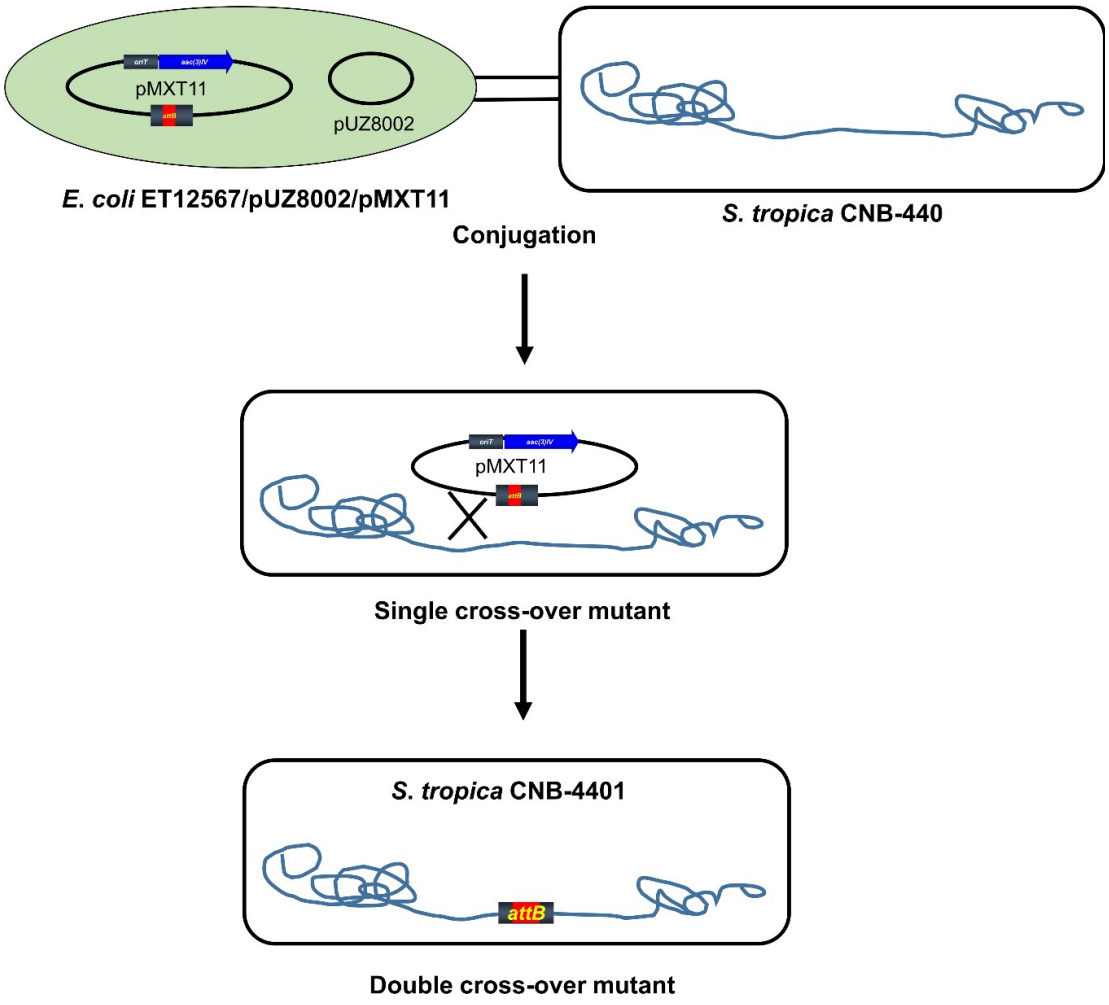


Figure S3. HPLC profiles of the supernatant extracts of CNB-4401 under the UV wavelength of (A) 210 nm, (B), 239 nm, (C) 254 nm, and (D) 280 nm.

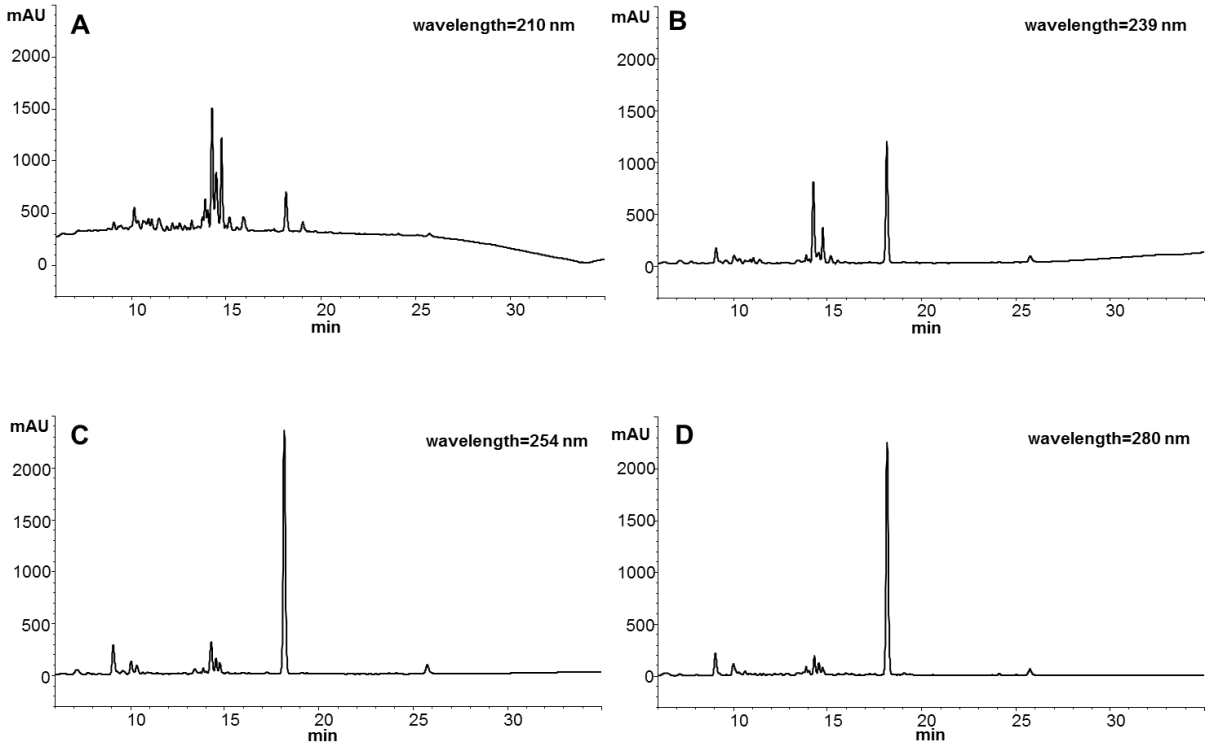


Figure S4. Conjugation of pMXT13 containing the *tlm* biosynthetic gene cluster to (A) *S. tropica* CNB-440 and (B) *S. tropica* CNB-4401. Exconjugants identified on the A1 agar plate are marked with red circles.

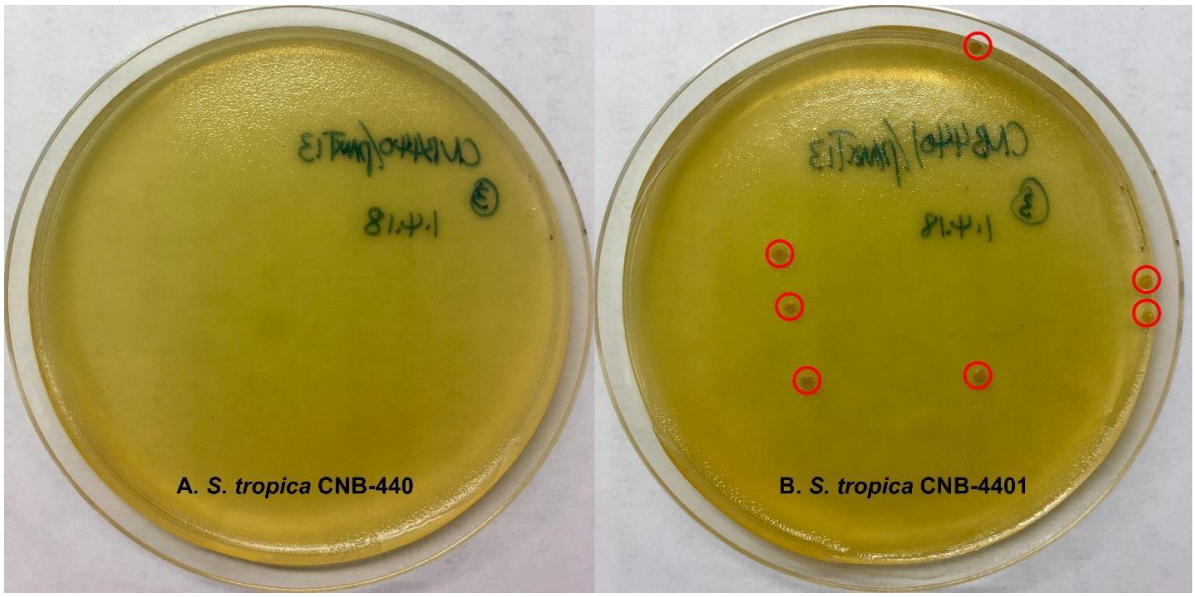
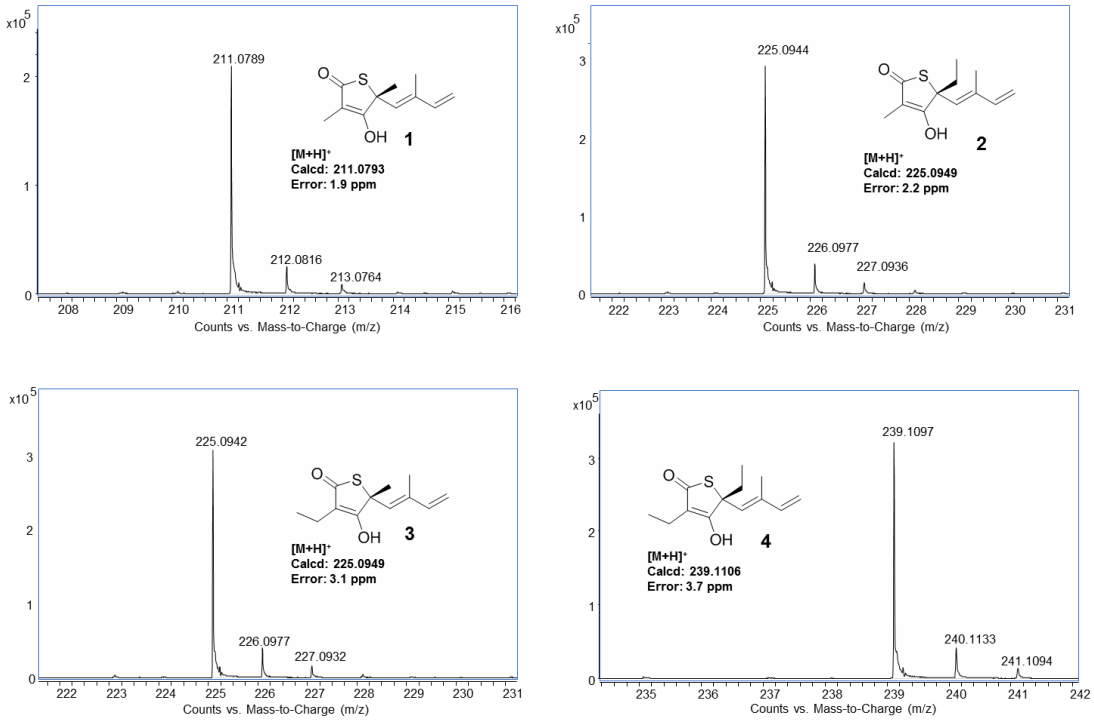


Figure S5. High-resolution mass spectrometry analysis of thiolactomycin (**1**), 10-methyl thiolactomycin (**2**), 11-methyl thiolactomycin (**3**), and thiotetromycin (**4**) produced by *S. tropica* CNB-4401/*tlm*.



2.3 Acknowledgement

Chapter 2, in full, is a reprint of the material as it appears in *Applied Microbiology and Biotechnology*, Zhang, J. J.; Moore, B. S.; Tang, X., 2018. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3. Broad-host-range heterologous expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria.

3.1 Abstract

Chapter 3 describes the development and utilization of a broad-host-range vector for direct cloning and heterologous expression of natural product BGCs in Gram-negative proteobacterial hosts. Using this vector, we express the BGC for the purple pigment violacein from the marine γ-proteobacterium *Pseudoalteromonas luteoviolacea* 2ta16 across proteobacterial hosts of varying phylogenetic relatedness. We conclude, based on our assessment of host "quality," that phylogeny is not an accurate predictor of compatibility, as natural product biosynthesis in a heterologous organism may rely on host factors such as regulatory and/or biosynthetic elements that may be difficult to predict. Furthermore, as the effectiveness of these elements is best determined empirically, we believe the ability to express BGCs across various hosts represents a valuable tool that greatly enhances the probability of successful heterologous reconstitution of BGCs, enabling discovery of new chemical entities and/or interrogation of novel biosynthetic mechanism.

3.2 Reprint of "Broad-host-range heterologous expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria."

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Broad-Host-Range Expression Reveals Native and Host Regulatory Elements That Influence Heterologous Antibiotic Production in Gram-Negative Bacteria

Jia Jia Zhang,[a] Xiaoyu Tang,[a,b] Michelle Zhang,[a] Darlene Nguyen,[a] Bradley S. Moore[a,c]

Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA[a]; Genomic Medicine, J. Craig Venter Institute, La Jolla, California, USA[b]; Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA[c]

ABSTRACT Heterologous expression has become a powerful tool for studying microbial biosynthetic gene clusters (BGCs). Here, we extend the transformation-associated recombination cloning and heterologous expression platform for microbial BGCs to include Gram-negative proteobacterial expression hosts. Using a broad-host-range expression platform, we test the implicit assumption that biosynthetic pathways are more successfully expressed in more closely related heterologous hosts. Cloning and expression of the violacein BGC from *Pseudoalteromonas luteoviolacea* 2ta16 revealed robust production in two proteobacterial hosts, *Pseudomonas putida* KT2440 and *Agrobacterium tumefaciens* LBA4404, but very little production of the antibiotic in various laboratory strains of *Escherichia coli*, despite their closer phylogenetic relationship. We identified a nonclustered LuxR-type quorum-sensing receptor from *P. luteoviolacea* 2ta16, PviR, that increases pathway transcription and violacein production in *E. coli* by ~60-fold independently of acyl-homoserine lactone autoinducers. Although *E. coli* harbors the most similar homolog of PviR identified from all of the hosts tested, overexpression of various *E. coli* transcription factors did not result in a statistically significant increase in violacein production, while overexpression of two *A. tumefaciens* PviR homologs significantly increased production. Thus, this work not only introduces a new genetic platform for the heterologous expression of microbial BGCs, it also challenges the assumption that host phylogeny is an accurate predictor of host compatibility.

IMPORTANCE Although Gram-positive heterologous hosts such as

Received 26 July 2017 **Accepted** 1 August 2017 **Published** 5 September 2017

Citation Zhang JJ, Tang X, Zhang M, Nguyen D, Moore BS. 2017. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. mBio 8:e01291-17. https://doi.org/10.1128/mBio.01291-17.

Editor Julian E. Davies, University of British Columbia

Address correspondence to Bradley S. Moore, bsmoore@ucsd.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Mohammad Seyedsayamdost, Princeton University; Emily Balskus, Harvard University.

Biosynthetic gene clusters (BGCs) for the production of specialized metabolites can occupy up to 10% of a bacterial genome (1), and DNA sequencing and new bioinformatics tools continue to uncover previously undetected biosynthetic potential (2). The metabolic cost of evolving, gaining, or maintaining a BGC is high, as some can exceed 100 kb in size and encode dozens of proteins that catalyze synchronized enzymatic reactions. Despite their prominence, most BGCs have not been characterized (3), and while the pharmacology of many natural products has been studied in animal or cancer cell lines, their biological role within their native context is often not understood. Thus, many questions regarding microbial BGCs remain to be explored.

Heterologous expression has become a powerful tool for the investigation of microbial BGCs for several reasons. First, BGCs have become easier to clone. Techniques that rely upon homologous recombination in yeast or *Escherichia coli* have greatly simplified the process of cloning large fragments of DNA (4, 5), and in the near future, many BGCs will likely be obtained directly from DNA synthesis. Second, heterologous expression of a defined set of genes allows for direct connection of BGCs to their small-molecule products and simultaneously establishes a boundary of essential elements. Third, utilizing a cloning and heterologous expression platform eliminates the need to develop new genetic tools for the manipulation of each new genus or species of interest. Instead, a set of universally applicable tools can be optimized and applied to BGCs from any source. Thus, targeted changes can readily be made to "cryptic" or "silent" BGCs that do not produce small molecules under normal laboratory culture conditions (3). Despite these advantages, successful heterologous production of small molecules from new BGCs remains a major bottleneck (6), especially for pathways that may utilize uncharacterized and/or nonclustered regulatory or biosynthetic elements.

One important consideration is the choice of an expression host. Despite numerous reports of successful heterologous expression of BGCs from various sources, researchers have historically relied upon a rather limited set of hosts, focusing heavily on the development and optimization of high-GC-content Gram-positive organisms such as *Streptomyces*, including genetically minimized strains of *Streptomyces coelicolor* and *Streptomyces avermitilis* (7, 8). These strains have been specifically engineered to optimize secondary metabolite production through deletion of native pathways to remove sinks for carbon and nitrogen and introduction of specific mutations in transcriptional and translational machinery that have been empirically shown to increase secondary metabolite production. Far less attention has been paid to Gram-negative hosts, despite growing interest in Gram-negative bacteria as sources of new antibiotics (9). Gram-negative hosts that have been successfully utilized for the heterologous expression of BGCs include *Myxococcus xanthus*, *Pseudomonas putida*, and *E. coli* (10). Although they have not been leveraged or optimized to the same extent as *Streptomyces* hosts, some of these organisms grow faster and are easier to work with.

A general assumption often stated or implied is that the best heterologous host will likely be an organism most similar or closely related to the original source of the BGC. Although theoretically appealing, this hypothesis has not been rigorously tested and there is limited empirical evidence to support this claim. Müller and coworkers hypoth- esized that the tubulysin gene cluster from *Cystobacter* sp. strain SBCb004 would be better expressed in *M. xanthus* than in *P. putida* because of the closer relationship between the two *Deltaproteobacteria*, and indeed, the authors observed 100-fold higher production in *M. xanthus* than in *P. putida* (11). However, production titers from *M. xanthus* were still lower than those from the native producer, and reasons for the observed differences in production level were not investigated. One possible rationale is that production of a natural product could involve chaperones or machinery not directly encoded within a gene cluster. Consequently, heterologous expression of BGCs often necessitates that host factors complement any missing activities. This leads to the impression that more closely related organisms will be more compatible heterologous hosts. However, whether this is generally the case remains to be established.

Previously, we and others constructed shuttle vectors that combine the capabilities of transformation-associated recombination (TAR) cloning in yeast, maintenance and

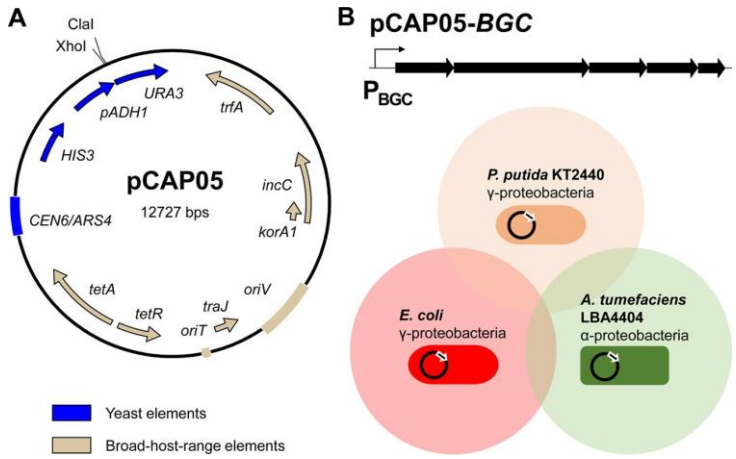


FIG 1 Direct cloning and proteobacterial broad-host-range heterologous expression of microbial BGCs. (A) Organization of TAR cloning and broad-host-range expression vector pCAP05 composed of elements for maintenance (*CEN6*/*ARS4*), selection (*HIS3*), and counterselection (*pAHD1*, *URA3*) in yeast (blue) and broad-host-range maintenance (*oriV*, *trfA*) and selection (*tetA*, *tetR*) in Gram-negative bacteria (tan). Additionally, *incC* and *korA1* regulate plasmid maintenance and *oriT* and *traJ* enable conjugal transfer of pCAP05. The vector can be linearized for cloning by digestion with ClaI and XhoI. (B) Schematic illustration of the broad-host-range heterologous expression concept. A generic microbial BGC cloned into pCAP05 is depicted above three proteobacterial expression hosts that were selected and validated in this study.

genetic manipulation in *E. coli*, and heterologous expression in *Streptomyces* hosts (12–14). Recently, we improved the efficiency of our cloning method and expanded our platform to include *Bacillus* hosts (15, 16). We have used this platform to successfully interrogate the biosynthesis and bioactivity of several natural-product BGCs (17–19). In this work, we extend the platform to include Gram-negative proteobacterial hosts, constructing a direct cloning and broad-host-range expression vector and validating three hosts: the alphaproteobacterium *Agrobacterium tumefaciens* LBA4404 and the *Gammaproteobacteria P. putida* KT2440 and *E. coli*. Using the BGC for the antibacterial pigment violacein from *Pseudoalteromonas luteoviolacea* 2ta16, which also belongs to the class *Gammaproteobacteria*, we (i) validate the extension of our genetic platform, (ii) test the assumption that more closely related hosts perform better, and (iii) investigate specific native and host factors that control violacein expression. Our results suggest that investigation of microbial BGCs will be aided by the development and use of diverse heterologous hosts, as the identity and activity of host elements that influence heterologous antibiotic production are difficult to predict.

RESULTS AND DISCUSSION

Construction of a direct cloning, broad-host-range expression vector for proteobacterial hosts. A new TAR cloning vector, pCAP05, was constructed by combining yeast elements for cloning with Gram-negative broad-host-range elements for heterologous expression (Fig. 1A). Broad-host-range elements originate from plasmid pRK442(H), a broad-host-range vector of the RK2 family. The RK2 replicon and its derivatives, which belong to the IncP incompatibility group, have been shown to replicate at low copy number (<30 in *E. coli*) in a wide range of Gram-negative bacteria through the *oriV* origin of replication and the essential *trfA* gene, which controls plasmid copy number and host range (20). Notably, however, RK2 replicons do not propagate in *Bacteroides* or *Myxococcus* (21). RK2 derivative plasmids have previously been utilized as versatile tools for broad-host-range screening of metagenomic DNA libraries (22, 23). Given the abundance of publicly available DNA sequence information and advances in bioinformatic detection of BGCs, direct cloning of targeted pathways has become an attractive alternative to untargeted library

generation and screening (3). Thus, pCAP05 includes elements for maintenance, selec- tion, and counterselection in *Saccharomyces cerevisiae* to leverage an improved TAR cloning method we described previously (15), which allows for the use of short capture arm sequences and improves cloning efficiency.

We assembled pCAP05 using Gibson Assembly and validated that the empty vector could be introduced into *S. cerevisiae* VL6-48N and selected against or for in histidine-deficient medium with or without 5-fluoroorotic acid (5-FOA), respectively (see Fig. S1A to D in the supplemental material). *HIS3* was chosen as the yeast selective marker to make pCAP05 orthogonal to our original series of capture vectors, which use TRP1 and are designed for the heterologous expression of BGCs in *Streptomyces* hosts (13, 15). We then validated that pCAP05 could be transferred to two proteobacterial expression hosts beyond *E. coli*, *P. putida* KT2440 and *A. tumefaciens* LBA4404, and selected for with tetracycline (Fig. S1E and F). *P. putida* KT2440 is a metabolically versatile and biosafety-certified host that has been successfully utilized for the heterologous expression of several BGCs (24). Its genome has been sequenced, and it is amenable to genetic manipulation. Further, KT2440 was shown to possess a promiscuous phosphopanteth- einyl transferase capable of posttranslationally activating carrier proteins involved in natural-product biosynthesis from myxobacteria and *Streptomyces*, which is an essential host trait for the heterologous expression of carrier protein-utilizing BGCs (25). *A. tu- mefaciens* LBA4404 is a biotechnological strain commonly used for plant transforma- tion. LBA4404 was constructed from its immediate precursor, *A. tumefaciens* LBA4213, by the deletion of a large region of the tumor-inducing (Ti) plasmid, rendering it avirulent (26). Taken together with various laboratory strains of *E. coli*, these hosts represent phylogenetically diverse *Proteobacteria* that are compatible with pCAP05 for broad-host-range heterologous expression testing of microbial BGCs (Fig. 1B).

Cloning and heterologous expression of *vio2ta16*. To validate our broad-host- range heterologous expression platform, we targeted the violacein BGC (*vio2ta16*) from the marine bacterium *P. luteoviolacea* 2ta16. *vio2ta16* was selected for several reasons. The BGC is compact in size (~8 kb), possesses a simple single-operon organization, and produces an easily detectable purple pigment. Violacein is produced by a range of *Betaproteobacteria* and *Gammaproteobacteria*, including *P. luteoviolacea*, but the violacein BGC is not broadly conserved across all members of the genus *Pseudoaltero-monas*. We efficiently TAR cloned *vio2ta16* into pCAP05 in yeast and transferred the construct to *E. coli* for verification by restriction digestion (Fig. S2A and B).

Within the phylum *Proteobacteria*, *A. tumefaciens* belongs to the taxonomic class *Alphaproteobacteria*, while *P. luteoviolacea*, *E. coli*, and *P. putida* all belong to *Gamma-proteobacteria*. Telescoping multiprotein phylogenetic analysis of 104 gammapro-teobacterial genomes resulted in a well-resolved phylogenetic tree that places the taxonomic orders to which *P. luteoviolacea* and *E. coli* belong (*Alteromonadales* and *Enterobacteriales*, respectively) within the same node, while the order of *P. putida* (*Pseudomonadales*) falls within a sister clade (27). Thus, broad-host-range expression of *vio2ta16* with the selected hosts allowed us to probe how host relationships across taxonomic order and class, as well as across different strains of *E. coli*, affect heterologous antibiotic production.

Despite the close phylogenetic relationship between *P. luteoviolacea* and *E. coli*, violacein was produced at very low levels by various laboratory strains of *E. coli*, including DH5*α* and BL21(DE3). Other *E. coli* strains, including Top 10, DH10B, BL21-Gold(DE3), Rosetta(DE3), and BW25113, were tested with very similar results (data not shown); thus, DH5*α* and BL21(DE3) were selected as two representative *E. coli* strains moving forward. Violacein production was not detectable by eye on any plates incu-bated at 37°C. However, we observed subtle differences in production that grew more pronounced over time at lower temperatures. At or below 30°C, low levels of violacein production could be visualized on plates of BL21(DE3) but not DH5*α*. Extraction of liquid cultures and quantification by high-performance liquid chromatography (HPLC) confirmed that BL21(DE3) was capable of producing approximately twice as much

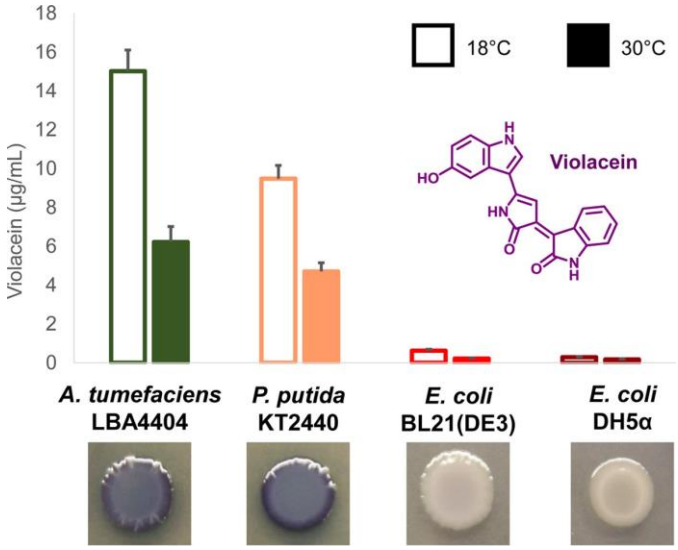


FIG 2 Heterologous expression of *vio2ta16* challenges the assumption that more closely related hosts perform better. Heterologous violacein production by four proteobacterial host strains was determined by extraction and HPLC quantification after 48 h of liquid culture in LB medium at 18 or 30°C. The data plotted are the mean ± standard error from two independent experiments (total $n = 6$). Plate images were taken after 48 h of growth at 30°C.

violacein as DH5*α* at 18°C (0.63 ± 0.07 µg/ml versus 0.28 ± 0.04 µg/ml), consistent with plate observations (Fig. 2). These observations indicated that even very minor changes in the host genotype at the strain level can result in fold changes in heterologous compound production.

Conversely, violacein production was readily detectable on plates of *P. putida* and *A. tumefaciens* transformed with pCAP05-*vio2ta16* and incubated at 30°C, the preferred growth temperature for these organisms. Extraction from *P. putida* liquid cultures and quantification confirmed the robust production of just under 10 µg of violacein/ml of culture after 48 h at 18°C (Fig. 2). Interestingly, the highest levels of production were obtained with *A. tumefaciens*, the least related host tested, which produced approximately 50% more violacein than *P. putida* at 18°C (Fig. 2). Production by all of the hosts was higher at 18°C than 30°C. Violacein production yield was not normalized to cell density, however, as the compound absorbs strongly at 600 nm and the *A. tume- faciens* heterologous host does not grow homogeneously in LB medium. Differences in growth or metabolism at different temperatures or in different host strains certainly affects heterologous production titers. Furthermore, the absolute plasmid copy number of pCAP05-*vio2ta16* in the hosts tested under the given cultivation conditions is not known either. Thus, our titer comparisons are only qualitative and not rigorously quantitative.

We also characterized violacein production levels from the native producer, *P. luteoviolacea* 2ta16, for qualitative comparison. Although production was relatively stable at 30°C (13.52 ± 1.36 µg/ml), it was highly inconsistent at 18°C (22.49 ± 11.77 µg/ml, Fig. S3). The reason for the spread at low temperature was not determined. Regardless, *P. luteoviolacea* 2ta16 is capable, at both growth temperatures, of producing more violacein than all of the heterologous hosts tested.

It was previously reported that the violacein BGC from *Pseudoalteromonas* sp. strain 520P1 could not be expressed from its native promoter in *E. coli* (28). To test whether promoter recognition was also responsible for low violacein production from *vio2ta16*, we replaced the native pathway promoter by *A* Red recombination-mediated PCR targeting (29), moving the biosynthetic operon into pET28a and placing it under the

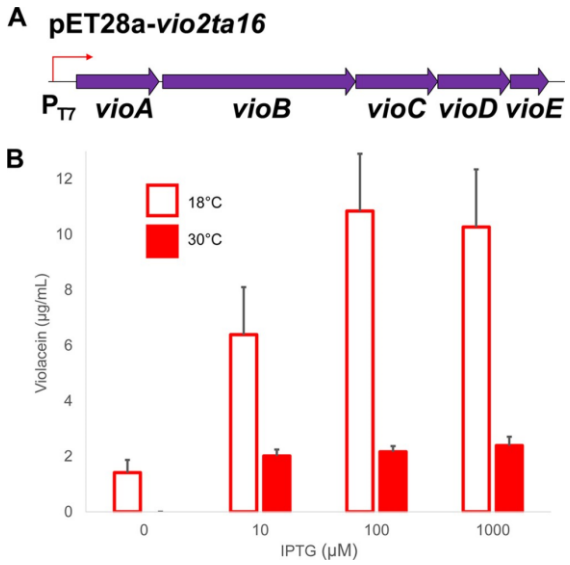


FIG 3 The *vio2ta16* native promoter is not efficiently activated in *E. coli*. (A) Schematic illustration of the pET28a-*vio2ta16* construct. The native *vio2ta16* promoter was replaced by transfer of the biosynthetic operon into expression vector pET28a, placing the pathway under the direct control of the T7 promoter. (A) Heterologous violacein production from *E. coli* BL21(DE3) harboring pET28a-*vio2ta16*. Cultures were grown to an OD_{600} of 0.5 to 0.6 at 37°C before induction with different concentrations of IPTG as shown. Following induction, liquid cultures were incubated for 24 h with shaking at 18 or 30°C before extraction and violacein quantification by HPLC. The data plotted are the mean ± standard error from two independent experiments (total $n = 6$).

direct control of the T7 promoter (Fig. 3A and S2C). Promoter replacement revealed that strain BL21(DE3) is indeed capable of supporting high levels of violacein production, up to approximately 11 μg/ml (Fig. 3B), strongly suggesting that the *vio2ta16* promoter is only active at low levels in *E. coli* hosts and that this is the primary reason for the low production level observed. Again, production levels were higher at 18°C than at 30°C, indicating that temperature plays an important role in heterologous violacein production and suggesting that the violacein biosynthetic enzymes are more stable or correctly folded at a lower temperature.

Regulation of the *vio2ta16* promoter. To gain a more detailed understanding of why the native *vio2ta16* promoter was well recognized in *P. putida* and *A. tumefaciens* but not in *E. coli*, we set out to identify specific regulatory elements involved in *vio2ta16* regulation. Although the violacein BGC from many *Betaproteobacteria* and *Gammapro- teobacteria* has been identified, a complete regulatory circuit controlling violacein expression has been fully characterized in only one bacterial species, *Chromobacterium violaceum*. In *C. violaceum*, violacein production is controlled by CviR, a LuxR-type quorum-sensing receptor that binds to and activates the transcription of the violacein promoter in the presence of its cognate autoinducer, an acyl-homoserine lactone (AHL) produced by CviI (30). Like the *luxI-luxR* quorum-sensing signal receptor pair from *Vibrio fischeri*, *cviI* and *cviR* are adjacent within the genome. It is notable that although the CviI-CviR pair has been confirmed to regulate violacein production in two strains of *C. violaceum* (ATCC 31532 and ATCC 12472), *C. violaceum* ATCC 31532 produces a C_6 homoserine lactone (C_6-HSL) and is inhibited by long-chain HSLs (C_{10} to C_{14}), while *C. violaceum* ATCC 12472 produces 3-hydroxy-C_{10}-HSL and is inhibited by short-chain HSLs (C_4 to C_8) (31, 32). Thus, the two circuits are functional opposites of one another.

We queried the genome of *P. luteoviolacea* 2ta16 with DELTA-BLAST (33) for CviR homologs that could be involved in violacein regulation. We identified seven candidates, which we named PLR1 to PLR7, four of which appear to be canonical LuxR-type

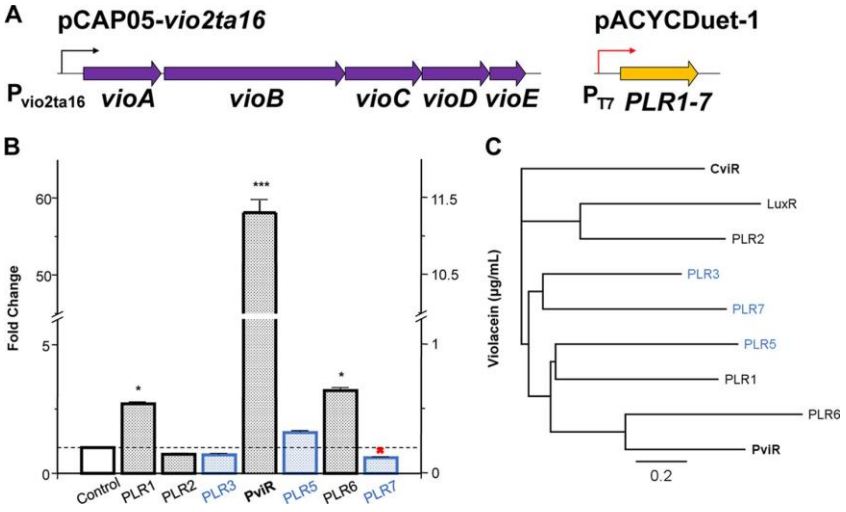


FIG 4 Coexpression of a LuxR-type receptor from *P. luteoviolacea* 2ta16 activates high levels of violacein production in *E. coli.* (A) Schematic illustrating coexpression of candidate quorum-sensing receptors from *P. lu- teoviolacea* with pCAP05-*vio2ta16* in *E. coli* BL21(DE3). (B) Quantification of violacein production in a coexpression experiment. Cultures were grown to an OD_{600} of 0.5 to 0.6, induced with 100 μM IPTG, and incubated at 18°C with shaking for 24 h before extraction and violacein quantification by HPLC. Candidate regulators are on the *x* axis and ordered by similarity to CviR as determined by DELTA-BLAST, and putative response regulators are blue. The data plotted are the mean ± standard error of total production on the right *y* axis and fold change compared to the control on the left *y* axis ($n = 3$). Statistical significance was determined with a one-tailed Student *t* test (*, $P < 0.01$; ***, $P < 0.0005$). PLR7 was not expressed solubly in *E. coli*, as indicated by the red X. (C) Neighbor-joining phylogenetic tree of quorum-sensing receptors LuxR and CviR and homologs identified from *P. luteoviolacea* 2ta16 generated with Geneious 5.1.7. An alignment was generated with a Blosum62 cost matrix, and the tree was built with a Jukes-Cantor genetic distance model in Geneious Tree Builder.

receptors and three of which appear to be response regulators in a two-component regulatory system, as evidenced by the presence of N-terminal phosphorylation sites (Fig. S4). Conserved domains for AHL autoinducer binding were not identified in any of the seven candidate receptors by the NCBI conserved-domain identifier (34). Of six amino acid residues important for AHL binding (35, 36), three or fewer residues are conserved in any single candidate receptor identified (Fig. S5). Furthermore, *P. luteovi- olacea* 2ta16 does not possess any homologs of the AHL autoinducer synthase CviI or LuxI.

Previous studies have shown that quorum-sensing signal receptors are not soluble or active when expressed in the absence of stabilizing autoinducer ligands (30, 37). However, cloning and expression of the candidate regulators revealed that six out of seven could be expressed as soluble proteins in *E. coli* BL21(DE3) without the addition of any *P. luteoviolacea* extracts or purified AHL autoinducers (Fig. S6A). When the candidate regulators were coexpressed with pCAP05-*vio2ta16*, three receptors induced a statistically significant increase in violacein production in *E. coli* and one receptor, which we have named PviR, increased production nearly 60-fold compared to the control, resulting in production levels similar to those achieved by T7 promoter refactoring (11.12 ± 0.17 μg/ml, Fig. 4B). PviR does not cluster near *vio2ta16* or any identifiable autoinducer synthase genes, nor is it the most similar CviR homolog identified from *P. luteoviolacea* 2ta16 (Fig. 4C).

We verified that PviR controls violacein production at the transcriptional level by quantitative PCR (qPCR). Four hours after the induction of PviR in *E. coli* BL21(DE3), *vio2ta16* transcription increased 16-fold (Fig. 5A). After 24 h, we measured a 61-fold increase in *vioA* transcript levels, suggesting that the transcripts accumulated over time (Fig. 5A). This activity was not dependent on the addition of any exogenous culture extracts or compounds. PviR aligns poorly with CviR along the N terminus of

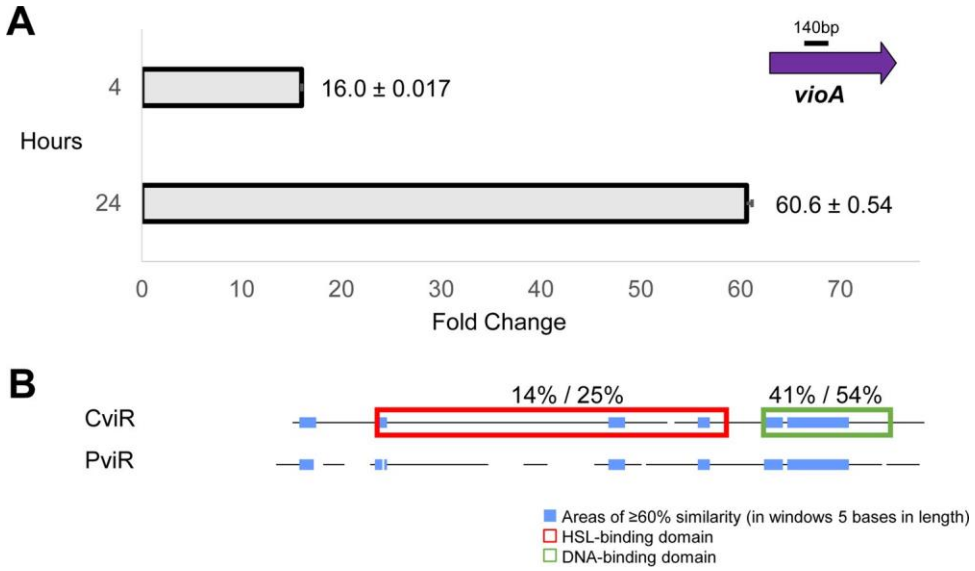


FIG 5 LuxR-type receptor PviR activates *vio2ta16* transcription independently of HSL autoinducers. (A) Fold changes in *vioA* transcript levels determined by qPCR. Changes were measured by the relative standard curve method with the *E. coli* gene *cysG* as the calibrator. Control and PviR coexpression cultures were grown as described previously (in the legend to Fig. 4) before total RNA isolation, DNase digestion, and cDNA generation with random hexamers for qPCR analysis. (B) Alignment of CviR and PviR sequences generated with Clone Manager. The values above the boxed domains are the amino acid identity/similarity percentages within those regions.

the protein (where the AHL-binding domain is located) and is similar primarily within the C-terminal DNA-binding domain (Fig. 5B). Only one of six amino acid residues important for AHL binding is conserved (Fig. S5). These results indicate that PviR activation of *vio2ta16* transcription does not depend on an AHL autoinducer. Previ- ously, it was reported that the LuxR-type transcription factor MalR from *Burkholderia thailandensis* activates a malleilactone promoter construct in *E. coli* independently of AHLs; however, MalR shows complete amino acid identity with AHL-binding residues (38). If PviR does require an autoinducer for transcriptional activation of the *vio2ta16* promoter, as is typical for quorum-sensing receptors, the molecule or a functional mimic is supplemented in LB medium or produced by *E. coli*. Recently, Bassler and coworkers showed that a ubiquitous autoinducer produced by *E. coli* can interact with a *Vibrio cholerae* quorum-sensing receptor to control the expression of genes involved in biofilm formation and toxin production (39).

Unlike PLR6 and PLR1, which also induced statistically significant, albeit smaller, increases in heterologous violacein production, PviR is conserved among all of the sequenced *Pseudoalteromonas* strains that possess the violacein BGC. As shown in Fig. 6, which lists 22 *Pseudoalteromonas* strains, including 16 strains of *P. luteoviolacea*, PLR6 is not found in 2 strains, while PLR1 is absent from 9. Interestingly, PviR is conserved at 91% amino acid identity in *P. luteoviolacea* ATCC 29581, even though the violacein biosynthetic genes have dropped to 54 to 77% identity. PviR appears to be specific to the genus

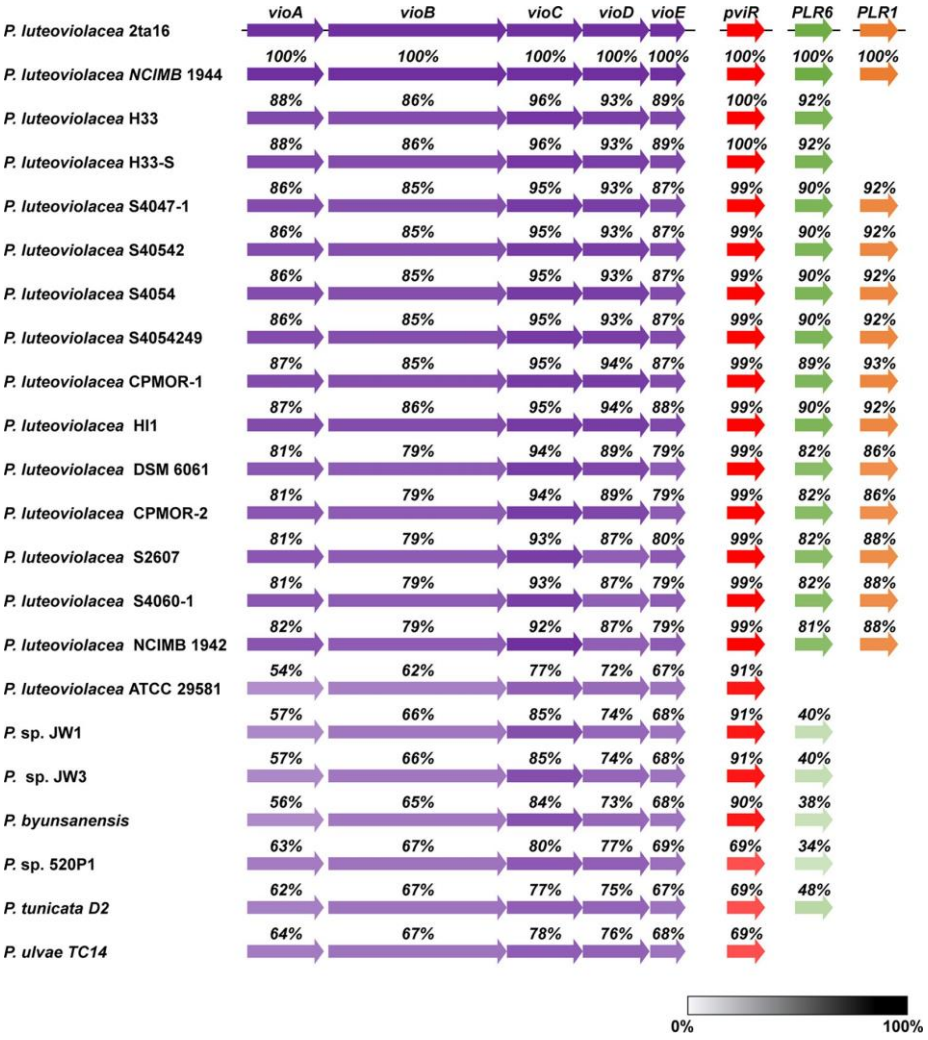


FIG 6 PviR is conserved across sequenced *Pseudoalteromonas* strains that possess the violacein BGC. Comparison of LuxR-type receptor conservation across all of the sequenced *Pseudoalteromonas* strains with the violacein BGC (*vioA* to *vioE*). All of the *Pseudoalteromonas vio* BGCs in the JGI and NCBI databases were identified by a protein BLAST search for VioC from *P. luteoviolacea* 2ta16 and referenced against BLAST hits for PviR, PLR6, and PLR1. Amino acid sequence identity percentages are listed above the arrows representing the genes and also indicated by color opacity.

activation during heterologous expression. Thus, the availability or activity of compatible regulators could explain, in part, the observed differences in heterologous violacein production. Using PviR as a probe, we queried the genomes of the four host strains by DELTA-BLAST. Surprisingly, the most similar candidate identified was the regulatory protein CsgD from *E. coli*. Alignment of the sequences of PviR and CsgD revealed approximately 30% amino acid identity spanning 63% of the protein sequence, with an expect (E) value 6 to 9 orders of magnitude lower than that of the top candidates identified in *A. tumefaciens* and *P. putida* (Fig. 7A to C and S7). This result is perhaps reflective of the closer phylogenetic relationship between *P. luteoviolacea* and *E. coli* than the other two hosts.

A total of 11 PviR homologs were identified in *E. coli* (Fig. S7A). Although the genomes of *E. coli* DH5*α* and BL21(DE3) are very similar, seven of the BLAST hits are not

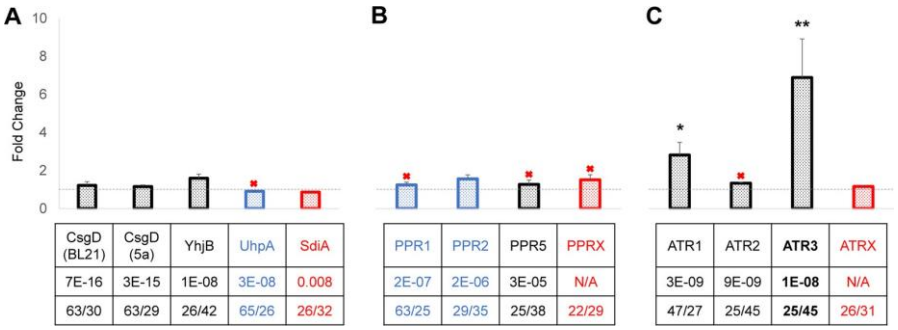


FIG 7 Host regulatory factors influence heterologous violacein production. Heterologous violacein production by *E. coli* BL21(DE3) through overexpression of PviR homologs from *E. coli* (A), *P. putida* KT2440 (B), and *A. tumefaciens* LBA4404 (C). Each table column lists the protein name, E value, and percent coverage/identity from top to bottom, respectively. Response regulators are blue. Regulatory proteins with E values below the significance threshold are red. Proteins that were not expressed solubly (see Fig. S6) are indicated by a red X. The data plotted are the mean ± standard error of the fold change in violacein production relative to the control in two independent experiments (total $n = 5$). Statistical significance was determined with a one-tailed Student *t* test (*, $P < 0.05$; **, $P < 0.005$).

identical proteins in the two strains, including CsgD, which contains two point mutations in BL21(DE3) compared to DH5*α* (S19P, A154V). A recent study revealed that metabolic physiology and gene expression can vary widely among different strains of *E. coli*, including BL21(DE3) and DH5*α* (40). Furthermore, the study showed that expression of *csgD* specifically is low in both strains under aerobic and anaerobic culture conditions. Thus, we hypothesized that overexpression of *csgD* or other PviR homologs in *E. coli* might rescue heterologous violacein production to levels that rival those in *P. putida* or *A. tumefaciens*.

We selected the top three PviR homologs from *E. coli* and tested their ability to increase heterologous violacein production when overexpressed in *E. coli* BL21(DE3). We also selected a LuxR homolog that was not a PviR BLAST hit as a control. In total, five *E. coli* regulatory genes were tested, *csgD* from BL21(DE3) and DH5*α*; *yhjB* and *uhpA*, which are identical proteins in BL21(DE3) and DH5*α*; and the control, *sdiA*, which is present in DH5*α* but absent from BL21(DE3). Despite significant sequence similarity to PviR, neither CsgD variant was able to raise heterologous violacein production above the baseline. Although YhjB improved violacein production approximately 1.5-fold, this result was not statistically significant, as determined with a one-tailed Student *t* test ($P < 0.05$) (Fig. 7A). According to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of proteins isolated at the time of extraction, UhpA was not overexpressed solubly under the conditions tested (Fig. S6B). Thus, we conclude that while *E. coli* regulatory proteins CsgD and YhjB are similar in sequence to PviR, they are not capable of activating the *vio2ta16* promoter. Although we cannot rule out the possibility that regulatory proteins not tested or identified are capable of activating *vio2ta16* expression, our results suggest that, despite their close phylogenetic relationship to *P. luteoviolacea*, laboratory *E. coli* strains are not able to express *vio2ta16* from the native pathway promoter.

As overexpression of the two CsgD variants from BL21(DE3) and DH5*α* produced very similar results, we reasoned that higher violacein titers from BL21(DE3) than from DH5*α* are the result of enhanced protein stability and not an increase in *vio2ta16* transcription. Differences in violacein expression could have arisen because of disparate activity or expression levels of transcription factors in the two strains; however, this does not seem to be the case. Furthermore, enhanced protein stability, which is a hallmark of BL21(DE3), is consistent with the finding that the violacein titers of BL21(DE3) were similar to those of DH5*α* at 30°C but higher at 18°C (Fig. 1).

Using the same methodology, we also identified and tested regulatory genes from *P. putida* KT2440 and *A. tumefaciens* LBA4404. Because violacein is readily expressed in

these hosts, we hypothesized that we could identify PviR homologs capable of *vio2ta16* activation from these organisms. We identified 15 homologs from *P. putida* KT2440, 11 of which (including 4 of the top 5) are two-component system response regulators (Fig. S7B). The two most similar BLAST hits, PPR1 and PPR2, along with the most similar LuxR-type receptor, PPR5, were tested for the ability to increase violacein expression in *E. coli*. Additionally, a LuxR homolog, PPRX, which was not a BLAST hit for PviR, was selected as a control. However, only PPR2 was expressed solubly in *E. coli* BL21(DE3) (Fig. S6C) and did not induce a statistically significant increase in violacein production relative to the baseline (Fig. 7B). Addition of freeze-dried *P. putida* KT2440 culture supernatants to a concentration of 1 mg/ml did not alter the ability of any of the regulatory proteins tested to enhance violacein production (Fig. S9A). Thus, we were unable to conclusively identify transcriptional regulators from *P. putida* KT2440 that can activate *vio2ta16* expression.

However, testing of PviR homologs from *A. tumefaciens* LBA4404, also identified by DELTA-BLAST (Fig. S7C), revealed two regulators, ATR1 and ATR3, that increased violacein production approximately 3- and 7-fold, respectively (Fig. 7C), resulting in visibly purple cultures (data not shown). In particular, ATR3 raised violacein production to more than 2.5 μg/ml, which is far greater than the levels achieved by coexpression of the *P. luteoviolacea* LuxR-type receptor PLR1 or PLR6 (Fig. 4B). Although ATR3 is not able to raise *E. coli* violacein production levels to those of the *A. tumefaciens* heterologous host, we would not necessarily expect it to do so, as coexpression of PviR does not raise *E. coli* production to levels that rival those of *P. luteoviolacea*. Furthermore, ATR3 was only overexpressed at relatively low levels according to SDS-PAGE analysis (Fig. S6D). Confirmation of this LuxR-type receptor as a regulator of *vio2ta16* is consistent with the ability of the *A. tumefaciens* heterologous host to support robust production of violacein from the native pathway promoter. Unlike PviR, ATR3 does possess a conserved domain for AHL autoinducer binding, and four out of six AHL-interacting amino acid residues have been preserved (Fig. S8), although ATR3 activity and solubility were not dependent on the addition of exogenous AHLs. *A. tumefaciens* possesses a characterized quorum-sensing regulatory pair involving an AHL autoinducer and LuxR-type receptor TraR (36), but DELTA-BLAST searches with PviR as a probe did not detect TraR (Fig. S7C). Addition of freeze-dried *A. tumefaciens* LBA4404 culture supernatants to a concentration of 1 mg/ml did not alter the ability of any of the regulatory proteins tested to enhance violacein production in a statistically significant way (Fig. S9B). Because ATR3 is not the most similar PviR homolog identified, we conclude that it is difficult to predict the exact identity of compatible regulatory elements on the basis of sequence comparison alone. Furthermore, it has not escaped our attention that regulators that were not tested or identified from the hosts could also possibly influence heterologous *vio2ta16* expression.

Conclusion. In this study, we validate the extension of our TAR cloning and heterologous expression platform to include Gram-negative proteobacterial expression hosts. Furthermore, our findings suggest that a more closely related host may not always be preferable for the heterologous expression of microbial BGCs. Heterologous antibiotic production remains challenging, particularly for pathways that utilize nonclustered and uncharacterized regulatory, biosynthetic, or resistance factors. For example, it was only recently discovered that glutamyl-tRNA^{Glu} is required for lantibiotic dehydration reactions during the biosynthesis of the antibiotic nisin (41); consequently, heterologous nisin production would require that host glutamyl-tRNAs be compatible with nisin biosynthetic machinery. Additionally, the antibiotic salinamide A inhibits bacterial RNA polymerase (RNAP) (42) and therefore must be expressed in hosts with resistant RNAP subunits. The phylogenetic relationship indicates a shared evolutionary history, and thus, it may theoretically provide a useful indication of host compatibility for the heterologous expression of BGCs in some cases. However, our study suggests that similarity is not directly synonymous with compatibility, either for an entire host organism or for an individual protein component. When working with challenging BGCs, broad-host-range

heterologous expression represents a viable approach used to maximize the likelihood of achieving successful heterologous production.

Many characterized microbial BGCs are regulated by LuxR-type transcription factors, some of which are encoded by cluster-situated regulatory genes. Previous studies have shown that transcriptionally silent BGCs can be activated by the overexpres- sion of genes encoding LuxR-type regulators (43, 44). Attempts to do so in a heterologous expression system would require the cognate autoinducer or binding ligand to be present and available in the host system. We identified a nonclustered LuxR-type regulator, PviR, that is capable of dramatic upregulation of *vio2ta16* expression independently of AHL autoinducers and is conserved among all of the sequenced *Pseudoalteromonas* host strains that harbor the *vio* BGC. Furthermore, we identified two PviR homologs from *A. tumefaciens* that are capable of robust activation of violacein production, consistent with the ability of the *A. tumefaciens* heterologous host to readily produce large amounts of violacein from the native pathway promoter. The cognate autoinducers, if they exist, for PviR and the
A. tumefaciens homologs are unknown and may be produced by *E. coli*, which is not unprecedented (39).

Finally, although replacement of the *vio2ta16* promoter was an easy way to overcome host regulatory incompatibility with this pathway, most BGCs are not so easily refactored. The operonic architecture of BGCs can be complex and is usually only inferred on the basis of gene organization. Some BGCs may possess long operons or even single genes that cannot be reliably expressed from the T7 promoter (10, 45). Broad-host-range expression creates the possibility of achieving heterologous antibiotic production in the absence of pathway refactoring. Although only alphaproteobacterial and gammaproteobacterial hosts were tested in this study, earlier work on the RK2 replicon suggests that it is functional in a wide range of Gram-negative genera, although it is not clear whether bacterial phyla such as *Cyanobacteria*, *Spirochetes*, or *Bacteroidetes* have been systematically assessed (46). Furthermore, the maintenance of very large and repetitive multimodular BGCs presents a formidable challenge for heterologous expression in any system and must be evaluated in future work. We believe that the characterization and utilization of additional host organisms will aid the investigation of cryptic BGCs and could lead to unanticipated findings related to natural-product biosynthesis, regulation, or bioactivity.

MATERIALS AND METHODS

General methods and materials. The primers and plasmids used in this study are reported in Table S1 in the supplemental material. Highly transformable *Saccharomyces cerevisiae* strain VL6-48N (*MATa his3-D200 trp1-Δ1 ura3-Δ1 lys2 ade2-101 met14 psi⁺ cirO*) was used as a host for TAR cloning. Yeast cells were grown in liquid YPD (yeast extract-peptone-dextrose) medium (2% D-glucose, 1% yeast extract, 2% peptone) supplemented with 100 mg/liter adenine and used for spheroplast preparation prior to TAR. Yeast transformants were selected for on synthetic histidine dropout (SD-His) agar containing 5-FOA (SD-Trp--5-FOA agar) consisting of 0.17% yeast nitrogen base without amino acids and ammonium sulfate (Sigma), 0.19% yeast synthetic dropout medium supplements without histidine (Sigma), 1 M sorbitol, 2% D-glucose, 0.5% ammonium sulfate, 100 mg/liter adenine, 2% agar, and 0.1% 5-FOA (Zymo Research). *E. coli* strains, *P. putida* KT2440, and *A. tumefaciens* LBA4404 were cultivated in LB medium (components purchased from BD Biosciences or Fisher Scientific) supplemented with the appropriate antibiotics. YM medium (yeast extract at 0.04%, mannitol at 1.0%, NaCl at 1.7 mM, $MgSO_4 \cdot 7H_2O$ at
0.8 mM, $K_2HPO_4 \cdot 3H_2O$ at 2.2 mM) was used for transformation and preculture of *A. tumefaciens* LBA4404. *P. luteoviolacea* 2ta16 was grown in Marine Broth 2216 (BD Biosciences). For plates, 1.5% agar (Fisher Scientific) was added to the appropriate medium. DNA isolation and manipulations were carried out in accordance with standard protocols. DNA fragments larger than 3 kb were amplified with PrimeSTAR Max (Clontech Laboratories, Inc.); all other PCR products were amplified with PrimeSTAR HS DNA polymerase (Clontech Laboratories, Inc.).

Construction and validation of pCAP05. To capture all of the broad-host-range elements from pRK442(H) (20), fragments of 5,121 and 3,943 bp were amplified with primer pairs pCAP05-1_5121F/R and pCAP05-2_3943F/R (Table S1; Fig. S1A), respectively. In parallel, a 2,024-bp fragment containing *CEN6_ARS4* and *HIS3* and a 1,639-bp fragment containing *pADH* and *URA3* were amplified from pARS-VN
(47) with primer pairs pCAP05-3_2024F/R and pCAP05-4_1639F/R (Table S1; Fig. S1A), respectively. The four PCR-amplified fragments were combined and assembled into a single construct with Gibson Assembly (New England Biolabs) to generate pCAP05. The circular construct was digested with NdeI for verification and transferred to *S. cerevisiae* VL6-48N and selected against or for on histidine-deficient agar

plates with or without 5-FOA (Fig. S1B to D). pCAP05 was transferred to *P. putida* KT2440 and *A. tumefaciens* LBA4404 by electroporation at 2,500 V with a 2-mm cuvette. To prepare electrocompetent *P. putida*, cells grown at 30°C to an optical density at 600 nm (OD_{600}) of 1.0 were washed twice with an ice-cold 10% glycerol solution and resuspended in a small volume of the glycerol solution before the addition of up to 1 μg of DNA. Separately, electrocompetent *A. tumefaciens* was prepared by washing cells grown at 30°C in YM medium to an OD_{600} of 0.5 with ice-cold 1 mM HEPES buffer (pH 7.5) twice before resuspension in 1 mM HEPES--10% glycerol and the addition of 100 μg of DNA. *P. putida*/pCAP05 was recovered for 2 h at 30°C in LB medium before selection on LB agar supplemented with 15 μg/ml tetracycline (Fig. S1E). *A. tumefaciens*/pCAP05 was recovered for 3 h in YM medium before selection on YM agar supplemented with 5 μg/ml tetracycline (Fig. S1F).

TAR cloning of *vio2ta16*. The 9-kb DNA region containing *vio2ta16* was PCR amplified from the genomic DNA of *P. luteoviolacea* 2ta16 in three fragments (each approximately 3 kb) with primer pairs vio2ta16-1_3121-40F/vio2ta16-1_3121-134R, vio2ta16-2_3058-134F/vio2ta16-2_3058-108R, and vio2ta16-3b_3027-482F/vio2ta16-3b_3027-40R (Table S1; Fig. S2A). These fragments were cloned into pCAP05 in *S. cerevisiae* VL6-48N by a previously reported protocol (15), with minor modifications.

S. cerevisiae VL6-48N was grown to an OD_{600} of 0.7 to 1.0 in 50 ml of YPD medium supplemented with adenine (100 mg/liter) at 30°C with shaking. The cells were harvested and washed with ice-cold water and osmotically stabilized in 1 M sorbitol at 4°C overnight prior to spheroplast preparation. Preparation of spheroplast cells was carried out with a lytic enzyme (Zymolyase 20T; MP Biomedicals) at a final concentration of 0.1 mg/ml with 30 to 40 min of incubation. A 0.5-μg sample of each PCR product and ClaI/XhoI-linearized pCAP05 were added to spheroplast cells, and the transformation was mediated by PEG 8000 (Sigma). The transformed spheroplasts were mixed with 10 ml of SD-His top agar (containing 3% agar) at 55°C and overlaid on SD-His--5-FOA agar. The plates were incubated at 30°C for 3 days. Hundreds of colonies appeared per plate, and four were picked, transferred onto new SD-His agar plates, and incubated for 24 h at 30°C. Cells were lysed with Zymolyase 20T at 37°C for 30 min and subsequently boiled at 98°C for 5 min. The captured *vio2ta16* BGC was screened with primer pair vio2ta16check_402F/R (Table S1). Plasmids were extracted from PCR-positive clones and transferred into *E. coli* Top 10 cells by electroporation. The plasmids were purified from tetracycline- resistant *E. coli* clones, and the resulting constructs were confirmed by restriction analysis; the vector containing *vio2ta16* BGC was designated pCAP05-*vio2ta16* (Fig. S2B).

Heterologous expression and quantification of violacein production. pCAP05-*vio2ta16* was introduced into *E. coli*, *P. putida*, and *A. tumefaciens* by electroporation as described above. Precultures were grown overnight at 30°C in LB (*E. coli* and *P. putida*) or YM (*A. tumefaciens*) medium supplemented with 5 μg/ml tetracycline. YM was used for *A. tumefaciens* precultures to prevent cell clumping. All of the cultures were seeded the next day with a standard inoculum of each strain to give a starting OD_{600} of 0.02 in LB with 5 μg/ml tetracycline. One-milliliter cultures were then incubated for 48 h with shaking at 220 rpm at 18 or 30°C before cells were harvested by centrifugation and extracted with methanol. For HPLC analysis, samples were injected onto a Phenomenex Kinetex XB-C18 reversed-phase HPLC column (2.6 μm, 150 by 4.6 mm [inside diameter]) and analyzed with an Agilent 1260 liquid chromatography system by gradient elution (A, CH_3CN with 0.1% trifluoroacetic acid [TFA]; B, H_2O with 0.1% TFA; 50 to 100% A over 5 min and 100% A from 5 to 8 min at 0.7 ml/min) at 575 nm. Violacein was quantified with a standard purchased from AdipoGen Life Sciences (San Diego, CA) with the formula $y = 3238.1x - 33.266$, where x is the concentration and y is the integrated UV absorbance at 575 nm ($R^2 = 0.9988$). Data were collected from two independent experiments, each time from triplicate cultures.

Wild-type *P. luteoviolacea* strain 2ta16 was grown in MB medium without any antibiotics. Violacein production was quantified as described above, and data were collected from three independent experiments, each time in triplicate.

***vio2ta16* promoter replacement and testing.** pET28a was amplified with primer pair pET28a-vioE-F/pET28a-vioA-R (Table S1), and the product was purified and digested with DpnI. The pCAP05 backbone was replaced in *E. coli* BW25113/pIJ790/pCAP05-*vio2ta16* with the PCR targeting system (29). The resulting plasmid, pET28a-*vio2ta16*, was confirmed by restriction analysis (Fig. S2C) and sequencing (Genewiz, San Diego, CA) and transformed into *E. coli* BL21(DE3). A fresh colony was picked and inoculated into LB with 50 μg/ml kanamycin and grown overnight at 37°C. Precultures were reinoculated into fresh LB with the same antibiotic and grown at 37°C to an OD_{600} of 0.5 to 0.6. Isopropyl-*β*-D- thiogalactopyranoside (IPTG) was added to a 0, 10, 100, or 1,000 μM final concentration before incubation at 18 or 30°C with shaking (220 rpm) for 24 h. Violacein production in 1-ml cultures was quantified by HPLC as described above. Two independent experiments were conducted, each time with triplicate cultures grown under each condition.

Cloning and testing of regulatory genes from *P. luteoviolacea* 2ta16. PLR1 to PLR7 were amplified with the primer pairs listed in Table S1. PCR products and pACYCDuet-1 were digested with the restriction enzyme pair NcoI/HindIII or NdeI/XhoI (as indicated in Table S1) and ligated with T4 DNA ligase (New England Biolabs). The resulting constructs were screened by restriction digestion, verified by sequencing, and transferred to BL21(DE3)/pCAP05*vio2ta16*. BL21(DE3)/pCAP05-*vio2ta16* with the empty vector was used as a control. Single clones were picked for precultures grown overnight at 37°C in LB medium supplemented with 15 μg/ml tetracycline and 34 μg/ml chloramphenicol. One hundred microliters of each preculture was transferred to 10 ml of fresh LB with the same antibiotics and grown at 37°C to an OD_{600} of 0.5 to 0.6. Cultures were induced with 100 μM IPTG and incubated for an additional 24 h at 18°C with shaking (220 rpm). Violacein production was again quantified by HPLC as described above. Protein expression was checked by SDS-PAGE.

Cloning and testing of regulatory genes from *E. coli*, *P. putida*, and *A. tumefaciens*. PviR homologs from *E. coli*, *P. putida*, and *A. tumefaciens* were cloned and tested as described above. Violacein production was again quantified by HPLC, and protein expression was checked by SDS-PAGE in two independent experiments. To determine whether small molecules secreted by *P. putida* and *A. tumefaciens* influence the ability of tested regulatory genes from the two hosts to enhance violacein production, 50 ml of culture supernatants from *P. putida* and *A. tumefaciens* grown to an OD_{600} of approximately 1.0 in LB were freeze-dried, resuspended in dimethyl sulfoxide (DMSO), and added to *E. coli* BL21(DE3) expression cultures to a final concentration of 1 mg/ml at the time of IPTG induction. Freeze-dried LB was added as a control, and violacein production by all of the cultures was quantified by HPLC.

qPCR analysis. Single clones of BL21(DE3)/pCAP05-*vio2ta16* with pACYCDuet-1-*pviR* and empty pACYCDuet-1 were picked and grown as described above for testing of violacein induction activity. Cells were harvested 4 and 24 h after IPTG induction for RNA isolation. Cells were stabilized with RNAprotect Cell Reagent (Qiagen). RNA was isolated with an RNeasy minikit (Qiagen), and cell lysis was obtained by homogenization with the FastPrep system (BIO 101 Inc.) (five cycles of 30 s at 5.5 speed). Turbo DNase (Thermo Fisher Scientific) treatment was carried out for 2 h in accordance with the manufacturer's instructions. The RNA was checked on an agarose gel, and the absence of genomic DNA in the sample was confirmed by PCR with primers for the housekeeping gene *cysG* (48). cDNA synthesis was carried out with SuperScript IV reverse transcriptase (Thermo Fisher Scientific), 50-pg/μl random hexamers, and 500 ng of RNA in accordance with the manufacturer's protocol. The relative standard curve method was used for relative quantification of *vioA* transcripts during PviR coexpression compared to the control. BL21(DE3)/pCAP05*vio2ta16*/pACYCDuet-1-*pviR* genomic DNA was used to generate 8-fold dilution series standard curves with the primers for *vioA* (qRT-vioA_140F/R) and the calibrator *cysG* (qRT-cysG_126F/R) listed in Table S1. The standard curves were run in triplicate, and each cDNA reaction was run in parallel five times, along with reverse transcriptase-negative cDNA and no-template control reactions. The qPCR mixtures contained each primer pair at 0.5 μM, 1 μl of 10-fold-diluted cDNA, and 1X SYBR green PCR master mix (Thermo Fisher Scientific) for a total volume of 10 μl. A Stratagene Mx3000p qPCR thermocycler was used with a cycling protocol consisting of 1 cycle of 50°C for 2 min; 1 cycle of 95°C for 2 min; 40 cycles of 95°C for 15 s and 60°C for 1 min; and 1 cycle of at 95°C for 1 min, 55°C for 30 s, and 95°C for 30 s to generate a melting curve.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01291-17.

FIG S1, PDF file, 0.3 MB.
FIG S2, PDF file, 0.3 MB.
FIG S3, PDF file, 0.1 MB.
FIG S4, PDF file, 0.2 MB.
FIG S5, PDF file, 0.1 MB.
FIG S6, PDF file, 0.3 MB.
FIG S7, PDF file, 0.6 MB.
FIG S8, PDF file, 0.1 MB.
FIG S9, PDF file, 0.1 MB.
TABLE S1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank V. Larionov (National Cancer Institute, NIH) for providing plasmid pARS-VN and D. K. Hanson (Argonne National Laboratory) for providing plasmid pRK442(H). We also thank P. Jordan (Scripps Institution of Oceanography, UCSD) for reviewing the manuscript and A. Orebic for assistance with experiments during his summer internship in the Research Scholars Program at UCSD.

This work was supported by a graduate fellowship from the National Science Foundation to J.J.Z. and United States. National Institutes of Health grants F31- AI129299 to J.J.Z. and R01-GM085770 and R01-AI117712 to B.S.M.

REFERENCES

1. Udwary DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, Fenical W, Jensen PR, Moore BS. 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. Proc Natl Acad Sci U S A 104:10376–10381. https://doi.org/10.1073/pnas.0700962104.
2. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412–421. https://doi.org/10.1016/j.cell.2014.06.034.
3. Ren H, Wang B, Zhao H. 2017. Breaking the silence: new strategies for discovering novel natural products. Curr Opin Biotechnol 48:21–27. https://doi.org/10.1016/j.copbio.2017.02.008.
4. Kouprina N, Larionov V. 2016. Transformation-associated recombination

(TAR) cloning for genomics studies and synthetic biology. Chromosoma 125:621– 632. https://doi.org/10.1007/s00412-016-0588-3.
1. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Müller R, Stewart AF, Zhang Y. 2012. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospect- ing. Nat Biotechnol 30:440 – 446. https://doi.org/10.1038/nbt.2183.
2. Rutledge PJ, Challis GL. 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol 13: 509 –523. https://doi.org/10.1038/nrmicro3496.
3. Gomez-Escribano JP, Bibb MJ. 2011. Engineering *Streptomyces coelicolor* for the heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4:207–215. https://doi.org/10.1111/j.1751-7915.2010 .00219.x.
4. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. 2010. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A 107:2646 –2651. https://doi.org/10.1073/pnas.0914833107.
5. Masschelein J, Jenner M, Challis GL. 2017. Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosyn- thetic highlights. Nat Prod Rep 34:712–783. https://doi.org/10.1039/c7np00010c.
6. Stevens DC, Hari TP, Boddy CN. 2013. The role of transcription in heterologous expression of polyketides in bacterial hosts. Nat Prod Rep 30:1391–1411. https://doi.org/10.1039/c3np70060g.
7. Chai Y, Shan S, Weissman KJ, Hu S, Zhang Y, Müller R. 2012. Heterolo- gous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagene- sis. Chem Biol 19:361–371. https://doi.org/10.1016/j.chembiol.2012.01 .007.
8. Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF. 2010. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers 93:833– 844. https://doi.org/10.1002/bip.21450.
9. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci U S A 111:1957–1962. https://doi.org/10.1073/pnas .1319584111.
10. Shao Z, Zhao H, Zhao H. 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37:e16. https://doi.org/10.1093/nar/gkn991.
11. Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS. 2015. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem Biol 10:2841–2849. https://doi.org/10.1021/acschembio .5b00658.
12. Li Y, Li Z, Yamanaka K, Xu Y, Zhang W, Vlamakis H, Kolter R, Moore BS, Qian PY. 2015. Directed natural product biosynthesis gene cluster cap- ture and expression in the model bacterium *Bacillus subtilis*. Sci Rep 5:9383. https://doi.org/10.1038/srep09383.
13. Jordan PA, Moore BS. 2016. Biosynthetic pathway connects cryptic ribo-somally synthesized posttranslationally modified peptide genes with pyr-roloquinoline alkaloids. Cell Chem Biol 23:1504–1514. https://doi.org/10 .1016/j.chembiol.2016.10.009.
14. Ross AC, Gulland LE, Dorrestein PC, Moore BS. 2015. Targeted capture and heterologous expression of the *Pseudoalteromonas* alterochromide gene cluster in *Escherichia coli* represents a promising natural product exploratory platform. ACS Synth Biol 4:414 – 420. https://doi.org/10 .1021/sb500280q.
15. Agarwal V, El Gamal AA, Yamanaka K, Poth D, Kersten RD, Schorn M, Allen EE, Moore BS. 2014. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. Nat Chem Biol 10:640 – 647. https://doi.org/10.1038/nchembio.1564.
16. Scott HN, Laible PD, Hanson DK. 2003. Sequences of versatile broad- host-range vectors of the RK2 family. Plasmid 50:74 –79. https://doi.org/10.1016/S0147-619X(03)00030-1.
17. Greener A, Lehman SM, Helinski DR. 1992. Promoters of the broad host range plasmid RK2: analysis of transcription (initiation) in five species of Gram-negative bacteria. Genetics 130:27–36.
18. Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF. 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse *Proteo- bacteria*. Appl Environ Microbiol 76:1633–1641. https://doi.org/10.1128/AEM.02169-09.
19. Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L, Völker U, Ellingsen TE, Valla S. 2009. A plasmid RK2-based broad-host-range clon- ing vector useful for transfer of metagenomic libraries to a variety of bacterial species. FEMS Microbiol Lett 296:149 –158. https://doi.org/10 .1111/j.1574-6968.2009.01639.x.
20. Loeschcke A, Thies S. 2015. *Pseudomonas putida*---a versatile host for the production of natural products. Appl Microbiol Biotechnol 99: 6197–6214. https://doi.org/10.1007/s00253-015-6745-4.
21. Gross F, Gottschalk D, Müller R. 2005. Posttranslational modification of myxobacterial carrier protein domains in *Pseudomonas* sp. by an intrinsic phosphopantetheinyl transferase. Appl Microbiol Biotechnol 68:66 –74. https://doi.org/10.1007/s00253-004-1836-7.
22. Henkel CV, den Dulk-Ras A, Zhang X, Hooykaas PJ. 2014. Genome sequence of the octopine-type *Agrobacterium tumefaciens* strain Ach5. Genome An- nounc 2:e00225-14. https://doi.org/10.1128/genomeA.00225-14.
23. Williams KP, Gillespie JJ, Sobral BW, Nordberg EK, Snyder EE, Shallom JM, Dickerman AW. 2010. Phylogeny of gammaproteobacteria. J Bacteriol 192:2305–2314. https://doi.org/10.1128/JB.01480-09.
24. Zhang X, Enomoto K. 2011. Characterization of a gene cluster and its putative promoter region for violacein biosynthesis in *Pseudoalteromo- nas* sp. 520P1. Appl Microbiol Biotechnol 90:1963–1971. https://doi.org/10.1007/s00253-011-3203-9.
25. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640 – 6645. https://doi.org/10.1073/pnas.120163297.
26. Swem LR, Swem DL, O'Loughlin CT, Gatmaitan R, Zhao B, Ulrich SM, Bassler BL. 2009. A quorum-sensing antagonist targets both membrane- bound and cytoplasmic receptors and controls bacterial pathogenicity. Mol Cell 35:143–153. https://doi.org/10.1016/j.molcel.2009.05.029.
27. McClean KH, Winson MK, Fish L, Taylor A, Chhabra SR, Camara M, Daykin M, Lamb JH, Swift S, Bycroft BW, Stewart GS, Williams P. 1997. Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein pro-duction and inhibition for the detection of N-acylhomoserine lactones. Microbiology 143:3703–3711. https://doi.org/10.1099/00221287-143-12 -3703.
28. Morohoshi T, Kato M, Fukamachi K, Kato N, Ikeda T. 2008. N-acylhomoserine lactone regulates violacein production in *Chromobacterium violaceum* type strain ATCC 12472. FEMS Microbiol Lett 279:124–130. https://doi .org/10.1111/j.1574-6968.2007.01016.x.
29. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. Biol Direct 7:12. https://doi.org/10.1186/1745-6150-7-12.
30. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily do- main architectures. Nucleic Acids Res 45:D200 –D203. https://doi.org/10 .1093/nar/gkw1129.
31. Fuqua C, Parsek MR, Greenberg EP. 2001. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. Annu Rev Genet 35:439 – 468. https://doi.org/10.1146/annurev.genet.35 .102401.090913.
32. Whitehead NA, Barnard AM, Slater H, Simpson NJ, Salmond GP. 2001. Quorum-sensing in Gram-negative bacteria. FEMS Microbiol Rev 25: 365–404. https://doi.org/10.1111/j.1574-6976.2001.tb00583.x.
33. Zhu J, Winans SC. 2001. The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. Proc Natl Acad Sci U S A 98:1507–1512. https://doi.org/10.1073/pnas.98.4.1507.
34. Truong TT, Seyedsayamdost M, Greenberg EP, Chandler JR. 2015. A *Burkholderia thailandensis* acyl-homoserine lactone-independent orphan LuxR homolog that activates production of the cytotoxin malleilactone. J Bacteriol 197:3456 –3462. https://doi.org/10.1128/JB.00425-15.
35. Papenfort K, Silpe JE, Schramma KR, Cong JP, Seyedsayamdost MR, Bassler BL. 2017. A *Vibrio cholerae* autoinducer-receptor pair that con- trols biofilm formation. Nat Chem Biol 13:551–557. https://doi.org/10 .1038/nchembio.2336.
36. Monk JM, Koza A, Campodonico MA, Machado D, Seoane JM, Palsson BO, Herrgård MJ, Feist AM. 2016. Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain pheno- types. Cell Syst 3:238 –251.e12. https://doi.org/10.1016/j.cels.2016.08.013.
37. Ortega MA, Hao Y, Zhang Q, Walker MC, van der Donk WA, Nair SK. 2015. Structure and mechanism of the tRNA-dependent lantibiotic dehydra- tase NisB. Nature 517:509 –512. https://doi.org/10.1038/nature13888.

1. Degen D, Feng Y, Zhang Y, Ebright KY, Ebright YW, Gigliotti M, Vahedian-Movahed H, Mandal S, Talaue M, Connell N, Arnold E, Fenical W, Ebright RH. 2014. Transcription inhibition by the depsipeptide antibiotic salinamide A. Elife 3:e02451. https://doi.org/10.7554/eLife.02451.
2. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, Aigle B. 2011. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. Proc Natl Acad Sci U S A 108:6258 – 6263. https://doi.org/10.1073/pnas.1019077108.
3. Li S, Li Y, Lu C, Zhang J, Zhu J, Wang H, Shen Y. 2015. Activating a cryptic ansamycin biosynthetic gene cluster to produce three new naphthalenic octaketide ansamycins with n-pentyl and n-butyl side chains. Org Lett 17:3706 –3709. https://doi.org/10.1021/acs.orglett.5b01686.
4. Mutka SC, Carney JR, Liu Y, Kennedy J. 2006. Heterologous production of epothilone C and D in *Escherichia coli*. Biochemistry 45:1321–1330. https://doi.org/10.1021/bi052075r.
5. Kües U, Stahl U. 1989. Replication of plasmids in Gram-negative bacteria. Microbiol Rev 53:491–516.
6. Noskov VN, Kouprina N, Leem SH, Ouspenski I, Barrett JC, Larionov V. 2003. A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. BMC Genomics 4:16. https://doi.org/10.1186/1471-2164-4-16.
7. Zhou K, Zhou L, Lim Q', Zou R, Stephanopoulos G, Too HP. 2011. Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. BMC Mol Biol 12:18. https://doi.org/10.1186/1471-2199-12-18.

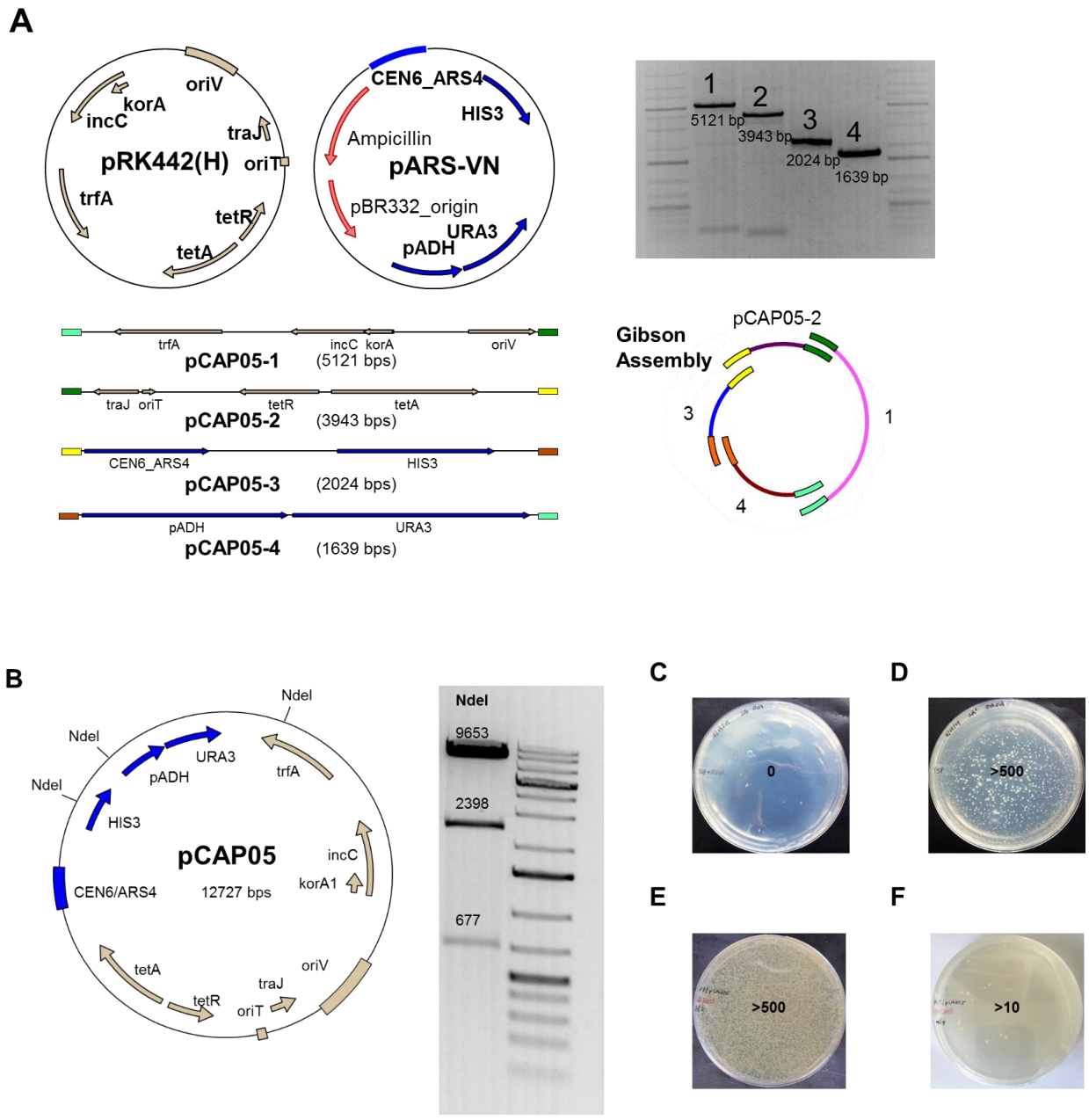


FIG S1

Construction and validation of pCAP05. (A) Approach used to combine elements from pRK442(H) and pARS-VN using Gibson Assembly. Elements maintained are shown in bold type. (B) Confirmation by restriction digestion with NdeI. (C) Transfer to *S. cerevisiae* VL6-48N and plating on selective medium without histidine and with 5-FOA. (D) Transfer to *S. cerevisiae* VL6-48N and plating on selective medium without histidine. (E) Transfer to *P. putida* KT2440 by electroporation with selection with tetracycline at 15 µg/ml in LB medium. (F) Transfer to *A. tumefaciens* LBA4404 by electroporation with selection with tetracycline at 5 µg/ml in YM medium.

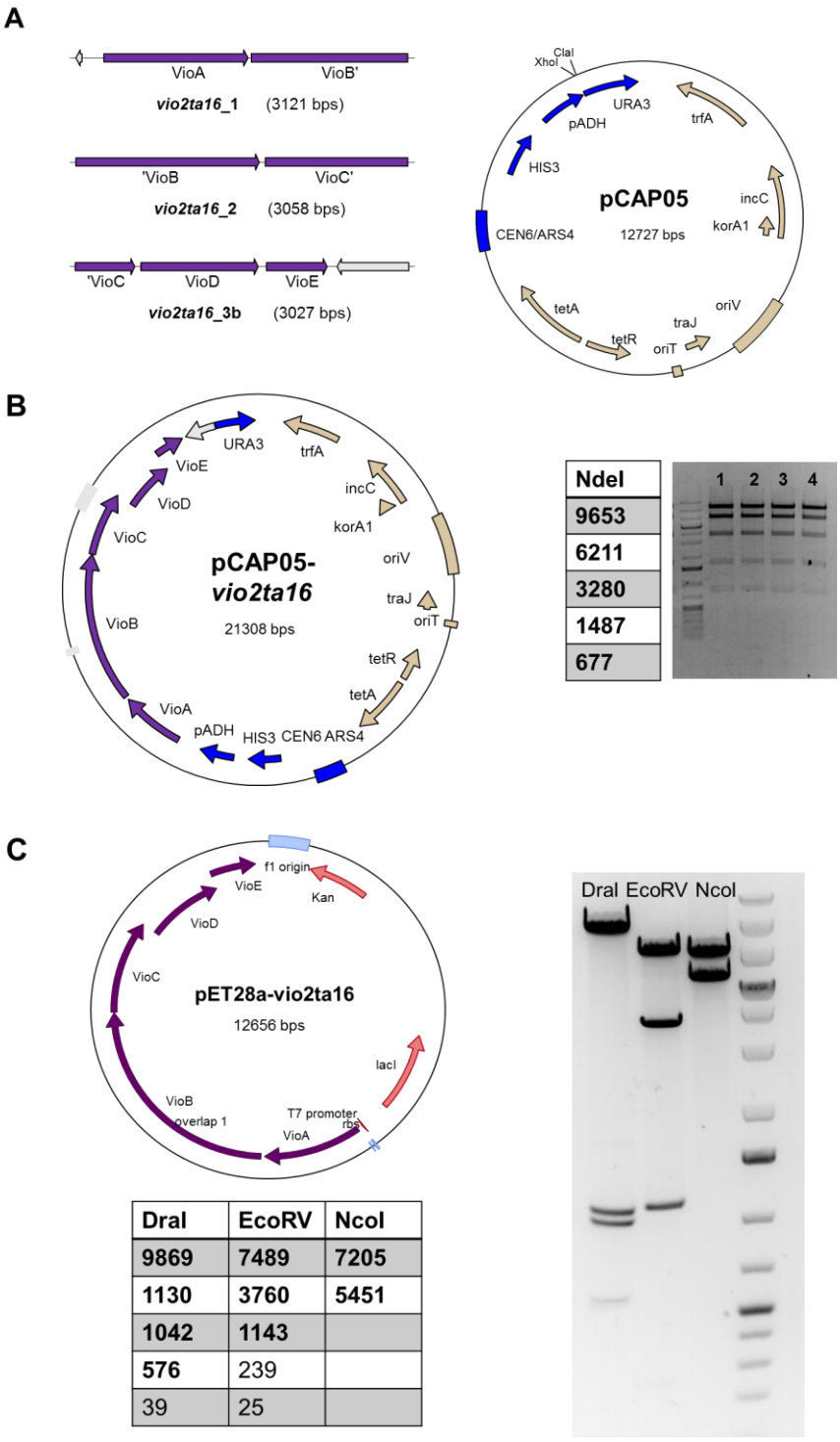


FIG S2

Validation of pCAP05-*vio2ta16* and pET28a-*vio2ta16*. (A) DNA used for TAR cloning. Three fragments was amplified from *P. luteoviolacea* 2ta16 gDNA and combined with pCAP05 digested with ClaI and XhoI. A 0.5-µg sample of each DNA fragment was transferred to *S. cerevisiae* VL6-48N spheroplasts and plated on selective medium without histidine and with 5-FOA. (B) Vector map of pCAP05-*vio2ta16* along with digestion confirmation from independent clones. (C) Expected restriction digestion fragments of pET28a-*vio2ta16*obtained with DraI, EcoRV, and NcoI are listed in a table, where bands that could be visualized in the gel are in bold. The construct was also sequenced at the insertion site with the T7 and T7 terminator primers (data not shown).

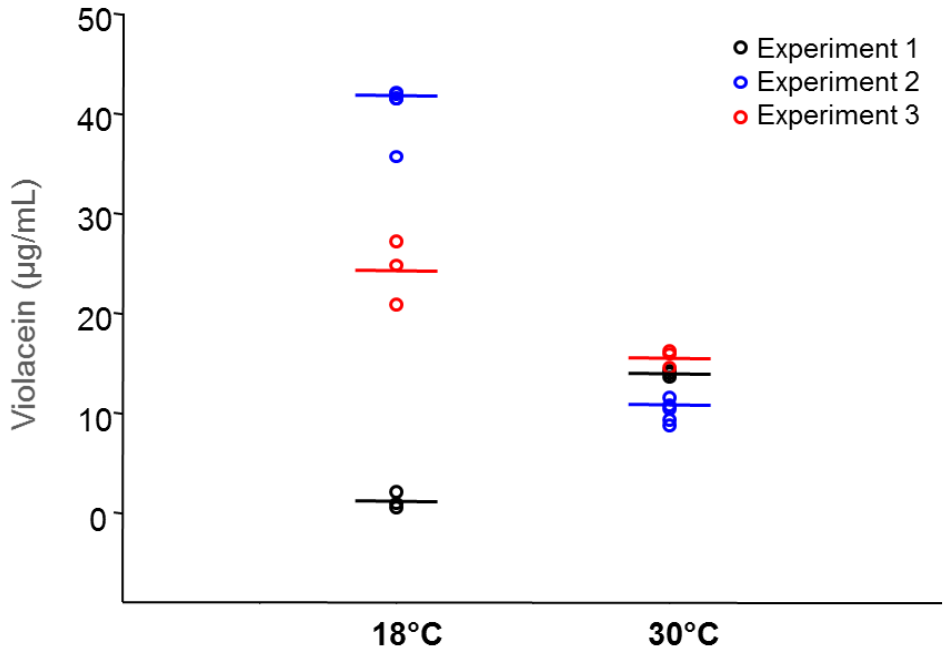
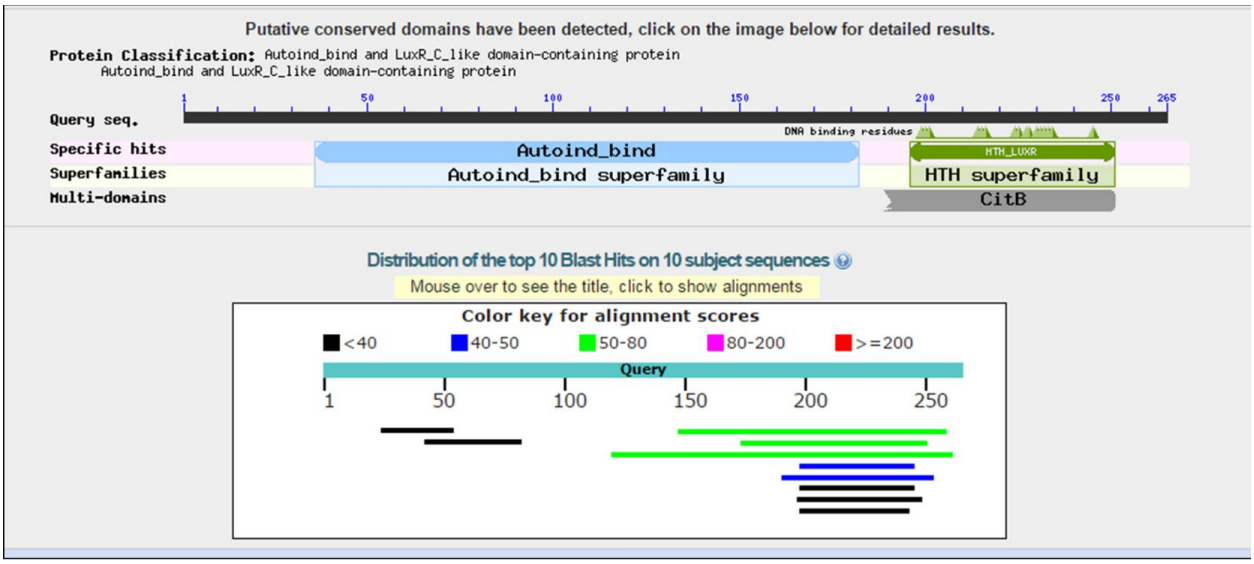


FIG S3

Quantification of violacein production by *P. luteoviolacea* 2ta16 grown at two different temperatures. The data from three independent experiments are in black, blue, and red.



Name	Locus Tag (PL2TA16_)	Score	Coverage (%)	E value	Identity (%)
PLR1	01227	73.9	41%	2e-16	36%
PLR2	00118	66.2	29%	9e-14	44%
PLR3	03988	56.2	53%	3e-10	30%
PLR4 (PviR)	04751	43.1	18%	9e-06	48%
PLR5	01039	40.4	23%	7e-05	33%
PLR6	04047	38.5	18%	3e-04	40%
PLR7	00814	38.1	19%	4e-04	40%

FIG S4

Results of a DELTA-BLAST search querying the genome of *P. luteoviolacea* 2ta16 by using CviR from *C. violaceum* ATCC 31532. All of the hits possess E values above the significance threshold. The top panel shows sequence alignments. The table lists candidates with putative two-component system response regulators in blue.

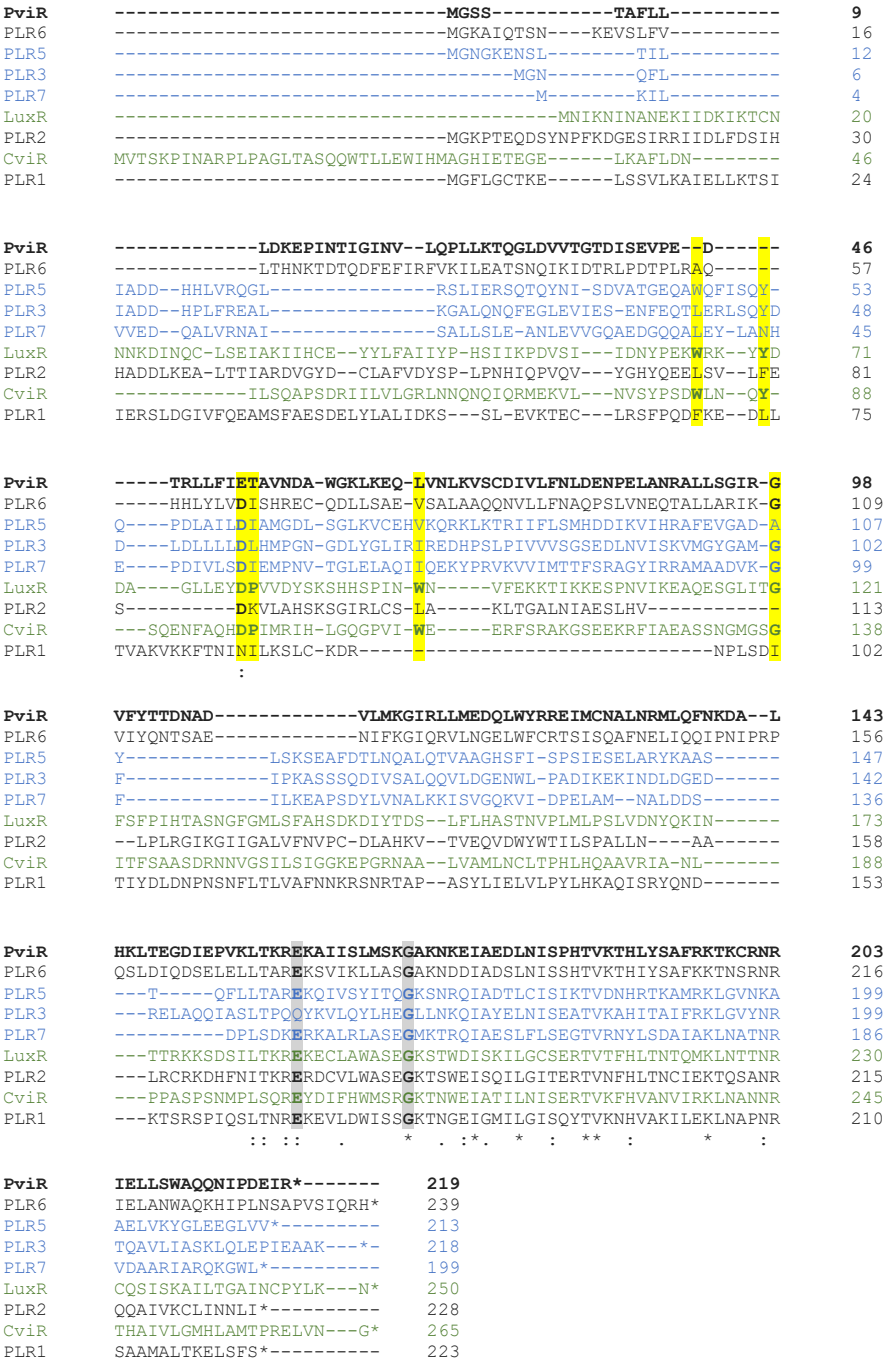
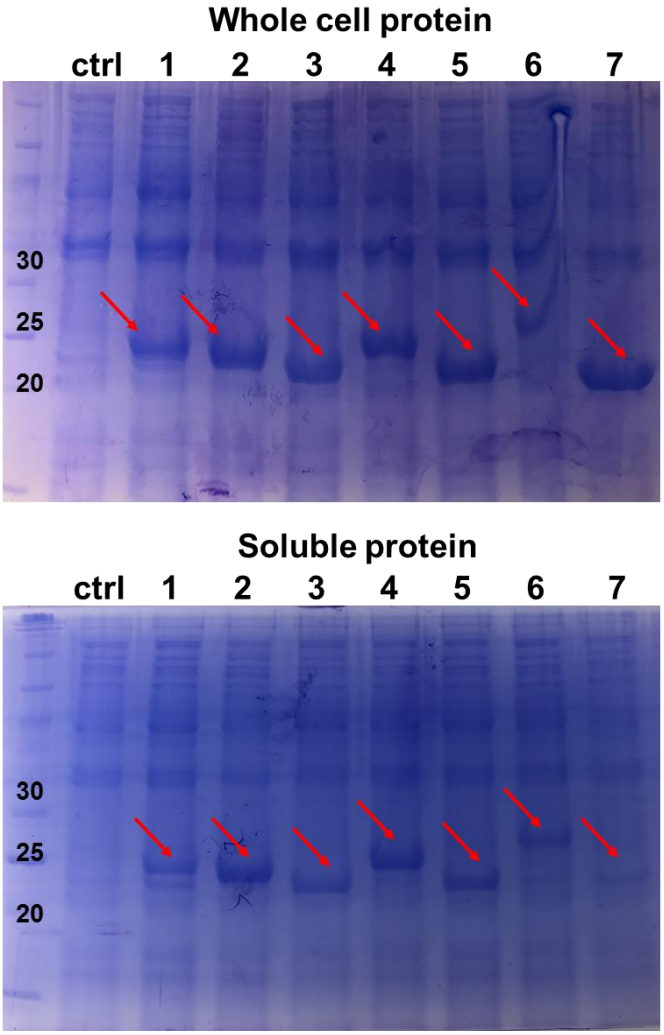


FIG S5

Multiple-sequence alignment of CviR, LuxR, and PLR1-7 generated with CLUSTAL O(1.2.4). Amino acid residues conserved across quorum-sensing signal receptors within the autoinducer- and DNA-binding domains are highlighted in yellow and gray, respectively.

A



PL2TA16: (kDa)

1. PLR1: 25.06

2. PLR2: 25.7

3. PLR3: 24.35

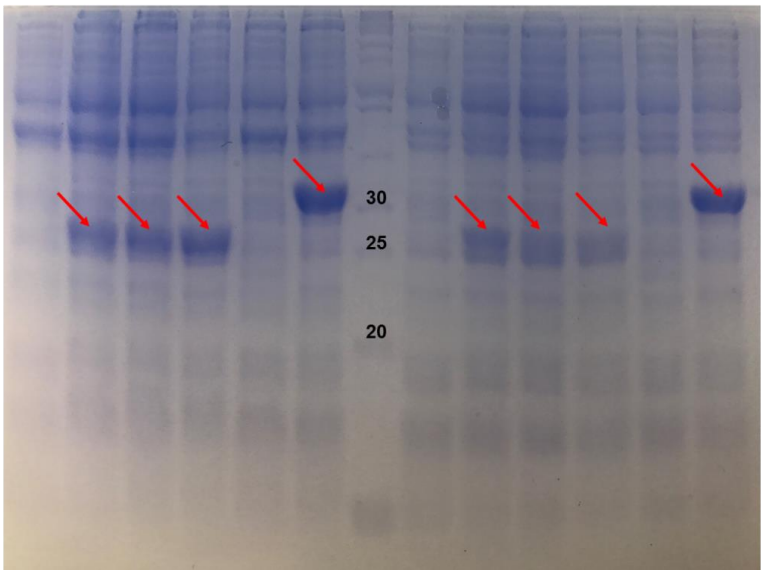
4. PLR4 (PviR): 24.88

5. PLR5: 23.74

6. PLR6: 26.97

7. PLR7: 21.95

B



***E. coli* : (kDa)**

1. CsgD_BL21: 24.98

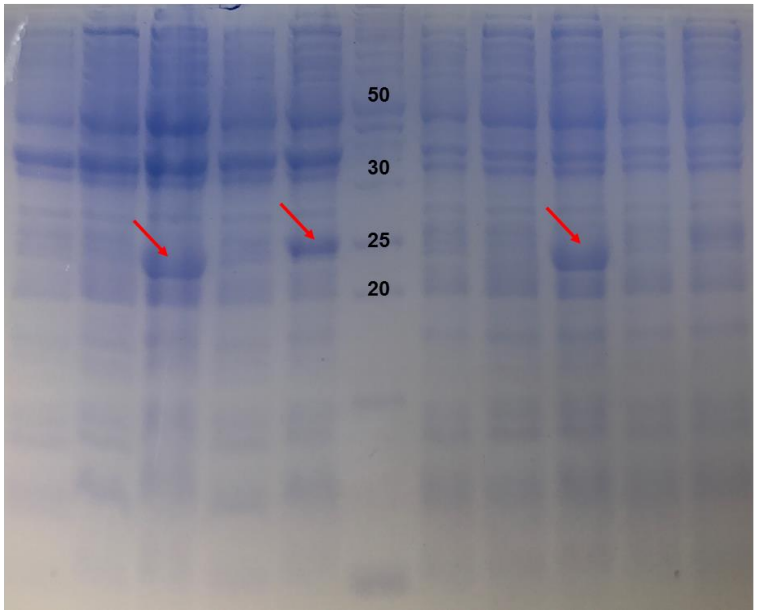
2. CsgD_5a: 24.94

3. YhjB: 22.67

4. UhpA: 20.95

5. SdiA: 28.18

C



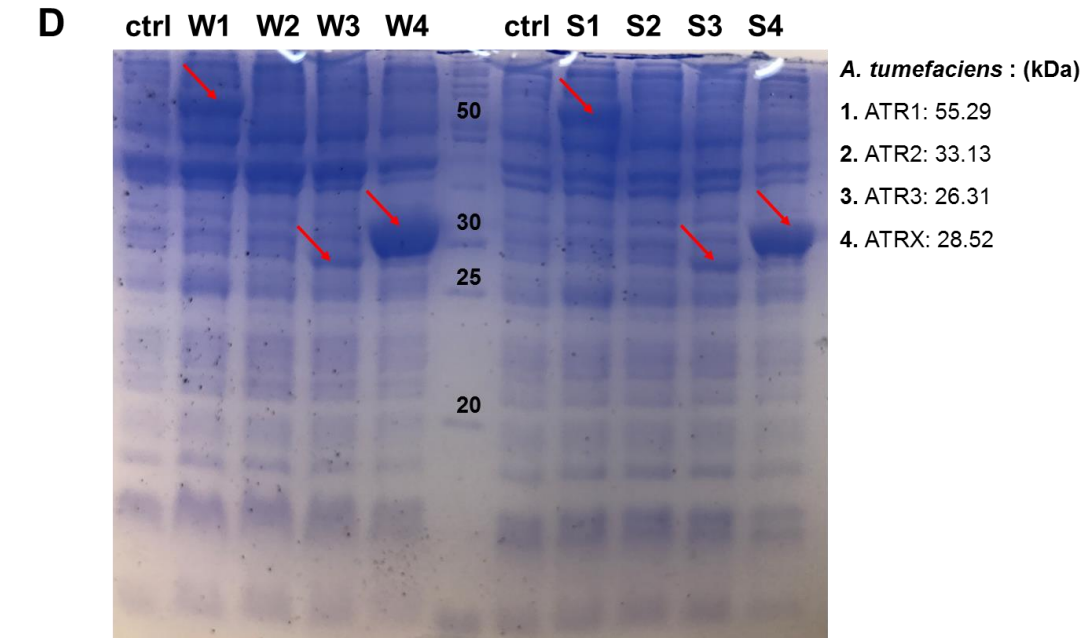
***P. putida* : (kDa)**

1. PPR1: 23.93

2. PPR2: 24.6

3. PPR5: 100.45

4. PPRX: 26.86



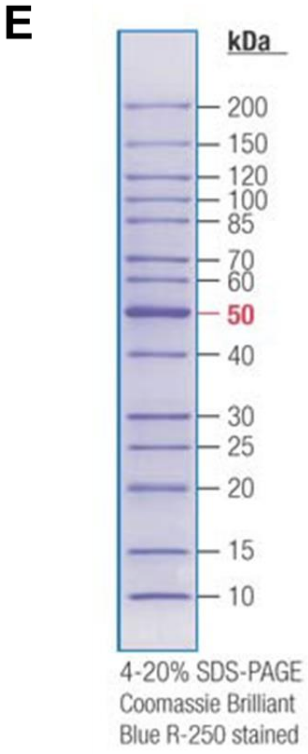
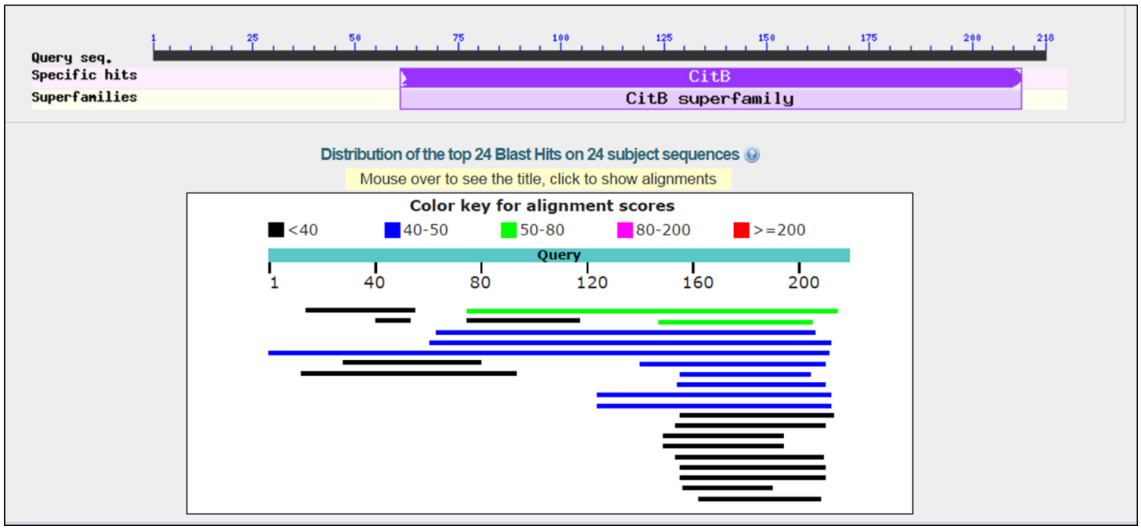


FIG S6

SDS-PAGE analysis of protein extracts of *E. coli* BL21(DE3) overexpressing candidate quorum-sensing signal receptors identified in native and heterologous hosts. Panels: A, *P. luteoviolacea* 2ta16; B, *E. coli*; C, *P. putida* KT2440; D, *A. tumefaciens* LBA4404; E, protein molecular size ladder used in all of the gels.

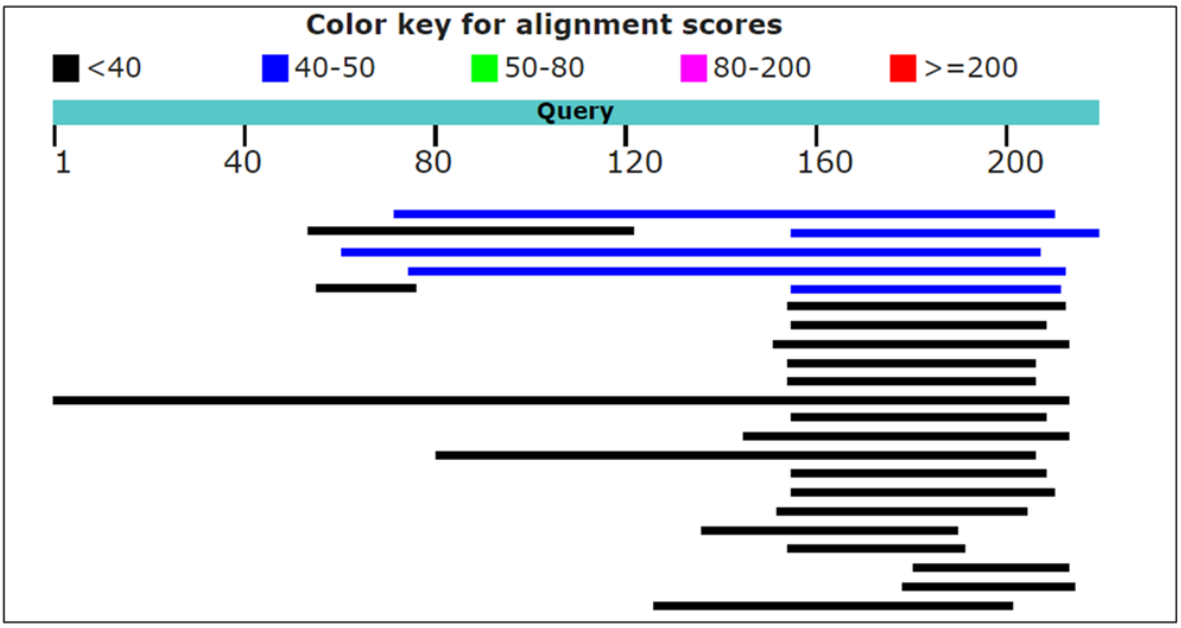
A



	Gene name	Locus Tag	Score	Coverage (%)	E value	Identity (%)
1	csgD	b1040	68.9	63	3e-15	29
2	yhjB	b3520	50.8	26	1e-08	42
3	uhpA	b3669	49.3	65	3e-08	26
4	narL	b1221	49.3	69	4e-08	29
5	fimZ	b0535	48.5	96	7e-08	22
6	uvrY	b1914	46.2	32	5e-07	35
7	narP	b2193	43.9	22	3e-06	43
8	N/A	EO53_03510	42.7	25	4e-06	41
9	malT	b3418	41.6	40	3e-05	27
10	evgA	b2369	39.3	26	9e-05	33
	sdiA	b1916	33.5	26	0.008	32

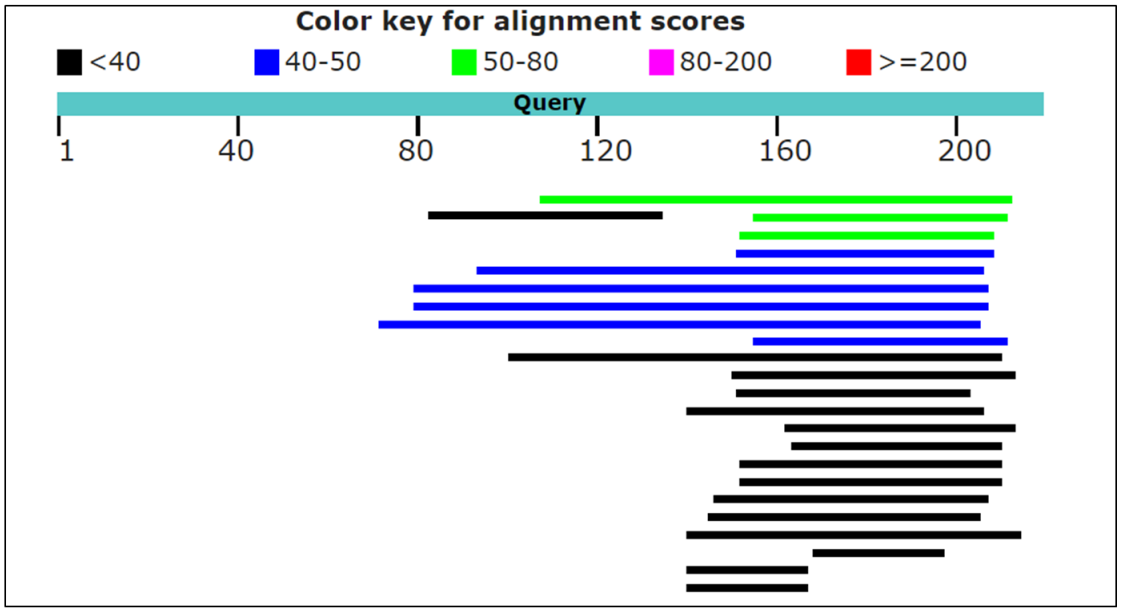
	Gene name	Locus Tag	Score	Coverage (%)	E value	Identity (%)	Identical protein?
1	csgD	B21_01044	70.5	63	7e-16	30	N
2	yhjB	B21_03321	50.8	26	8e-09	42	Y
3	uhpA	B21_03495	49.3	65	2e-08	26	Y
4	narL	B21_01209	49.3	69	3e-08	29	Y
5	fimZ	B21_00490	48.5	96	6e-08	22	N
6	N/A	B21_RS09895	46.6	32	2e-07	35	N
7	uvrY	ECD_01879	46.2	32	2e-07	35	N
8	narP	B21_02079	43.9	22	2e-06	43	N
9	malT	B21_03222	41.6	40	2e-05	27	N
10	evgA	B21_02240	39.3	26	7e-05	33	Y

B



Name	Locus Tag	Score	Coverage (%)	E value	Identity (%)
PPR1	PP_2126	47.4	63	2e-07	25
PPR2	PP_2665	44.3	29	2e-06	35
PPR3	PP_2672	43.1	66	6e-06	23
PPR4	PP_1635	42.4	62	9e-06	21
PPR5	**PP_0767**	**41.6**	**25**	**3e-05**	**38**
PPR6	PP_1090	38.9	26	1e-04	31
PPR7	PP_3847	39.3	24	1e-04	42
PPR8	PP_3412	37.4	28	5e-04	29
PPR9	PP_5241	37	23	9e-04	35
PPR10	PP_0574	36.6	23	1e-03	37
PPR11	PP_2101	36.2	97	1e-03	19
PPR12	PP_2587	36.6	24	1e-03	36
PPR13	PP_0410	36.2	31	1e-03	29
PPR14	PP_4099	35.8	57	1e-03	26
PPR15	PP_3717	35.8	24	2e-03	32
PPRX	PP_4647	28.9	22	N/A	29

C



	Locus Tag	Score	Coverage (%)	E value	Identity (%)
ATR1	**X971_4629**	**53.9**	**47**	**3e-09**	**27**
ATR2	**X971_3376**	**52.4**	**25**	**9e-09**	**45**
ATR3	**X971_2241**	**50.8**	**25**	**1e-08**	**45**
ATR4	X971_3295	49.7	26	4e-08	44
ATR5	X971_2958	47.4	51	2e-07	29
ATR6	X971_2949	47	58	3e-07	28
ATR7	X971_0445	42	61	1e-05	23
ATR8	X971_4153	40	25	5e-05	39
ATR9	X971_0488	39.7	50	9e-05	30
ATR10	X971_0487	39.7	28	1e-04	32
ATR11	X971_2373	36.2	23	1e-03	35
ATR12	X971_4763	35.4	30	3e-03	35
ATR13	X971_2720	32.3	23	4e-03	31
ATRX	**X971_0678**	**30.8**	**26**	**N/A**	**31**

FIG S7

Results of a DELTA-BLAST search querying the genomes of *E. coli* (A), *P. putida* KT2440 (B), and *A. tumefaciens* LBA4213 (C) with PviR from *P. luteoviolacea* 2ta16. The top panels show sequence alignments of homologs to PviR, displayed in order below the query sequence. The tables list BLAST hits, with genes selected for testing in bold and underlined, putative two-component system response regulators in blue, and hits with E values below the significance threshold in red. In panel A, K-12 homologs are shown in the top table and BL21(DE3) homologs are shown in the bottom table, with BLAST hits that differ highlighted in peach.

FIG S8

Multiple-sequence alignment of CviR, LuxR, ATR1, and ATR3 generated with CLUSTAL O(1.2.4). Amino acid residues conserved across quorum-sensing signal receptors within the autoinducer- and DNA-binding domains are highlighted in yellow and gray, respectively.

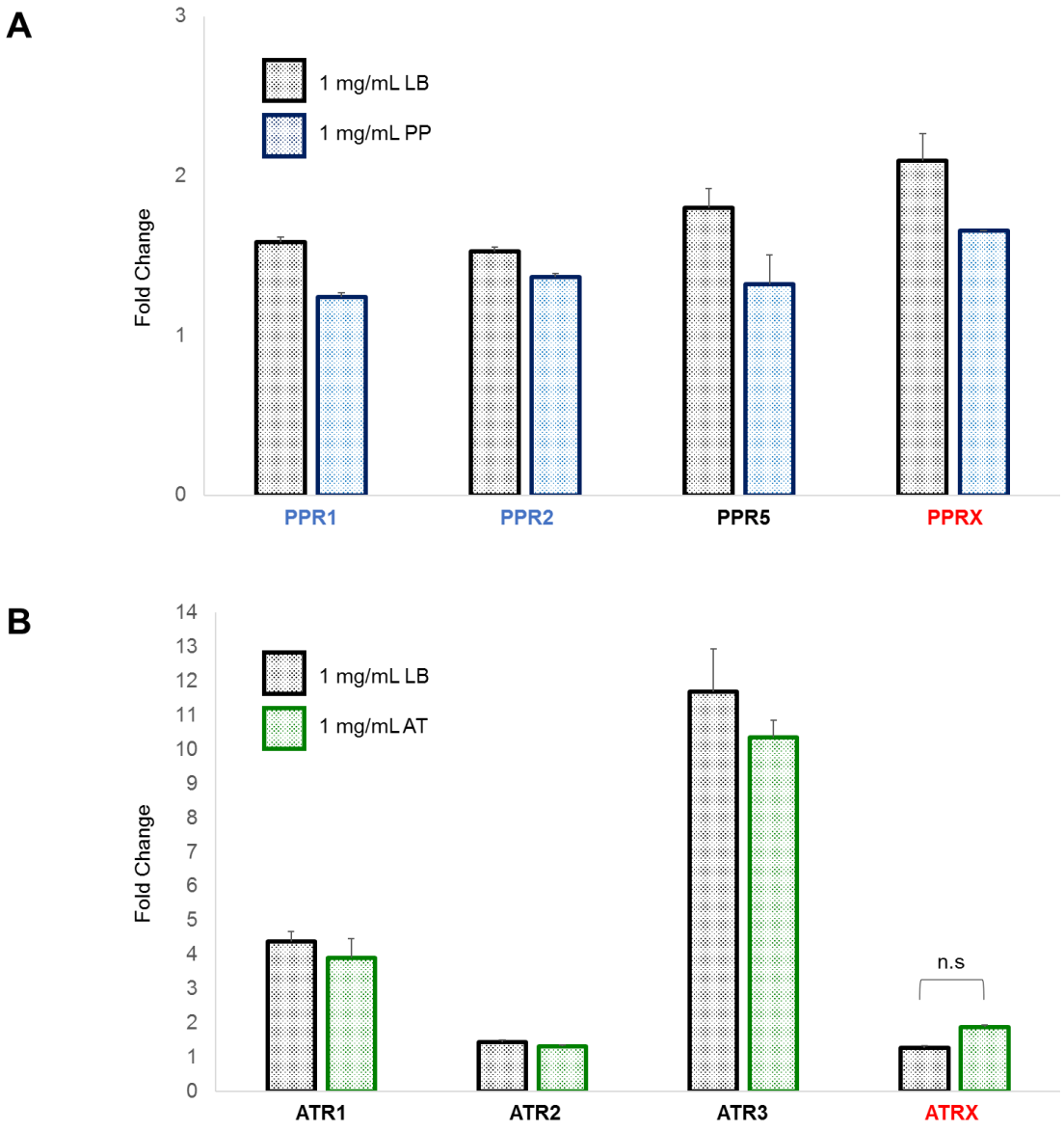


FIG S9

Culture supernatants from *P. putida* and *A. tumefaciens* do not enhance receptor activity on the *vio2ta16* promoter in an *E. coli* expression system. Heterologous violacein production by *E. coli* BL21(DE3) through overexpression of PviR homologs from *P. putida* KT2440 (A) and *A. tumefaciens* LBA4404 (B) with addition of culture supernatants (PP and AT, respectively). Freeze-dried supernatants from log-phase *P. putida* KT2440 and *A. tumefaciens* LBA4404 cultures grown in LB at 30°C were dissolved in DMSO and added to *E. coli* BL21(DE3) coexpressing putative regulatory genes with *vio2ta16* at the time of IPTG induction. Freeze-dried LB medium prepared in the same way was also added as a control. The data are plotted as the mean ± standard error of the fold change compared to violacein production from *E. coli* BL21(DE3)/pCAP05-*vio2ta16*/pACYCDuet-1 treated with either control or culture supernatants. Significance was determined with a one-tailed Student t test ($n = 2$, $P < 0.05$). Addition of culture supernatants resulted in an increase in violacein production for only one regulatory gene, ATRX, which was not statistically significant (n.s.).

TABLE S1

Primers and plasmids used in this study.

Name	Sequence
pCAP05-1_5121F	**AGCTCCAATTCGCCCTATAG**ATCACGAGCAATAAACCCGC
pCAP05-1_5121R	**CTGGAAAGCGGGCAGTGAGC**GGCCTTTTGCTCACATGTTC
pCAP05-2_3943F	**GAACATGTGAGCAAAAGGCC**GCTCACTGCCCGCTTTCCAG
pCAP05-2_3943R	**CCTATTTGTTTATTTTTCTA**ACGATTGATGGCGGTCCTGG
pCAP05-3_2024F	**CCAGGACCGCCATCAATCGT**TAGAAAAATAAACAAATAGG
pCAP05-3_2024R	**GGGTTAATTCCGAGCTTGGC**TATTTCACACCGCATATGAT
pCAP05-4_1639F	**ATCATATGCGGTGTGAAATA**GCCAAGCTCGGAATTAACCC
pCAP05-4_1639R	**GCGGGTTTATTGCTCGTGAT**CTATAGGGCGAATTGGAGCT
vio2ta16-1_3121-40F	**CTCGGTTTGACGCCTCCCATGGTATAAATAGTGGCTCGAG**CCTAAACTAACCGCCATGTG
vio2ta16-1_3121-134R	CTGTAGTAGCTGCGTATGTCCAGTGACTGT
vio2ta16-2_3058-134F	TGCTCAGAGCGGCACACTTG
vio2ta16-2_3058-108R	TCTGGCGGGTAATGATCTGG
vio2ta16-3b_3027-482F	GCGGTTTGCCTTCCTTATACGGGTCAAGTG
vio2ta16-3b_3027-40R	**AGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATCGAT**CCGCGGGTTTCCTAAATTAC
vio2ta16check_402F	CTCGGGAGATGCCAGAAGTC
vio2ta16check_402R	GCTCGCCCAATGCCATAAGG
pET28a-vioE-F	**TGAAATTTTTCAACCGACTCCATTAAAACAACCTGAGTAA**GCACTCGAGCACCACCACCA
pET28a-vioA-R	**CCGCTCCGATGATTGAGATGGTGTTTTCTCTAACCGTCAT**GTATATCTCCTTCTTAAAGT
PLR1(NcoI)_669F	**ACCTGCCCATGGGC**TTTTTGGGCTGCACAAAAGA
PLR1(HindIII)_669R	**GGCCGCAAGCTT**TTAGCTAAAGCTCAGCTCTTTG
PLR2(NcoI)_684F	**ACCTGCCCATGGGC**AAACCTACCGAACAAGATAG
PLR2(HindIII)_684R	**GGCCGCAAGCTT**CTAAATAAGGTTGTTTATTAAACACT
PLR3(NcoI)_654F	**ACCTGCCCATGGGC**AATCAGTTTTTAATTGCGGATGA
PLR3(HindIII)_654R	**GGCCGCAAGCTT**CTACTTTGCCGCCTCGATT
PLR4(NcoI)_657F	**ACCTGCCCATGGGC**AGCAGTACTGCATTTTTATTG

PLR4(HindIII)_657R	**GGCCGC<u>AAGCTT</u>**TTAACGTATCTCGTCAGGGA
PLR5(NdeI)_642F	**CGGAATTC<u>CATATG</u>**GGTAACGGAAAAGAAAATAG
PLR5(XhoI)_642R	**AACCG<u>CTCGAG</u>**CTACACAACTAGCCCTTCTT
PLR6(NcoI)_717F	**ACCTGC<u>CCATGG</u>GC**AAAGCGATTCAAACTAGCAA
PLR6(HindIII)_717R	**GGCCGC<u>AAGCTT</u>**TTAATGACGCTGTATTGAGA
PLR7(NdeI)_600F	**CGGAATTC<u>CATATG</u>**AAAATTTTAGTTGTAGAAGATCAA
PLR7(XhoI)_600R	**AACCG<u>CTCGAG</u>**TTACAGCCAACCCTTTTGTC
qRT-cysG_126F	TCCACAGTTCACCGCATGGG
qRT-cysG_126R	CGCTGGTTAAGCGCGTCATC
qRT-vioA_140F	CTGACAATGCTGGCTATGAG
qRT-vioA_140R	CGCAGTATCAGCTTCTAGAC
csgD(NdeI)_651F	**CGGAATTC<u>CATATG</u>**TTTAATGAAGTCCATAGTATTCATG
csgD(XhoI)_651R	**AACCG<u>CTCGAG</u>**TTATCGCCTGAGGTTATCGT
yhjB(NcoI)_603F	**ACCTGC<u>CCATGG</u>GC**CAAATAGTCATGTTTGACAG
yhjB(HindIII)_603R	**GGCCGC<u>AAGCTT</u>**TCAGGAGGAGATATTTAACA
uhpA(NcoI)_591F	**ACCTGC<u>CCATGG</u>GC**ATCACCGTTGCCCTTATAGA
uhpA(HindIII)_591R	**GGCCGC<u>AAGCTT</u>**TCACCAGCCATCAAACATGC
sdiA(NcoI)_723F	**ACCTGC<u>CCATGG</u>GC**CAGGATAAGGATTTTTTCAGCT
sdiA(HindIII)_723R	**GGCCGC<u>AAGCTT</u>**TCAAATTAAGCCAGTAGCGG
PPR1(NdeI)_660F	**CGGAATTC<u>CATATG</u>**ACCTGTAGACTACTGCT
PPR1(XhoI)_660R	**AACCG<u>CTCGAG</u>**TCAGTCATCCAGGCTGATGA
PPR2(NcoI)_666F	**ACCTGC<u>CCATGG</u>GC**TACAAAATTCTGATTGCCGAC
PPR2(HindIII)_666R	**GGCCGC<u>AAGCTT</u>**TCAACGGCGCAGATAATCA
PPR5(NdeI)_2718F	**CGGAATTC<u>CATATG</u>**ACAGATCTGTCCCGTACCCA
PPR5(XhoI)_2718R	**AACCG<u>CTCGAG</u>**TCACGCCAGCAACCCCAGG
PPRX(NcoI)_708F	**ACCTGC<u>CCATGG</u>GC**CTTCACTGGAAACCCGA
PPRX(HindIII)_708R	**GGCCGC<u>AAGCTT</u>**TCAGATCCAGCCACGCAAGG
ATR1(NdeI)_1479F	**CGGAATTC<u>CATATG</u>**GACGACAGCGTAAACAT
ATR1(XhoI)_1479R	**AACCG<u>CTCGAG</u>**TTATTTTTCCCTATGGCCGG

ATR2(NdeI)_930F	**CGGAATTCCATATG**AGCGGGCAAGGCGCAG
ATR2(XhoI)_930R	**AACCGCTCGAG**TCACCCCTGATGCAAAATATGCGCG
ATR3(NcoI)_717F	**ACCTGCCCATGGGC**CCGAAGACGGAGCGCCAGG
ATR3(HindIII)_717R	**GGCCGCAAGCTT**TCAGCTTATCAGGCCAAGGCG
ATRX(NcoI)_756F	**ACCTACCATGGGC**AAGAACCGCTTCATGCA
ATRX(HindIII)_756R	**GGCCGCAAGCTT**CTATTTAATCAGACCGACCC

Name	Description	Reference
pARS-VN	Plasmid used to amplify selective, counter-selective, and maintenance elements for yeast TAR cloning in construction of pCAP05	[1]
pRK442(H)	Plasmid used to amplify broad-host-range origin of replication and maintenance elements for Gram-negative bacteria in construction of pCAP05	[2]
pCAP05	TAR cloning and broad-host-range heterologous expression vector; CEN6_ARS4, *HIS3*, p*AHD1*, *URA3*, *oriV*, *traJ*, *oriT*, *tetR*, *tetA*, *trfA*, *incC*, *korA*	This study
pCAP05-vio2ta16	pCAP05 containing violacein gene cluster from *P. luteoviolcea* 2ta16; Tet^R	This study
pET28a-vio2ta16	pET28a containing violacein gene cluster from *P. luteoviolcea* 2ta16 under control of T7 promoter; Kan^R	This study
pACYCDuet-1	Commercial expression vector used for co-expression of LuxR homologs; Cm^R	Novagen

1 Noskov, V. N., Kouprina, N., Leem, S. H., Ouspenski, I., Barrett, J. C. & Larionov, V. A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics* **4,** 16, doi:10.1186/1471-2164-4-16 (2003).

2 Scott, H. N., Laible, P. D. & Hanson, D. K. Sequences of versatile broad-host-range vectors of the RK2 family. *Plasmid* **50,** 74-79 (2003).

3.3 Acknowledgement

Chapter 3, in full, is a reprint of the material as it appears in *mBio*, Zhang, J. J.; Tang, X.; Zhang, M.; Nguyen, D.; Moore, B. S., 2017. The dissertation author was the primary investigator and author of this paper.

CHAPTER 4. Avant-garde assembly line biosynthesis expands diversity of cyclic lipodepsipeptide products.

4.1 Abstract

Modular nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymatic assembly lines are large and dynamic protein machines that generally effect a linear sequence of catalytic cycles. Here we report the heterologous reconstitution and comprehensive characterization of two hybrid NRPS-PKS assembly lines that defy many standard rules of assembly line biosynthesis to generate a large combinatorial library of cyclic lipodepsipeptide protease inhibitors called thalassospiramides. We generate a series of precise domain-inactivating mutations in thalassospiramide assembly lines and present compelling evidence for an unprecedented biosynthetic model that invokes inter-module substrate activation and tailoring, module skipping, and pass-back chain extension, whereby the ability to pass the growing chain back to a preceding module is flexible and substrate-driven. Expanding bidirectional inter-module domain interactions could represent a viable mechanism for generating chemical diversity without increasing the size of biosynthetic assembly lines and challenges our understanding of the potential elasticity of multi-modular megaenzymes.

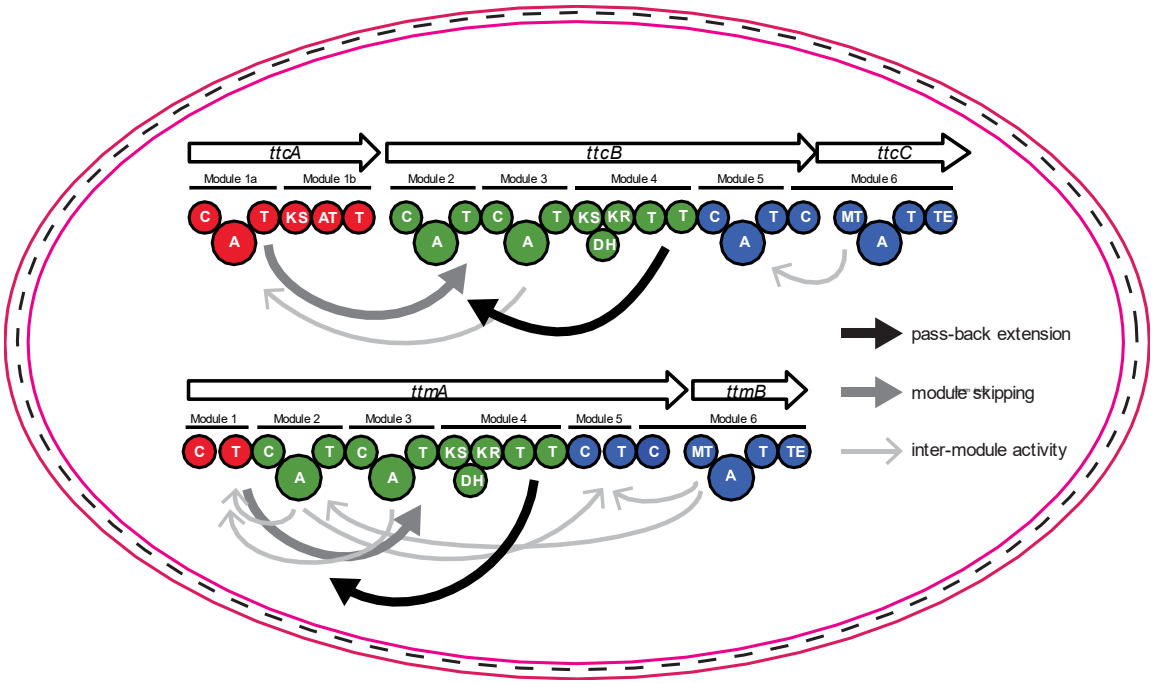


Figure 4.1 Graphical abstract of the chapter.

4.2 Introduction

Modular nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymes are molecular-scale assembly lines that construct complex polymeric products, many of which are useful to humans as life-saving drugs. The first characterized assembly lines exhibited an elegant co-linear biosynthetic logic, whereby the linear arrangement of functional units, called modules, along an NRPS/PKS polypeptide directly correlates to the chemical structure of the product[1]. The PKS giving rise to the antibiotic erythromycin[2] and the NRPS producing the antibiotic daptomycin[3] are two such examples. The core components of an NRPS or PKS assembly line elongation module include, respectively, condensation (C) or ketosynthase (KS) domains catalyzing chain extension, adenylation (A) or acyltransferase (AT) domains for substrate selection, and thiolation (T) domains for covalent substrate tethering. Optional tailoring domains such as methyltransferase (MT), ketoreductase (KR), or dehydratase (DH) domains, if present, chemically modify building blocks or chain-extension intermediates. This one-to-one correlation between product moieties and assembly line modules with the requisite catalytic domains is one feature that makes NRPS/PKS enzymes among the largest proteins found in nature.

However, it is now clear that many assembly lines do not strictly abide by the rules of co-linearity. A phylogenetically distinct class of modular PKSs are *trans*-AT PKSs, which do not directly encode AT domains within modules but instead, as stand-alone enzymes that act *in trans*[4]. A separate type of NRPS, referred to as a nonlinear NRPS, deviates from the standard core domain arrangement of C-A-T and reuses a single domain more than once[5]. Several nonlinear NRPSs possess modules missing A domains that are presumably loaded by A domains from upstream modules[6-10]. Leveraging domain activities *in trans* or from different modules reduces the size of biosynthetic assembly lines and thus may represent a mechanism for minimizing modular assembly lines without sacrificing product complexity.

A particularly intriguing set of hybrid NRPS-PKS assembly lines that are both nonlinear and *trans*-AT are those responsible for the biosynthesis of the thalassospiramides, a large group of immunosuppressive cyclic lipodepsipeptides[11-13]. Thalassospiramide NRPS-PKS genes have been identified in several marine *Rhodospirillaceae* bacteria and exhibit distinct architectures that range in domain and module "completeness"[13]. While it is still unclear whether all configurations are functional, thalassospiramide assembly lines representing the most "complete" and "incomplete" architectures are both capable of producing a large combination of lipodepsipeptides that vary in fatty acid and amino acid composition, order, and length[12]. Previous work identifying these NRPS-PKS genes and structurally characterizing their numerous and diverse chemical products[11-13] led us to hypothesize that these assembly lines must operate with an unprecedented degree of nonlinearity. Furthermore, we posited that in order to generate their chemical products, thalassospiramide assembly lines must catalyze one or two rounds of pass-back chain extension, where the chain-extension intermediate is passed from a downstream module back to an upstream module within the same polypeptide.

Here, we report a comprehensive characterization of the thalassospiramide biosynthetic machinery from α-proteobacteria *Thalassospira* sp. CNJ-328 and *Tistrella mobilis* KA081020-065, which represent the most "complete" and "incomplete" assembly line architectures, respectively[13]. We present an experimentally supported and mechanistically novel biosynthetic model that invokes inter-module substrate activation and tailoring, module skipping, and pass-back chain extension, whereby the ability to pass the growing chain forward or backward is flexible and influenced by the identity and chain length of the chemical intermediate. These newly described features accentuate the potential bidirectionality and flexibility of multi-modular megaenzymes and reveal new engineering opportunities and structural considerations.

4.3 Results

4.3.1 Cloning and heterologous expression of *ttc* and *ttm*.

Close inspection of the thalassospiramide *ttc* and *ttm* gene clusters revealed that the NRPS/PKS genes from *Thalassospira* and *Tistrella*, respectively, occupy completely different genomic contexts (**Figure 4.2, Tables 4.1-4.2**). *Ttc* also includes a 4'-phosphopantetheinyl transferase (PPTase), TtcD, for which there is no homolog in *ttm.* Notably absent from the genomic vicinity of both pathways are any genes encoding stand-alone AT or A domains (**Tables 4.1-4.2**), although both assembly lines possess a *trans*-AT PKS module and Ttm contains two NRPS modules without A domains (**Figure 4.3**). We chose to clone a broader range for *ttc* and a more limited range for *ttm* (**Figure 4.2**).

To directly clone these gene clusters, a new bacterial artificial chromosome (BAC) based transformation-associated recombination (TAR) cloning vector, pCAP-BAC (pCB), was designed and constructed to enable stable maintenance of large constructs in *Escherichia coli* (**Figure 4.7**). pCB lacks host-specific integration elements, which can be introduced after cloning, to make it easier to retrofit cloned pathways with integration elements for different hosts. Following successful cloning of *ttc*, heterologous expression was attempted but never achieved in *E. coli*, despite efforts to perform promoter refactoring, stabilize protein expression, and co-express the pathway with various promiscuous PPTases (**Figure 4.8**). Thus, we constructed a *Pseudomonas* integration cassette containing the Int-B13 site-specific recombinase[14] and introduced it into the vector backbone to generate pCB-*ttc*-int (**Figure 4.7**). This construct was successfully integrated into the genome of *Pseudomonas putida* EM383[15] (**Figure 4.9**). The same procedure was used for *ttm*, and expression of both gene clusters in *P. putida* was successful as evidenced by detection of the representative product thalassospiramide A (**1**) (**Figure 4.2, Figure 4.10**). This validated that the Ttc and Ttm assembly lines do not require additional pathway-specific

enzymatic components, beyond what was transferred to the host and supplied through primary metabolism. To our knowledge, this is the first report of successful heterologous expression of a *trans*-AT pathway without co-transfer of a cognate AT. Our current hypothesis as to why expression was successful in *P. putida* but not in *E. coli* is that the *P. putida* primary metabolic AT is capable of interfacing with the *trans*-AT PKS modules, while the AT from *E. coli* is not.

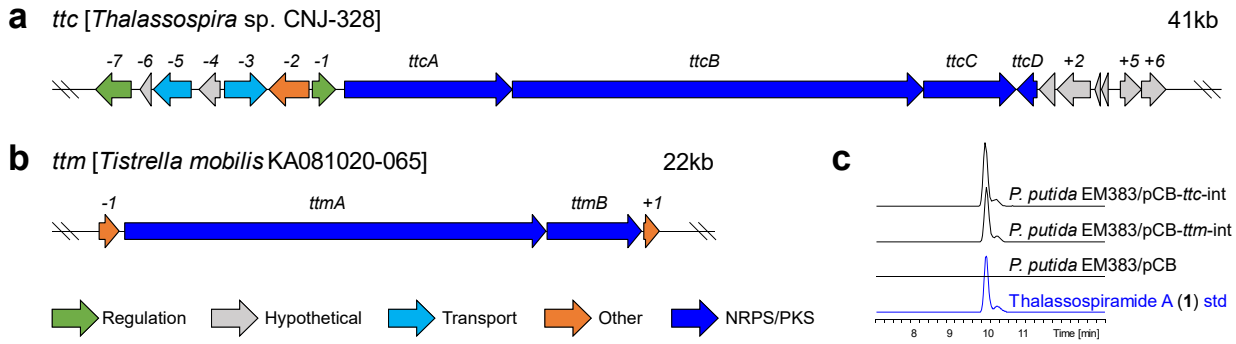


Figure 4.2 Heterologous reconstitution of thalassospiramide biosynthetic gene clusters in a *P. putida* host. Annotated genomic loci encompassing the thalassospiramide assembly line genes from (a) *Thalassospira* sp. CNJ-328 and (b) *Tistrella mobilis* KA081020-065 targeted for cloning and heterologous expression. (c) LC-MS analysis of extracts from an empty *P. putida* EM383 host and hosts with genomically integrated *ttc* and *ttm* pathways compared against an authentic thalassospiramide A (**1**) standard. Extracted ion chromatograms (EIC) of m/z 958.5496. See **Figure 4.10** for associated MS/MS spectra.

4.3.2 Reconstitution of thalassospiramide structural diversity.

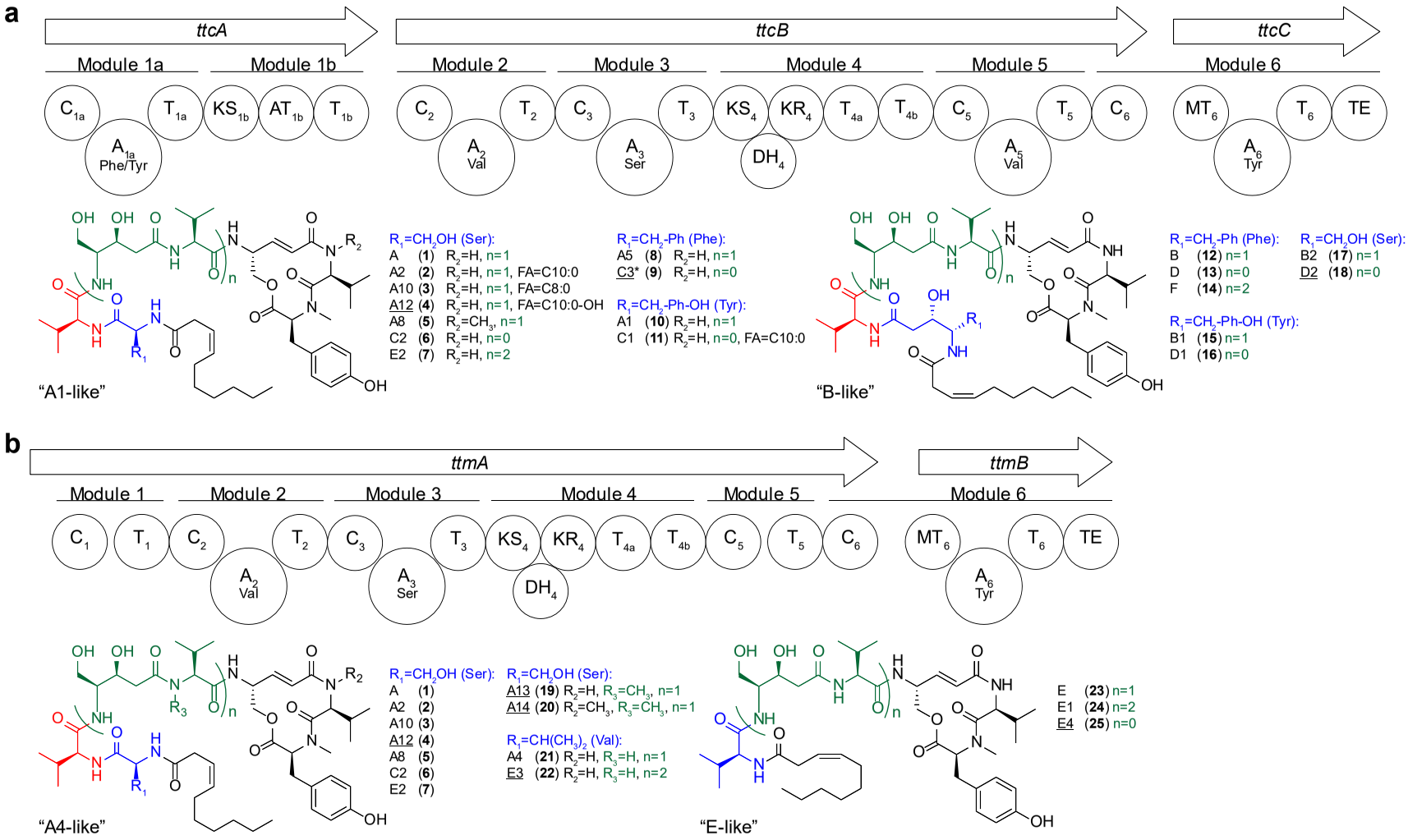
We next explored whether the heterologously expressed Ttc and Ttm could reproduce the full suite of thalassospiramide chemical diversity. Thalassospiramide cyclic lipodepsipeptides can be grouped into four categories based on their chemical structures and biosynthetic origin[11-13] (**Figure 4.3**). Products of Ttc can incorporate serine, phenylalanine, or tyrosine as the first amino acid residue (colored blue in **Figure 4.3**), which can be extended by a single ketide unit to generate “B-like” as opposed to “A1-like” thalassospiramides. Alternatively, Ttm incorporates serine or valine in the first position, then includes or omits a valine residue (colored red in **Figure 4.3**) to generate “A4-like” or “E-like” thalassospiramides, respectively. Both assembly lines

produce “A-like” thalassospiramides at greater relative abundance than their “B-like” or “E-like” counterparts.

Further elements that contribute to structural diversity include the N-terminal fatty acid, which is predominantly an atypical C10:1(Δ3) fatty acid, and the pattern of N-methylation, which is predominantly limited to the final tyrosine residue of the cyclic peptide core but can also extend to the adjacent valine residue for products of Ttc and further to valine within the linear peptide for Ttm (**Figure 4.3**). Finally, the number of linear Ser-C2-Val units (4-amino-3,5-dihydroxy-N-pentanyl-valine, colored green in **Figure 4.3**) can be 0, 1, or 2 for both assembly lines, which presumably arises from passage of chain-extension intermediates from module 4 back to module 2 or module 5 back to module 3. Based on the different possible combinations of these variables, each assembly line can theoretically generate well over 100 compound analogs; several dozen compounds are routinely detected from small 50 mL cultures of producing organisms using mass spectrometry.

LC-MS-MS analysis revealed that nearly all analogs detected from the native producers are also produced by the heterologous host, with some subtle differences in relative production levels (**Figure 4.11**). Overall titers of most analogs are comparable between native strains and host and in some cases greater in the host (**Figure 4.11**).

Figure 4.3 Ttc and Ttm assembly lines and structures of associated cyclic lipodepsipeptide products. (a) Ttc assembly line and structures of a representative set of associated chemical products. (b) Ttm assembly line and representative associated chemical products. Analogs not previously reported are underlined; see **Table 4.3** for HR-MS data. Analogs detected by LC-MS from the native producer but not the heterologous host are marked with an asterisk. C, condensation; A, adenylation; T, thiolation; KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; MT, methyltransferase; TE, thioesterase.


a
ttcA
ttcB
ttcC
Module 1a
Module 1b
Module 2
Module 3
Module 4
Module 5
Module 6
C1a T1a KS1b AT1b T1b
A1a Phe/Tyr
C2 T2 C3 T3 KS4 KR4 T4a T4b C5 T5 C6
A2 Val
A3 Ser
DH4
A5 Val
MT6 T6 TE
A6 Tyr
"A1-like"
R1=CH2OH (Ser):
A (1) R2=H, n=1
A2 (2) R2=H, n=1, FA=C10:0
A10 (3) R2=H, n=1, FA=C8:0
A12 (4) R2=H, n=1, FA=C10:0-OH
A8 (5) R2=CH3, n=1
C2 (6) R2=H, n=0
E2 (7) R2=H, n=2
R1=CH2-Ph (Phe):
A5 (8) R2=H, n=1
C3* (9) R2=H, n=0
R1=CH2-Ph-OH (Tyr):
A1 (10) R2=H, n=1
C1 (11) R2=H, n=0, FA=C10:0
"B-like"
R1=CH2-Ph (Phe):
B (12) n=1
D (13) n=0
F (14) n=2
R1=CH2-Ph-OH (Tyr):
B1 (15) n=1
D1 (16) n=0
R1=CH2OH (Ser):
B2 (17) n=1
D2 (18) n=0
b
ttmA
ttmB
Module 1
Module 2
Module 3
Module 4
Module 5
Module 6
C1 T1 C2 T2 C3 T3 KS4 KR4 T4a T4b C5 T5 C6
A2 Val
A3 Ser
DH4
MT6 T6 TE
A6 Tyr
"A4-like"
R1=CH2OH (Ser):
A (1)
A2 (2)
A10 (3)
A12 (4)
A8 (5)
C2 (6)
E2 (7)
R1=CH2OH (Ser):
A13 (19) R2=H, R3=CH3, n=1
A14 (20) R2=CH3, R3=CH3, n=1
R1=CH(CH3)2 (Val):
A4 (21) R2=H, R3=H, n=1
E3 (22) R2=H, R3=H, n=2
"E-like"
E (23) n=1
E1 (24) n=2
E4 (25) n=0

4.3.3 Characterization of non-assembly line genes.

To determine whether co-transferred genes beyond the core NRPS/PKS genes affect thalassospiramide titer or product distribution, we performed targeted deletions of all non-assembly line genes within *ttc*. Previous work using the broad-host-range vector, pCAP05[16], demonstrated that no upstream genes (*-7* through *-1*) are essential, although heterologous expression using this vector produced low yields (**Figure 4.12**) and proved to be genetically unstable over time. Targeted deletion of *ttc -1*, *+1*, *+2*, *+3*, and *+4* in the stable pCB-*ttc*-int expression construct had no impact on thalassospiramide production; however, deletion of the gene encoding the putative PPTase TtcD resulted in an approximately five-fold reduction in thalassospiramide A (**1**), which was restored upon genetic complementation of *ttcD* (**Figures 4.12-4.13**). This observation suggests that the single native *P. putida* PPTase is capable of activating Ttc carrier proteins, albeit not as effectively as TtcD. The *P. putida* PPTase is clearly capable of activating Ttm carrier proteins, as no cognate PPTase is encoded within *ttm*, and thalassospiramides are still produced in the heterologous host. However, co-expression of *ttcD* with *ttm* resulted in an approximately two-fold increase in levels of thalassospiramide A (**1**) (**Figures 4.12-4.13**). Quantitative analysis suggests that the PPTase TtcD favorably biases production of analogs that incorporate one or more linear Ser-C2-Val units (n≥1), particularly for Ttm, as *ttcD* co-expression actually decreases production of C2 (**6**) and E4 (**25**) (n=0) (**Figure 4.13**). This result suggests that TtcD-catalyzed phosphopantetheinylation of TtmA carrier proteins predisposes TtmA to perform pass-back chain extension at the expense of linear assembly through an unknown mechanism.

4.3.4 Inactivation and testing of assembly line domains.

Taken together, the results of the heterologous expression and gene deletion experiments strongly suggest that thalassospiramide structural diversity is generated directly from the multi-modular assembly line itself and does not involve accessory enzymes beyond what is provided

from primary metabolism. Although certain modules appear to be missing domains based on retro-biosynthetic analysis of thalassospiramide chemical structures, all necessary core and tailoring domains are present somewhere along the assembly line, including a DH domain in module 4 of both systems that was not previously annotated (**Figure 4.3, Figures 20-21**). Thus, we set out to investigate the predicted inter-module activity of assembly line domains.

Our initial approach leveraged gene deletion and complementation tools, focusing first on the smaller *ttcC* that encodes the last four domains of terminal module 6. Despite it harboring the lone assembly line MT domain, some thalassospiramides, such as A8 (**5**), are unusual in containing a 'misplaced' penultimate N-methylated valine residue in addition to the conserved terminal N-methylated tyrosine. Although MTs are usually positioned at the C-terminus of interrupted A domains[17], MT_6 is positioned at the N-terminus of A_6 (**Figure 4.17**). To explore whether MT_6 can methylate amino acids activated by both the A_5 and A_6 adenylation domains, we deleted *ttcC*, resulting in complete loss of thalassospiramide production. Complementation with wild-type *ttcC* restored thalassospiramide production, while complementation with a mutant encoding TtcC-G234D, in which MT_6 has been selectively inactivated, resulted in dramatic reduction and complete loss of thalassospiramide A (**1**) and A8 (**5**), respectively, and concomitant formation of a new product with HR-MS and MS/MS spectra consistent with desmethyl thalassospiramide A, or thalassospiramide A15 (**26**) (**Table 4.3, Figure 4.18**). The same result was observed for all analogs, resulting in the formation of a series of desmethyl cylic lipodepsipeptides. This confirms our hypothesis that MT_6 can act within the upstream module 5 of TtcB. Furthermore, it suggests that pass-back chain extension occurs between modules 4 and 2 as opposed to 5 and 3, as promiscuous methylation is confined to the cyclic valine residue for Ttc. If linear valine residues were installed by module 5, promiscuous methylation should extend to these positions; however, we never observe this for products of Ttc. As module 5 of Ttm does not possess an A domain and presumably borrows the activity of A_2, this could explain why promiscuous methylation can extend to upstream valine residues for Ttm but not Ttc.

We attempted to use the same approach to characterize other thalassospiramide domains; however, efforts to perform PCR mutagenesis in *ttcA*, which is over 6.5 kb, and *ttcB* and *ttmA*, which are both over 15.5 kb, were ultimately unsuccessful. Therefore, we established new methodology combining oligo recombineering with CRISPR-Cas9 counter selection for facile introduction of point mutations to large DNA constructs cloned into the pCB vector backbone (**Figure 4.19**). Using this method, we selectively inactivated a series of assembly line domains (**Figure 4.20**) to directly interrogate their role in thalassospiramide biosynthesis.

Consistent with our annotation of C_{1a} of TtcA as a starter C domain responsible for fatty acylation of the first amino acid residue[18], C_{1a} inactivation completely abolished production of all thalassospiramides (**Figure 4.4**). We did not detect any masses corresponding to core peptides lacking an N-terminal fatty acid, suggesting that the assembly line can only generate lipopeptide products. In contrast, inactivation of A_{1a} using two different point mutations, TtcA-G631D and TtcA-K972A, resulted in essentially complete loss of all thalassospiramides incorporating phenylalanine or tyrosine as the first residue but maintained or enhanced production of analogs incorporating serine in the first position (**Figure 4.4**). Inactivation of A_3 using the analogous lysine to alanine mutation (TtcB-K2045A) resulted in complete loss of thalassospiramide production, as did selective inactivation of T_{1a} (**Figure 4.4**). These results suggest that the serine residue adenylated by A_3 is directly loaded onto T_{1a} during biosynthesis of analogs that incorporate serine as the first amino acid residue.

We hypothesized that “B-like” thalassospiramides from Ttc arise through PKS module 1b and, correspondingly, that “A1-like” thalassospiramides arise through module 1b skipping. Consistent with that hypothesis, mutation of the active site cysteine of KS_{1b} to alanine resulted in complete loss of “B-like” analogs but maintained production of “A1-like” analogs, although some yields were slightly reduced (**Figure 4.4**). Surprisingly, although AT_{1b} inactivation (TtcA-S1728A) reduced production of “B-like” analogs, almost all could still be detected by LC-MS at ~8-24% of wild-type production levels (**Figure 4.4**). As the AT_{1b} mutation is expected to abolish enzymatic

activity and the assembly line has no other AT domains, we propose that the module 4-interacting *trans*-AT can partially complement AT_{1b}.

We previously hypothesized that the tandem T domains in module 4 might be important for determining whether chain-extension intermediates are passed forward or backward[12], although literature precedents indicated that multiple T domains do not change product identity but instead increase flux or yield[19]. We attempted to use the same oligo recombineering/CRISPR-Cas9 method to selectively inactivate T_{4a} and T_{4b}. However, CRISPR-Cas9 targeting did not result in oligo incorporation but instead recombination across the two very similar T domain sequences to generate an in-frame deletion of residues 3511 to 3596 in TtcB. The resultant mutant protein contained only a single chimeric T domain composed of 46% of the N-terminus of T_{4a} and 54% of the C-terminus of T_{4b} (**Figure 4.20**). Transfer of this construct to the heterologous host revealed that all thalassospiramide analogs could still be produced, including those incorporating linear Ser-C2-Val units arising from pass-back chain extension, albeit in decreased yields (**Figure 4.4**). This result proves that tandem T domains are not essential for bi-directional chain extension. Moreover, production of all analogs was affected equally, supporting the flux hypothesis.

Finally, we predicted that “E-like” thalassospiramides produced by Ttm arise from module 2 skipping, analogous to module 1b skipping in Ttc. Thus, if pass-back chain extension occurs through modules 5 and 3, C_2 inactivation in Ttm would preserve production of all “E-like” analogs, as C_2 is completely skipped in this model. While C_1 inactivation abolished production of all thalassospiramides, C_2 inactivation dramatically reduced and completely abolished production of thalassospiramides E (**23**) and E1 (**24**), respectively (**Figure 4.4**). This result confirms our hypothesis that most pass-back events occur between modules 4 and 2, as levels of “E-like” analogs were not preserved. Furthermore, we observed a more than 100-fold increase in levels of thalassospiramide E4 (**25**), suggesting that the inability to pass growing chains back via C_2 forces the assembly line to pass intermediates forward, resulting in enhanced production of the “premature” termination product E4 (**25**) (**Figure 4.4**). However, we also observed the formation

of two new compounds not previously detected with HR-MS and MS/MS spectra consistent with the structures of thalassospiramides E5 (**27**) and E6 (**28**), indicating that pass-back between modules 5 and 3 can occur but correlates with additional substrate dehydration by module 4 (**Figure 4.4, Table 4.3**).

a

		"A1-like"						"B-like"					
		A (**1**)	C2 (**6**)	E2 (**7**)	A5 (**8**)	A1 (**10**)	C1 (**11**)	B (**12**)	D (**13**)	F (**14**)	B1 (**15**)	B2 (**17**)	D2 (**18**)
Domain	AA1	Ser	Ser	Ser	Phe	Tyr	Tyr	Phe	Phe	Phe	Tyr	Ser	Ser
	n	1	0	2	1	1	0	1	0	2	1	1	0
C_{1a}	TtcA H154A	n.d. *	n.d. **	n.d. **	n.d. **	n.d. *	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **
$A_{1a\ Phe/Tyr}$	TtcA G631D	1.00	1.92 *	1.49 *	n.d. **	n.d. *	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	4.42 **	4.39 *
$A_{1a\ Phe/Tyr}$	TtcA K972A	1.29	1.85 *	1.39 *	0.01 **	0.01 *	n.d. **	0.01 **	0.02 **	n.d. **	0.02 **	4.26 **	4.17 **
T_{1a}	TtcA S1041A	n.d. *	n.d. **	n.d. **	n.d. **	n.d. *	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **
KS_{1b}	TtcA A1273G C1274A	0.90	0.69	1.46	0.61	0.63	0.36 *	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **
AT_{1b}	TtcA S1728A	0.65	1.12	0.93	1.04	0.84	1.06	0.08 **	0.17 **	0.24 **	0.10 **	0.15 **	0.11 **
$A_{3\ Ser}$	TtcB K2045A	n.d. *	n.d. **	n.d. **	n.d. **	n.d. *	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **	n.d. **
T_4	TtcB Δ3511-3596	0.65	0.71 *	0.56 *	0.96	1.09	0.64 *	0.73	0.66 *	0.93	0.87	0.93	0.75

b

		"E-like"		
		E (**23**)	E1 (**24**)	E4 (**25**)
Domain	AA_1	Val	Val	Val
	n	1	2	0
C_1	TtmA H163A	n.d. **	n.d. **	n.d. *
C_2	TtmA H700A	0.08 **	n.d. **	123 **

c

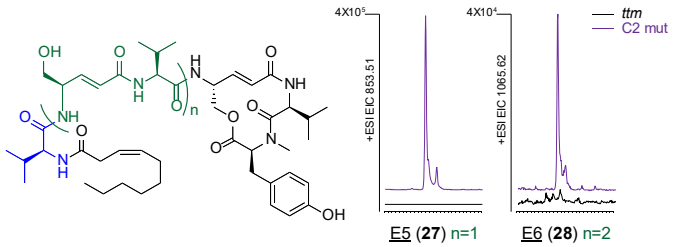


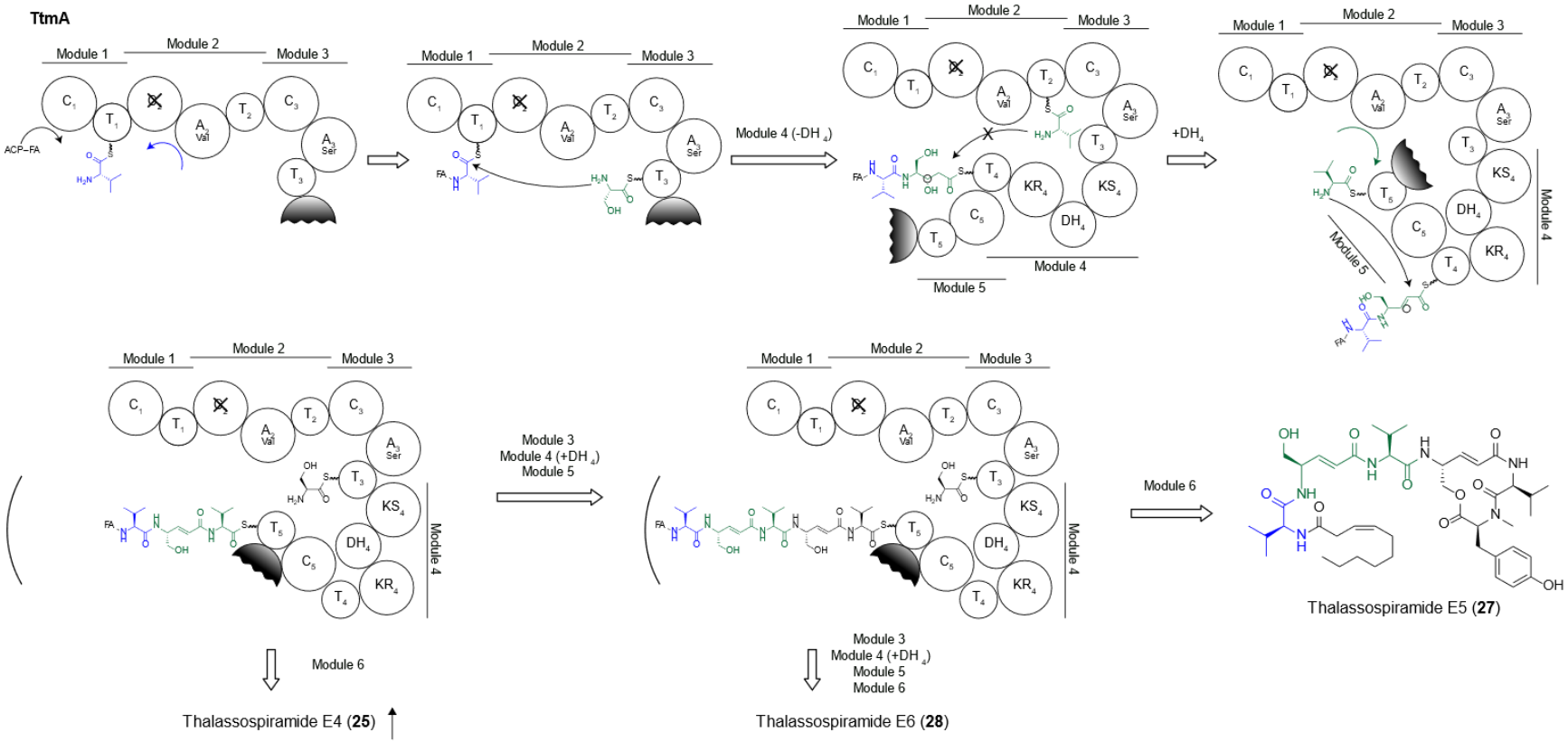
Figure 4.4 Selective inactivation of assembly line enzymatic domains alters product formation. Changes in production level of thalassospiramide analogs from (a) Ttc and (b) Ttm assembly lines upon selective inactivation of specific enzymatic domains. Domain and precise amino acid mutations are listed in the first two columns. Fold-change in MS ion intensity is indicated by number and color intensity, while white boxes indicate analogs were not detected (n.d.) from the mutant. Statistical significance was calculated using a two-tailed Student's t-test; n=3 biologically independent samples, *p<0.05, **p<0.005. (**c**) Structures and EICs of new thalassospiramide analogs produced upon TtmA C_2 inactivation; see **Table 4.3** for HR-MS data.

4.3.5 Models for thalassospiramide biosynthesis.

The results of Ttm C_2 inactivation provide a clear mechanistic insight into the flexibility and control of pass-back chain extension during thalassospiramide biosynthesis, as illustrated in **Figure 4.5**. Selective inactivation of C_2 does not affect early stages of "E-like" thalassospiramide

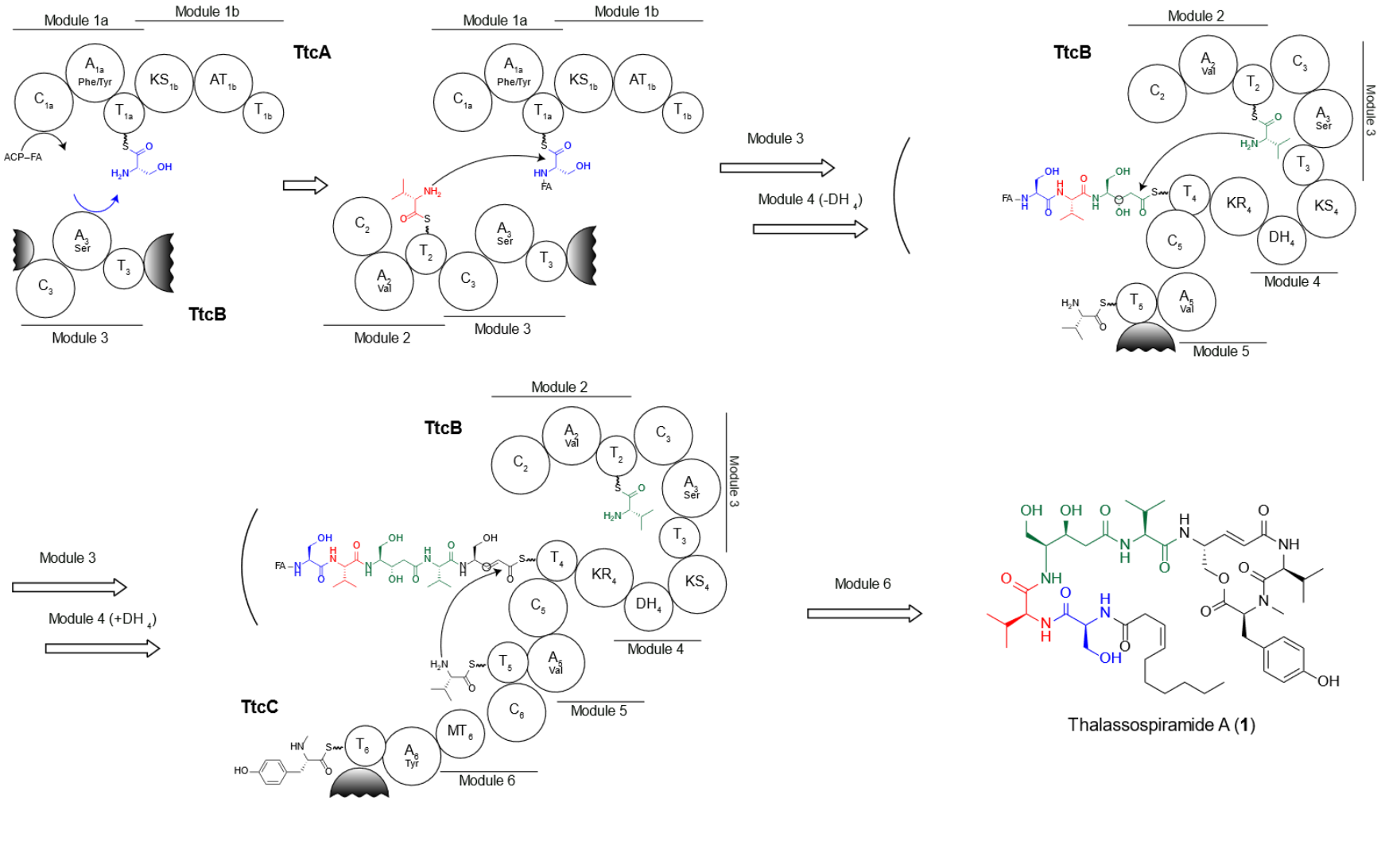
biosynthesis, during which A_2 loads T_1 with valine and then module 2 is skipped following appendage of the N-terminal fatty acid. Assembly then proceeds through modules 3 and 4, but the chain-extension intermediate does not undergo immediate dehydration and may not be directly accessible to DH_4. Now, the assembly line would normally pass the chain-extension intermediate from module 4 back to module 2 via C_2, which is favored based on product distribution, as E4 (**25**) is normally produced at very low abundance. We propose the assembly line C domains play an important role in "measuring" intermediate chain length, promoting donation of the module 4 intermediate backward to module 2 instead of forward to module 5. However, C_2 inactivation forces the chain-extension intermediate forward, making it accessible to DH_4 before it enters module 5. Once within module 5, the intermediate would normally proceed directly to module 6, resulting in formation of thalassospiramide E4 (**25**). Consistent with this proposal, C_2 inactivation drives a substantial increase in levels of E4 (**25**) compared to wild-type. However, chain length can also be "measured" at the donor site of C_6 (as C_6 usually accepts donor substrates of longer chain length), prompting the assembly line to catalyze pass-back of a subset of intermediates that have already undergone dehydration from module 5 back to module 3, resulting in eventual formation of new products E5 (**27**) and E6 (**28**). This result provides additional evidence that the substrate becomes accessible to DH_4 just as it is passed forward to module 5 and dehydration is a passive result of forward chain extension, since a model that invokes DH_4 gating of module 5 entry would not be consistent with the observed increase in levels of E4 (**25**) or formation of E5 (**27**) and E6 (**28**). Furthermore, it suggests that biosynthesis is flexible and controlled by a mechanism that measures intermediate chain length. If we assume that dehydration of the hydroxyl group within the linear unit is a signature for module 5 progression, there is evidence that pass-back between modules 5 and 3 occurs at low frequency under normal conditions, perhaps as an additional checkpoint, as we can detect thalassospiramides E7 (**29**), E8 (**30**), and E9 (**31**) by LC-MS (**Table 4.3**), which are all produced by wild-type Ttm and have undergone additional rounds of "premature" dehydration.

Figure 4.5 Model for thalassospiramide biosynthesis by Ttm C_2 inactivation mutant. Valine is loaded onto T_1 by A_2 and C_1 catalyzes addition of an activated fatty acid (FA) bound to an acyl carrier protein (ACP) or coenzyme A. Module 2 is skipped, and the fatty valine is extended directly onto serine-loaded T_3 via C_3. Chain extension proceeds normally to module 4, where the substrate is not immediately dehydrated and would normally be passed back to module 2 via C_2, perhaps based on chain length. However, C_2 inactivation forces the intermediate forward, where it becomes transiently accessible to DH_4 and is dehydrated before extension onto T_5, which is loaded with valine by A_2. Direct progression through module 6 results in formation of thalassospiramide E4 (**25**), which is substantially increased as a result of C_2 inactivation. Alternatively, passage from module 5 back to module 3 results in generation of new thalassospiramide analogs E5 (**27**) and E6 (**28**), which undergo one or two additional rounds of chain extension, respectively, through modules 3-5.


TtmA
Module 1
Module 2
Module 3
ACP–FA
Module 4 (-DH 4)
+DH4
Module 4
Module 5
Module 3
Module 4 (+DH 4)
Module 5
Module 6
Thalassospiramide E5 (27)
Thalassospiramide E4 (25)
Module 3
Module 4 (+DH 4)
Module 5
Module 6
Thalassospiramide E6 (28)

We can also propose a full model for thalassospiramide A (**1**) biosynthesis via Ttc (**Figure 4.6**). T_{1a} is preferentially adenylated with serine through the downstream A_3 domain. To our knowledge, this is the first report of an A domain activating a carrier protein within an upstream module that already possesses its own active A domain[12]. Subsequently, module 1b is skipped during formation of "A1-like" analogs. Normal chain extension proceeds from modules 2 to 4, at which point we hypothesize that the substrate is sequestered from DH_4 activity and module 5 entry based on chain length. We propose that during the formation of "B-like" thalassospiramides, ketoreduction to generate the upstream statine-like amino acid residue occurs at this time, analogous to dual ketoreduction of two disparate positions catalyzed by KR_3 of PksJ during bacillaene biosynthesis[20]. The chain-extension intermediate within module 4 is then passed back to module 2 via C_2, where it undergoes another round of linear chain extension through modules 2 to 4 (**Figure 4.6**). Now, the intermediate has reached "sufficient" chain length and becomes accessible to DH_4 before transfer to module 5 via C_5. MT_6 promiscuously methylates valine residues installed by module 5 to generate analogs such as thalassospiramide A8 (**5**). Finally, normal progression through modules 5 and 6 results in the formation of thalassospiramide A (**1**), the most abundant product of both Ttc and Ttm.

Figure 4.6 Model for thalassospiramide A biosynthesis by Ttc. Serine is loaded onto T1a by A3 and C1a catalyzes addition of an activated fatty acid (FA) bound to an ACP or coenzyme A. Module 1b is skipped, and the fatty serine is passed directly to valine-loaded T2 via C2. Chain extension proceeds normally to module 4, where the substrate is sequestered from DH4, perhaps within an enzyme binding pocket. The intermediate is passed from module 4 back to module 2 via C2, where it undergoes another round of chain extension through modules 3 and 4. Upon return to module 4, the intermediate is now acted upon by DH4, perhaps due to a conformational change associated with longer chain length, and extended forward through modules 5 and 6 to generate thalassospiramide A (**1**).


Module 1a
Module 1b
TtcA
TtcB
ACP–FA
Module 3
Module 2
Module 3
Module 3
Module 4 (-DH 4)
Module 2
Module 3
Module 4
Module 5
Module 3
Module 4 (+DH 4)
Module 2
TtcB
Module 3
Module 4
TtcC
Module 5
Module 6
Module 6
Thalassospiramide A (1)

4.4 Discussion

Modular assembly lines are large and dynamic enzymes that undergo dramatic domain conformational rearrangements during a single catalytic cycle[21-24]. However, whether multi-modular enzymes adopt a rigid, supermodular architecture[25] or a more flexible configuration[26] remains in debate. In this work, we demonstrate the ability of multi-modular assembly lines to catalyze bidirectional and nonlinear passage of chain-extension intermediates, favoring a more flexible arrangement.

Several new features of multi-modular NRPS/PKS biosynthesis are described in this work. Thalassospiramide assembly lines catalyze several instances of inter-module substrate activation and tailoring. While A domain supplementation has been previously reported[6-10], prior examples have been limited to upstream domains supplementing downstream modules, often encoded on separate proteins, and only within modules that lack their own A domains. For Ttc, A_3 adenylates T_{1a} with serine more frequently than A_{1a} does with phenylalanine or tyrosine, although A_{1a} is active and participates in biosynthesis of numerous thalassospiramide analogs. For Ttm, both A_2 and A_3 deliver substrates to T_1, but only A_3 adenylates T_5. Furthermore, only A_2 loads T_1 if module 2 is skipped, as all "E-like" analogs incorporate valine as the first amino acid, perhaps due to the specificity of C_3[27]. MT_6 promiscuously methylates valine residues activated by A_5 of Ttc and A_2 of Ttm. Finally, the *trans*-AT interacting with module 4 can partially supplement lost AT_{1b} activity in Ttc. Complementation of lost *cis*-AT activity with *cis*-ATs from other modules or non-cognate *trans*-ATs has been previously observed in the 6-deoxyerythronolide B PKS synthase[28].

Thalassospiramide assembly lines also catalyze programmed module skipping. Forward module skipping has been previously observed in PKS[29-31] and NRPS[32-34] systems, both naturally and as a byproduct of engineering. Ttc and Ttm catalyze analogous skipping of PKS module 1b and NRPS module 2, respectively, although NRPS module 2 does not sit at an enzyme junction but is embedded within a very large polypeptide. Perhaps as a result, skipping is favored in Ttc but disfavored in Ttm based on product distribution.

Finally, thalassospiramide assembly lines catalyze pass-back chain extension. To our knowledge, this mechanism has not been previously reported in modular assembly line systems, although it is analogous to "iteration" observed in the fungal beauvericin and bassianolide synthetases[35], where intermediates are passed back and forth between adjacent modules. Consistent with previous findings, tandem T domains in thalassospiramide assembly lines increase flux[19] but are not mechanistic determinants of pass-back chain extension. This result is also consistent with the observation that the thalassospiramide assembly line from *Oceanibaculum pacificum* contains only a single T domain in module 4 and can still produce thalassospiramide A (**1**)[13]. There is no evidence that additional T domains promote "stalling" to allow for additional tailoring reactions, as having a single T domain in module 4 does not change product distribution but decreased levels of all analogs equally.

Under normal conditions, thalassospiramide intermediates are passed from module 4 back to module 2 via C_2. Upon C_2 inactivation, we observe pass-back chain extension from module 5 back to module 3 via C_3. This suggests that the assembly line is flexible and can respond to perturbation, trying to "correct" for aberrant intermediate chain length caused by C_2 inactivation by passing back through C_3, which, like C_6, is specific for accepting intermediates with C-terminal valine residues. Our results are consistent with recent findings that multi-modular NRPSs can be "mixed and matched" if the specificity and relative position of downstream C domains are maintained[27]. It also suggests that assembly lines must possess some symmetry for pass-back chain extension to occur. While all NRPS modules possess mechanisms to control the timing of chain extension to prevent misinitiation[36,37], thalassospiramide modules have the added ability to accept longer intermediates from nonsequential modules.

Perhaps more importantly, how physical proximity between upstream C domains and downstream T domains is achieved during pass-back chain extension remains unclear. We cannot predict the plasticity of the hybrid thalassospiramide assembly line proteins based on their primary amino acid sequence alone (**Figures 4.14-4.17**). Although we can assume that TtcA,

TtcB, and TtmA are homodimeric due to the dimeric nature of KS and many linker domains, we do not know whether oligomerization impacts thalassospiramide assembly, although it is tempting to suggest that higher-order architecture helps facilitate nonlinear transfer.

It is curious that although the early stages of thalassospiramide biosynthesis are flexible, resulting in production of lipopeptides with a high degree of N-terminal structural diversity, the final stages are rather fixed. The C-terminal cyclic depsipeptide core of all thalassospiramide analogs is highly conserved, particularly the 12-membered ring and the α,β-unsaturated carbonyl moiety that together form the pharmacophore responsible for calpain protease inhibition[38]. Thus, the assembly line constructs a series of chemical products in which the structural elements that confer activity are maintained, while accessory elements such as the fatty acid and linear chain composition and length, which may confer target specificity, are variable. It has been speculated that biosynthetic promiscuity resulting in chemical diversity may be evolutionarily advantageous[39,40]. Furthermore, the type of combinatorial biosynthesis observed in this study expands the portfolio of small molecules produced without introducing new assembly line modules or domains, or even additional tailoring enzymes, thus representing a means of expanding chemical diversity while minimizing genomic space.

Our work was facilitated by new methodology combining oligo recombineering with CRISPR-Cas9 counter selection for facile editing of large DNA constructs. Although our method was used solely for domain inactivation in this study, it can be easily applied to perform other forms of multi-modular assembly line engineering, for example to alter A domain specificity[41,42]. Future efforts to understand the specific elements that enable and control assembly line flexibility will hopefully enhance efforts to engineer NRPS/PKS proteins and possibly expand their biosynthetic repertoire.

4.5 Methods

4.5.1 General methods

A complete list of the primers, plasmids, and strains used in this study can be found in **Tables 4.4-4.6**. DNA fragments larger than 3 kb were amplified with PrimeSTAR Max (Clontech Laboratories, Inc.); all other PCR products were amplified with PrimeSTAR HS DNA polymerase (Clontech Laboratories, Inc.). DNA isolations and manipulations were carried out using standard protocols. *Thalassospira* sp. CNJ-328 and *T. mobilis* KA081020-065 were grown in GYP media (glucose 10 g/L, yeast extract 4 g/L, peptone 2 g/L, sea salt 25 g/L). *S. cerevisiae* VL6-48N was grown in YPDA media (yeast extract 10 g/L, peptone 20 g/L, dextrose 20 g/L, adenine 100 mg/L) or selective histidine drop-out media containing 5-FOA (yeast nitrogen base without amino acids and ammonium sulfate 1.7 g/L, yeast synthetic dropout medium without histidine 1.9 g/L, sorbitol 182 g/L, dextrose 20 g/L, ammonium sulfate 5 g/L, adenine 100 mg/L, 5-FOA 1 g/L). *E. coli* and *P. putida* strains were grown in LB. *E. coli* TOP10 and DH5α λpir were used for standard cloning procedures. *E. coli* BW25113/pIJ790 was used for λ Red PCR targeting, and *E. coli* HME68 was used for oligo recombineering and CRISPR-Cas9 counter selection. *P. putida* EM383 was used for heterologous expression. All strains were grown at 30 °C except TOP10 and DH5α λpir, which were grown at 37 °C. Liquid cultures were grown shaking at 220 r.p.m. When necessary, *E. coli* (and *P. putida*) cultures were supplemented with the following antibiotics: 50 µg/mL kanamycin (150 µg/mL for *P. putida*), 10 µg/mL gentamycin (30 µg/ml for *P. putida*), 50 µg/mL apramycin, 100 µg/mL ampicillin, 25 µg/mL chloramphenicol.

4.5.2 Cloning and heterologous expression of *ttc* and *ttm*

The overall workflow for genetic manipulation and heterologous expression of the *ttc* and *ttm* pathways is outlined in **Figure 4.7**. Biosynthetic gene clusters were cloned from genomic DNA

using a TAR cloning protocol described previously[43]. Cluster-specific capture vectors were generated through a one-step PCR amplification of pCAP-BAC (pCB) using primers pCB-ttcCV_F/R for *ttc* and pCB-ttmCV_F/R for *ttm*. Yeast clones were screened, and PCR positive constructs were purified and transferred to *E. coli* TOP10 for verification by restriction digestion. pCB was miniprepped from at least 25 mL of VL6-48N and at least 10 mL of TOP10. Expression was attempted but never achieved in *E. coli* strains, including BL21(DE3) and BAP1[44] (**Figure 4.8**). pJZ001, containing the *intB13* cassette, was assembled by Gibson Assembly using four PCR fragments (amplified using primers CEN6/ARS4_608F/R, intB13_2330F/R, aacC1_1257F/R, and pADH_597F/R) and the pACYCDuet-1 vector backbone, linearized using HindIII and XhoI. Fully assembled pJZ001 was digested using HindIII and XhoI, and the 5155 bp fragment was gel purified and used for λ Red recombination[45] to knock-in the *intB13* cassette into the pCB vector backbone, replacing several yeast genes no longer necessary. Retrofitted constructs were then transferred to *P. putida* by electroporation as described previously[16] and selected for using kanamycin and gentamicin. IntB13-mediated integration of *ttc* into the genome of *P. putida* was characterized by PCR and chemical analysis (**Figure 4.9**). Edited constructs were similarly introduced to *P. putida* and then tested for heterologous production of lipopeptide products.

4.5.3 Extraction and LC-MS analysis

Precultures were inoculated with colonies picked from plates and grown overnight before being inoculated into full 50 mL cultures in 250 mL Erlenmeyer flasks (in triplicate). Full cultures were grown for 5 hours before addition of 1.5 g of autoclaved XAD7HP resin per 50 mL of culture. After 24 hours of additional incubation, culture ODs were measured and recorded at 600 nm and supernatant and cells were decanted. Resin was washed three times with Milli-Q water before extraction with 20 mL of ethyl acetate. Extracts were dried under nitrogen, resuspended in 200 µL of methanol, and filtered through a 0.22 µm filter prior to LC-MS-MS analysis.

An Agilent 1100 series HPLC system (Palo Alto, CA, U.S.A.) was coupled to a Bruker Impact II Q-TOF mass spectrometer (Billerica, MA, U.S.A.) for LC-MS analysis. An Agilent ZORBAX 300SB-C18 LC column (300 Å, 5 µm, 150 × 0.5 mm) was used for LC separation. Mobile phase A was H_2O in 0.1% FA and mobile phase B was ACN in 0.1% FA. The LC gradient was: t=0.00 min, 70%A; t=3.00 min, 70%A; t=5.00 min, 57%A; t=35.00 min, 57%A; t=49.00 min, 20%A; t=51.00 min, 0%A; t=52.00 min, 0%A; t=57.00 min, 70%A; t=60.00 min, 70%A. A post time of 10 min was set to re-equilibrate the column. For shorter runs, the LC gradient was: t=0.00 min, 70%A; t=3.00 min, 70%A; t=23.00 min, 20%A; t=24.00 min, 0%A; t=27.00 min, 70%A; t=30.00 min, 70%A. A post time of 3 min was set to re-equilibrate the column. Flow rate was 20 µL/min. Sample injection volume was 2 µL.

MS conditions for MS/MS spectra generation were set as follows: capillary voltage, 4500; nebulizer gas flow, 0.8 Bar; dry gas, 5.0 L/min at 180 ºC; funnel 1 RF 150 Vpp; funnel 2 RF, 300 Vpp; isCID energy, 0 eV; hexapole RF: 50 Vpp; Quadrupole ion energy, 4 eV; low mass 50 m/z; collision cell energy, 20 – 50 eV; pre pulse storage 5.0 µs; collision RF, ramp from 350 to 800 Vpp; transfer time ramp from 50 to 100 µs; detection mass range 25 to 1000 m/z; MS/MS spectra collection rate was 2.0 Hz.

All samples analyzed by comparison were run at the same time and under the same conditions. Values were normalized by culture ODs and compared only for peaks with identical MS spectra and retention time. HR-MS data for thalassospiramide analogs analyzed in this study are provided in **Table 4.3** and **Figures 4.21-4.34**.

4.5.4 Gene deletion and complementation experiments

Gene deletions were made using λ Red PCR targeting as described previously[46]. Primers used to amplify the *aac(3)IV* cassette and confirm gene deletions are listed in **Table 4.4**. Deletions using this cassette were made after addition of the *intB13* cassette, which contains a gentamycin resistance gene, as the apramycin resistance gene *aac(3)IV* confers resistance to gentamycin in

E. coli. For complementation, *ttcD* and *ttcC* were amplified using primers Tn7-ttcD_F/R and Tn7-ttcC_F/R, respectively, and cloned into the mini-Tn7 vector pUC18R6K-mini-Tn7T-Gm[47]. Cloned constructs were introduced to *P. putida* by electroporation[16] along with the helper plasmid pTNS1[47] and selected for using gentamycin. Complemented *P. putida* clones were then made electrocompetent and pCB constructs were transferred by electroporation and selected for using kanamycin and gentamycin. For complementation of TtcC-G234D (MT_6 inactivation), the mutation was generated by amplification of pTn7::*ttcC* using primers Tn7-ttcC-g702a_F/R and confirmed by sequencing. Although a *ttcA* deletion construct was generated and a mini-Tn7 vector containing *ttcA* was prepared, the latter ultimately could not be transferred to *P. putida* for complementation, as no clones were obtained even after multiple attempts, likely due to the large size of the gene (>6.5 kb).

4.5.5 Inactivation and testing of assembly line enzymatic domains

Assembly line domain active sites are shown in **Figures 4.14-4.17**. Motifs and active sites were identified by sequence alignment against annotated NRPS/PKS domains[22-24,48-55]. pJZ002, an ampicillin resistant version of the pCas9[56] vector, was constructed as follows. First, the BsaI restriction site was first removed from the ampicillin resistance gene *bla* via PCR amplification of pKD20 using primers pKD20-g848a_F/R. The resulting construct was PCR amplified using primers ts-repA101_F/R and combined with a fragment amplified from pCas9 using primers pCas9_5058F/R by Gibson assembly. Spacer sequences were cloned into pJZ002 as described previously[56]. Spacer sequences and targeting oligos used to target specific domains are listed in **Table 4.4**. Oligo recombination and CRISPR-Cas9 counter selection were performed as described previously[56,57], with several modifications. The general workflow is shown in **Figure 24**. pCB constructs were first transferred to *E. coli* HME68 by electroporation and selected for using kanamycin. A transformant was picked and grown at 30 °C to OD_{600} 0.4-0.5 and heat shocked for 15 minutes at 42 °C in a shaking water bath before being chilled on ice for 10 minutes. The cells

were pelleted and washed with ice-cold water before being resuspended in a small volume of ice-cold water. 100 ng of pJZ002 containing the appropriate spacer was mixed with 100 ng of targeting oligo and the DNA mixture was introduced to the cells prior to electroporation at 2.5 kV in a 2 mm gap electroporation cuvette. Cells were recovered for 2 hours at 30 °C shaking and plated on LB with kanamycin and ampicillin. Four clones of each mutant were picked, miniprepped, and screened by sequencing. If no correct mutant was identified, a new spacer sequence was designed and cloned into pJZ002 and the method was retried. Very subtle mutations could be recovered efficiently using an effective spacer sequence, although the effectiveness of the spacer could only be determined empirically. In total, four of 12 spacer sequences were redesigned to achieve successful editing (**Table 4.4**). Correctly edited constructs were transferred to TOP10 and confirmed by restriction digestion (**Figure 4.20**) before being transferred to *P. putida* for heterologous expression.

4.6 Data availability

The *ttc* and *ttm* biosynthetic gene cluster sequences are available in the MIBiG database (accession BGC0001050 and BGC0001876). Plasmids pCAP-BAC (#120229), pJZ001 (#120230), and pJZ002 (#120231) are available at Addgene.

4.7 Supplementary Information

4.7.1 Supplementary Tables

Table 4.1 Annotation and BLAST results of *ttc* biosynthetic pathway genes.

Gene	Size (aa)	Proposed function	Description (top BLAST hit)	Identity	NCBI accession
ttc -7	453	Regulation	Sigma-54-dependent Fis family transcriptional regulator [Thalassospira]	99%	WP_082824890.1
ttc -6	149	Hypothetical	TonB-dependent receptor [Thalassospira lucentensis]	99%	WP_062953563.1
ttc -5	478	Transport	TonB-dependent receptor [Thalassospira lucentensis]	99%	WP_062953563.1
ttc -4	270	Hypothetical	ferric iron reductase protein FhuF [Thalassospira xiamenensis]	99%	SOC29973.1
ttc -3	537	Transport	ABC transporter ATP-binding protein/permease [Thalassospira xiamenensis]	99%	WP_114109924.1
ttc -2	499	Other	methylmalonate-semialdehyde dehydrogenase (CoA acylating) [Thalassospira xiamenensis]	99%	WP_062959055.1
ttc -1	295	Regulation	LysR family transcriptional regulator [Thalassospira xiamenensis]	99%	WP_114109923.1
ttcA	2126	NRPS/PKS	hybrid non-ribosomal peptide synthetase/type I polyketide synthase [Thalassospira xiamenensis]	97%	WP_097053431.1
ttcB	5186	NRPS/PKS	non-ribosomal peptide synthetase [Thalassospira lucentensis]	96%	WP_062953557.1
ttcC	1174	NRPS/PKS	alpha/beta fold hydrolase [Thalassospira xiamenensis]	98%	WP_062959059.1
ttcD	253	NRPS/PKS PPTase	hypothetical protein [Thalassospira lucentensis]	95%	WP_062953555.1
ttc +1	186	Hypothetical	DUF697 domain-containing protein [Thalassospira sp. MCCC 1A03138]	100%	WP_085646013.1
ttc +2	420	Hypothetical	Prohibitin family protein [Thalassospira]	99%	WP_062953553.1
ttc +3	63	Hypothetical	hypothetical protein [Thalassospira]	100%	WP_062953552.1
ttc +4	79	Hypothetical	DUF697 domain-containing protein [Thalassospira]	100%	WP_085646011.1
ttc +5	260	Hypothetical	hypothetical protein [Thalassospira xiamenensis]	99%	WP_097053437.1
ttc +6	324	Hypothetical	MFS transporter [Thalassospira xiamenensis]	100%	WP_097053438.1

Table 4.2 Annotation and BLAST results of *ttm* biosynthetic pathway genes.

Gene	Size (aa)	Proposed function	Description (top BLAST hit)	Identity	NCBI accession
ttm-6	1220	Hypothetical	hypothetical protein [Erythrobacter sp. YT30]	48%	WP_067603228.1
ttm-5	423	Other	type II toxin-antitoxin system HipA family toxin [Roseomonas rosea]	86%	WP_073135195.1
ttm-4	81	Regulation	transcriptional regulator [Sandarakinorhabdus cyanobacteriorum]	90%	WP_094472265.1
ttm-3	707	Other	catalase [Catalinimonas alkaloidigena]	71%	WP_089678827.1
ttm-2	163	Other	CBS domain-containing protein [Mesorhizobium delmotii]	66%	SJM34906.1
ttm-1	242	Other	uracil-DNA glycosylase [Sphingomonas sp. PR090111-T3T-6A]	51%	WP_019834353.1
ttmA	5215	NRPS/PKS	TtbA [Tistrella bauzanensis]	63%	AGC65516.1
ttmB	1166	NRPS/PKS	TtbB [Tistrella bauzanensis]	69%	AGC65517.1
ttm+1	189	Other	HNH endonuclease [Defluviimonas alba]	67%	WP_084739999.1
ttm+2	104	Other	antitoxin of toxin-antitoxin stability system [Rhizobium sp. 60-20]	71%	OJY78535.1
ttm+3	82	Other	type II toxin-antitoxin system ParD family antitoxin [Neorhizobium galegae]	75%	WP_038540781.1
ttm+4	98	Other	type II toxin-antitoxin system RelE/ParE family toxin [Phyllobacteriaceae bacterium SYSU D60010]	48%	WP_119273623.1
ttm+5	594	Hypothetical	hypothetical protein AMS22_02705 [Thiotrichales bacterium SG8_50]	37%	KPK56010.1
ttm+6	304	Regulation	LysR family transcriptional regulator [Paraburkholderia nodosa]	69%	WP_051481188.1
ttm+7	324	Other	nitronate monooxygenase [Alcaligenes faecalis]	80%	WP_083053698.1
ttm+8	189	Other	cytochrome b561 [Alcaligenes faecalis]	71%	WP_026482775.1
ttm+9	395	Hypothetical	hypothetical protein A3D94_00820 [Alphaproteobacteria bacterium RIFCSPHIGHO2_12_FULL_66_14]	37%	OFX06623.1

Table 4.3 HR-ESI-MS data of thalassospiramide analogs analyzed in this chapter.

Compound	Exact [M+H]+	Observed [M+H]+	Molecular formula	Error (ppm)	Reference
A (**1**)	958.5496	958.5507	$C_{48}H_{75}N_7O_{13}$	1.15	[11, 12]
A2 (**2**)	960.5652	960.5676	$C_{48}H_{77}N_7O_{13}$	2.50	[12]
A10 (**3**)	932.5339	932.5348	$C_{46}H_{73}N_7O_{13}$	0.97	[13]
A12 (**4**)	976.5602	976.5604	$C_{48}H_{77}N_7O_{14}$	0.20	**Fig. 4.21**
A8 (**5**)	972.5652	972.5660	$C_{49}H_{77}N_7O_{13}$	0.82	[13]
C2 (**6**)	728.4229	728.4227	$C_{38}H_{57}N_5O_9$	-0.27	[13]
E2 (**7**)	1188.6762	1188.6768	$C_{58}H_{93}N_9O_{17}$	0.50	[13]
A5 (**8**)	1018.5860	1018.5870	$C_{54}H_{79}N_7O_{12}$	0.98	[12]
C3* (**9**)	788.4593	788.4575	$C_{44}H_{61}N_5O_8$	-2.28	**Fig. 4.22**
A1 (**10**)	1034.5809	1034.5826	$C_{54}H_{79}N_7O_{13}$	1.64	[12]
C1 (**11**)	806.4698	806.4708	$C_{44}H_{63}N_5O_9$	1.24	[12]
B (**12**)	1062.6122	1062.6137	$C_{56}H_{83}N_7O_{13}$	1.41	[11, 12]
D (**13**)	832.4855	832.4857	$C_{46}H_{65}N_5O_9$	0.24	[12]
F (**14**)	1292.7388	1292.7424	$C_{66}H_{101}N_9O_{17}$	2.78	[12]
B1 (**15**)	1078.6071	1078.6084	$C_{56}H_{83}N_7O_{14}$	1.21	[12]
D1 (**16**)	848.4804	848.4815	$C_{46}H_{65}N_5O_{10}$	1.30	[12]
B2 (**17**)	1002.5758	1002.5770	$C_{50}H_{79}N_7O_{14}$	1.20	[12]
D2 (**18**)	772.4491	722.4503	$C_{40}H_{61}N_5O_{10}$	1.66	**Fig. 4.23**
A13 (**19**)	972.5652	972.5665	$C_{49}H_{77}N_7O_{13}$	1.34	**Fig. 4.24**
A14 (**20**)	986.5808	986.5821	$C_{50}H_{79}N_7O_{13}$	1.32	**Fig. 4.25**
A4 (**21**)	970.5860	970.5879	$C_{50}H_{79}N_7O_{12}$	1.96	[12]
E3 (**22**)	1200.7126	1200.7143	$C_{60}H_{97}N_9O_{16}$	1.42	**Fig. 4.26**
E (**23**)	871.5176	871.5186	$C_{45}H_{70}N_6O_{11}$	1.15	[12]
E1 (**24**)	1101.6442	1101.6451	$C_{55}H_{88}N_8O_{15}$	0.82	[12]
E4 (**25**)	641.3909	641.3905	$C_{35}H_{52}N_4O_7$	-0.62	**Fig. 4.27**
A15 (**26**)	944.5339	944.5350	$C_{47}H_{73}N_7O_{13}$	1.16	**Fig. 4.28**
E5 (**27**)	853.5070	853.5064	$C_{45}H_{68}N_6O_{10}$	-0.70	**Fig. 4.29**
E6 (**28**)	1065.6231	1065.6241	$C_{55}H_{84}N_8O_{13}$	0.94	**Fig. 4.30**
E7 (**29**)	1170.6657	1170.6655	$C_{58}H_{91}N_9O_{16}$	-0.17	**Fig. 4.31**
E8 (**30**)	1182.7021	1182.7031	$C_{60}H_{95}N_9O_{15}$	0.85	**Fig. 4.32**
E9 (**31**)	1083.6336	1083.6352	$C_{55}H_{86}N_8O_{14}$	1.48	**Fig. 4.33-4.34**

Analogs colored in blue were also compared by retention time against authentic standards validated previously by NMR[11-13]. New compounds underlined; asterix* indicates analog was detected from native strain but not heterologous host.

Table 4.4 Primers used in this chapter.

Name	Sequence
pCB-ttcCV_F	ATTTCCCCGAAAAGTGCCACCTGGGTCCTTTTCATCACGTGCTATAAAAAATGTCGAAAGCTACATATAAGGAACGT
pCB-ttcCV_R	TTTGCCGATTGGGCCGATTGATGCCGTGGCGGAGCCGCCGTACCAAACGCCCACTATTTATACCATGGGAGGC
pCB-ttmCV_F	CCATGTCCGCAAAAGTCGTGTTGACGATAGAGCTGGAACCGCAACTGCGTATGTCGAAAGCTACATATAAGGAACGT
pCB-ttmCV_R	ATCCGCGGCGTCGGTGCCGGCCGCGACGGTTCCGCCAAGGTCGCCGAGGTCCACTATTTATACCATGGGAGGC
CEN6/ARS4_608F	ATTCGAGCTCGGCGCGCCTGCAGGTCGACAAGCTTTCTATCGCCTTCTTGACGAG
CEN6/ARS4_608R	TGGCATCCATAGACCCACACCCAGCAGAATTCGGCCGCTTGCCTGTAACTTACAC
intB13_2330F	AGTTACAGGCAAGCGGCCGAATTCTGCTGGGTGTG
intB13_2330R	TTGTCACAACGCCGCTTGTGCAGGCGGCGCATATC
aacC1_1257F	CGCCGCCTGCACAAGCGGCGTTGTGACAATTTAC
aacC1_1257R	ACGGATACGATGGAGCGTTAGCCATGAGGGTTTAG
pADH_597F	CTCATGTTTAACGAACTAAACCCTCATGGCTAACGCTCCATCGTATCCGTATTCC
pADH_597R	AATTTCGCAGCAGCGGTTTCTTTACCAGACTCGAGTTTCGCGCAAGACGATTGAC
spacer-seq_F	CTGAAGTATATTTTAGATGA
d-ttc(-1)_F	**TTGGGACCGCATTCGCATCTTTCTGGCCGTCGCACGACAG**GGTCGGGCTGGGAAGTTCC
d-ttc(-1)_R	**GATCGGCGACGGCGGCGACCCCGCGGGTGTCCTTGGTATC**GTAGGCTGGAGCTGCTTCG
d-ttc(-1)-check_F	CGCAAAGCAGGTCGATTATG
d-ttc(-1)-check_R	CGTGCAGATACAGCAATCAG
d-ttc(+1)_F	**CATCGAAGAGACTGAAGAAGTGGCCGTCGCGAAGGCGGAT**GGTCGGGCTGGGAAGTTCC
d-ttc(+1)_R	**TCGCACCTGCTGCCGGTTTGGCAGTGGTTTTGCCGCCTTT**GTAGGCTGGAGCTGCTTCG
d-ttc(+1)-check_F	GATGATCAATCAGGCCTTGG
d-ttc(+1)-check_R	GTGTTGCACGTTGCTACTTC
d-ttc(+2) _F	**CCTGCTGGTCGCTATTTTCGCGTTGGCGCCGCAGATTTTC**GGTCGGGCTGGGAAGTTCC

Table 4.4 Primers used in this chapter, continued.

Name	Sequence
d-ttc(+2) _R	**CGCCCCCGCTACCACTGGATTGATCAAAGACCGGCGCAAT**GTAGGCTGGAGCTGCTTCG
d-ttc(+2)-check_F	GCCACTTCTTCAGTCTCTTC
d-ttc(+2)-check_R	ATACACCCGACGACGAGATG
d-ttc(+3+4)_F	**GGGGATATTTGCCAATCTTGCCAAGGTGATGCCGGTTGTC**GGTCGGGCTGGGAAGTTCC
d-ttc(+3+4)_R	**AATCAAGGTCTTTAAGCTTTTCACGAAGCTCCCGCGACGG**GTAGGCTGGAGCTGCTTCG
d-ttc(+3+4)-check_F	ATGCTGATCGCCCGTCAGTC
d-ttc(+3+4)-check_R	TCCGTGATCGCGGCCTTAAC
d-ttcD_F	**CCCGAAGGGCACCGGCACATTTGTGTCGTCCGATATCCTG**GGTCGGGCTGGGAAGTTCC
d-ttcD_R	**GATCCGGCAATGTCAGGCCAAGCCACCCCTGCGCACCGGG**GTAGGCTGGAGCTGCTTCG
d-ttcD-check_F	CGAAGGCGATCAATATAGCG
d-ttcD-check_R	GCGTCCGAATTTGTCAAAGG
Tn7-ttcD_F	TTTGAAGCTAATTCGATCATGCATGAGCTCATAGCGGCTGGCAAAGGGTC
Tn7-ttcD_R	GGTTGGCCTGCAAGGCCTTCGCGAGGTACCGCAAGGCGTCCGAATTTGTC
d-ttcC_F	**CATCGCAACCCGCGATGAAACCGGCGAAATCAGCTATACT**GGTCGGGCTGGGAAGTTCC
d-ttcC_R	**TCCGGCAAACCTGCCCTGGGCCTGATTTTTCAGTCCATCA**GTAGGCTGGAGCTGCTTCG
d-ttcC-check_F	GAATGGCACGCCCGATGTTC
d-ttcC-check_R	GGTTCGATCCGGCAAACCTG
Tn7-ttcC_F	TTTGAAGCTAATTCGATCATGCATGAGCTCTCTGTTCCTGACCGAAGAAG
Tn7-ttcC_R	GGTTGGCCTGCAAGGCCTTCGCGAGGTACCTGCCAATTTCGGGCAGTATC
Tn7-ttcC-g702a_F	TGCCCGGGTGCTTGAAATCG**A**CTGCGCGTCCGGCTTCACCA
Tn7-ttcC-g702a_R	TGGTGAAGCCGGACGCGCAG**T**CGATTTCAAGCACCCGGGCA
ttcC-g702a-check_F	AGTGCCTTTACCGGGCAACC
ttcC-g702a-check_R	CTGCCGACCGATGACATGTG

Table 4.4 Primers used in this chapter, continued.

Name	Sequence
d-ttcA_F	**ATTTCAAAAACCGCTTGCTGCATGGTGCAATCAACGTTTG**GGTCGGGCTGGGAAGTTCC
d-ttcA_R	**TGGTAGGCGCGCTGTTTTTATCAAAGGATGCAATCAGGTT**GTAGGCTGGAGCTGCTTCG
d-ttcA-check_F	AGGCACCTAAACAGCGTACC
d-ttcA-check_R	TCCCTGTTCGCGTGTCGAAG
Tn7-ttcA_F	TTTGAAGCTAATTCGATCATGCATGAGCTCCTGTTCAGGCACCTAAACAG
Tn7-ttcA_R	GGTTGGCCTGCAAGGCCTTCGCGAGGTACCCATGCAGGCTTAGCACAATG
pKD20-g848a_F	TCTGGAGCCGGTGAGCGTGG**A**TCTCGCGGTATCATTGCAGC
pKD20-g848a_R	GCTGCAATGATACCGCGAGA**T**CCACGCTCACCGGCTCCAGA
bla*-check_214F	GATGGAGGCGGATAAAGTTG
bla*-check_214R	ACCTATCTCAGCGATCTGTC
ts-repA101_F	**CAAAAGAAGAGTAGTGTGATCGTCCATTCC**TAAGCTAGCCCATGGGTATGGA
ts-repA101_R	**TATCATTCTACATTTAGGCGCTGCCATCTT**CAGGTGGCACTTTTCGGGGA
pCas9_5058F	AAGATGGCAGCGCCTAAATG
pCas9_5058R	GGAATGGACGATCACACTAC
ttcA-C1a_H154A_F	CAAGGGACGTTTCTGGTGGGTCCGGGTGTATCA**CGCA**CT**C**GTCTGTGACGGATATGCAGGCCATCTGATG
ttcA-C1a-spacer_F	AAACGTTTCTGGTGGGTCCGGGTGTATCATCATCG
ttcA-C1a-spacer_R	AAAACGATGATGATACACCCGGACCCACCAGAAAC
ttcA-C1a-check_232F	GGTTCTGATCGCCCATAACG
ttcA-C1a-check_232R	GAATGTGCTTTCGGGCACTG
ttcA-A1a_G631D_F	GTGACGATATCGCCTTTGTCTTCCATACGTCGG**ACAGC**ACCGGGCAACCAAAACCGGTTCCCGTTCATCA
ttcA-A1a(G)-spacer_F	AAACTGACGATATCGCCTTTGTCTTCCATACGTCG
ttcA-A1a(G)-spacer_R	AAAACGACGTATGGAAGACAAAGGCGATATCGTCA
ttcA-A1a(G)-check_380F	GATTGCCGACTGGTTCTGAG
ttcA-A1a(G)-check_380R	TGGGTCATCAGCCACATACG

Table 4.4 Primers used in this chapter, continued.

Name	Sequence
ttcA-A1a_K972A_F	TGGGTGGAAACGCTTCCTTTGCTACCGTC**C**GG**GGCG**AT**A**GA**C**CGCAAGGCAC TTGCCCAATTTGCTCAGG
ttcA-A1a(K)-spacer_F	AAACGGCAAGTGCCTTGCGATCGATTTTGCCAGAG
ttcA-A1a(K)-spacer_R	AAAACTCTGGCAAAATCGATCGCAAGGCACTTGCC
ttcA-A1a(K)-check_235F	GACGGAAGCCATTCGAAACG
ttcA-A1a(K)-check_235R	TGCGCATCTGATCGGGTTTG
ttcA-T1a_S1041A_F	TCGATACCAACCTGTTTGAAGCTGGCGCCCA**CG**C**CC**T**C**CTGGTACCGCGTGC ACAGTTTGCCCTGTCAAA
ttcA-T1a-spacer_F	AAACGCAAACTGTGCACGCGGTACCAGCAATGAAG
ttcA-T1a-spacer_R	AAAACTTCATTGCTGGTACCGCGTGCACAGTTTGC
ttcA-T1a-check_297F	CAAACAGCCTGTTGCACAAG
ttcA-T1a-check_297R	TCGGGAACTGTCTGGTTATC
ttcA-KS1b_A1273G C1274A_F	CCTGACCGGCCCAGCCGTCGCGTCCTCGACCG**GTGCG**TCGACCGGCCTTGT CAATATTGCGCTTGCCGTC
ttcA-KS1b-spacer_F	AAACCAAGCGCAATATTGACAAGGCCGGTCGAACG
ttcA-KS1b-spacer_R	AAAACGTTCGACCGGCCTTGTCAATATTGCGCTTG
ttcA-KS1b-check_293F	TTGGTTTCCCGACCTATCTG
ttcA-KS1b-check_293R	GCCCTGCGTGAAGTAATATC
ttcA-AT1b_S1728A_F	CGGCATTTCACCCGCCGCGCTGGCCGGGCAC**GCA**ATTGGCGAATATGTCGCA GCCTGCATTGGCGGGGTC
ttcA-AT1b-spacer_F	AAACTTCACCCGCCGCGCTGGCCGGGCACAGCATG
ttcA-AT1b-spacer_R	AAAACATGCTGTGCCCGGCCAGCGCGGCGGGTGAA
ttcA-AT1b-spacer2_F	AAACAGGCTGCGACATATTCGCCAATGCTGTGCCG

Table 4.4 Primers used in this chapter, continued.

Name	Sequence
ttcA-AT1b-spacer2_R	AAAACGGCACAGCATTGGCGAATATGTCGCAGCCT
ttcA-AT1b-check_283F	AACATCAAGCCGCATGGCAC
ttcA-AT1b-check_283R	AAAGGCCGACAGCAGATCCC
ttcB-A3_K2045A_F	CCCTTGATCATCTCCCGCTTACCAGCTCAGGTGCAGTGGACCGAAAGGCCCTTTCGGGCACACCGATGGC
ttcB-A3-spacer_F	AAACTTGATCATCTCCCGCTTACCAGCAGCGGCAG
ttcB-A3-spacer_R	AAAACTGCCGCTGCTGGTAAGCGGGAGATGATCAA
ttcB-A3-spacer2_F	AAACAAAGGGCCTTTCGGTCGACCTTGCCGCTGCG
ttcB-A3-spacer2_R	AAAACGCAGCGGCAAGGTCGACCGAAAGGCCCTTT
ttcB-A3-check_208F	AGGCGACCTGACCATTGGTG
ttcB-A3-check_208R	ATCGGGCTTTGCCGCGATAG
ttcB-T4a_S3509V_F	CTCGGATGATGATTTCTTTGCTCTTGGCGGTGACGTTATTACAGGCATGCAGATCGTCGACCGCATCAAT
ttcB-T4a-spacer_F	AAACTGATGCGGTCGACGATCTGCATGCCGGTTAG
ttcB-T4a-spacer_R	AAAACTAACCGGCATGCAGATCGTCGACCGCATCA
ttcB-T4b_S3595V_F	CGGACGAGGATTTCTATGCTTTGGGCGGTGACGTTATTACAGGCATGCAGATCGTTGATCGCATGAATGC
ttcB-T4b-spacer_F	AAACTCATGCGATCAACGATCTGCATGCCGGTGAG
ttcB-T4b-spacer_R	AAAACTCACCGGCATGCAGATCGTTGATCGCATGA
ttcB-T4-check_566F	GTTTCCGCACCGGTATCTCC
ttcB-T4-check_566R	GGCGATGGTGTCGCTGTTTC
ttmA-C1_H163A_F	GCCCGATCGCCATCGCTGGATCCGCTGCTATCACGCCCCTCATCCTGGATGGTCAGGGTGGCATGATCCTG
ttmA-C1-spacer_F	AAACGCTGGATCCGCTGCTATCATCATCTGATCCG

Table 4.4 Primers used in this chapter, continued.

Name	Sequence
ttmA-C1-spacer_R	AAAACGGATCAGATGATGATAGCAGCGGATCCAGC
ttmA-C1-spacer2_F	AAACTGACCATCCAGGATCAGATGATGATAGCAGG
ttmA-C1-spacer2_R	AAAACCTGCTATCATCATCTGATCCTGGATGGTCA
ttmA-C1-check-194F	TTGCCCAAGAGGCCTTCGAG
ttmA-C1-check-194R	GATGAGCGCGGTGTAGATCC
ttmA-C2_H700A_F	GGACCAACCCGAGCGGTTGCGGGTGGTGGTCGAT**GCC**CTGGTCTTCGACGGCGAAAGCCGCACGGTGTTC
ttmA-C2-spacer_F	AAACCCGAGCGGTTGCGGGTGGTGGTCGATCATCG
ttmA-C2-spacer_R	AAAACGATGATCGACCACCACCCGCAACCGCTCGG
ttmA-C2-spacer2_F	AAACGCGGGTGGTGGTCGATCATCTGGTCTTCGAG
ttmA-C2-spacer2_R	AAAACTCGAAGACCAGATGATCGACCACCACCCGC
ttmA-C2-check_435F	ATGACGATGCGGCGGAAATG
ttmA-C2-check_435R	ACAGGGCGGTGGTTGCAATG
ttmB-MT6_G230D_F	TCACGCCCGCCAGCCGGGTGCTGGAAATCG**AT**TGCGCCTCGGGCTTCACGCTCCGTGCCCTGGCGCCGCT
ttmB-MT6-spacer_F	AAACCCAGCCGGGTGCTGGAAATCGGCTGCGCCTG
ttmB-MT6-spacer_R	AAAACAGGCGCAGCCGATTTCCAGCACCCGGCTGG
ttmB-MT6-check_555F	ATATCGCCCATCTCGTCTGC
ttmB-MT6-check_555R	CGCGGATATTGCCCAGAAAG

Table 4.5 Plasmids used in this chapter.

Name	Description	Ref.
pCAP-BAC (pCB)	Yeast-*E. coli* artificial chromosome shuttle vector; HIS3 and Kan^R; **[Addgene ID: 120229]**	This chapter
pACR11	pCAP01 carrying *ttc*	This chapter
pRMR6K-Gm	R6K replicon, Pcirc-lacO4-*intB13*, *oriT*, FRT, MCS, Gm^R	[14]
pJZ001	pACYCDuet-1 carrying *intB13* cassette for pCB knock-in; **[Addgene ID: 120230]**	This chapter
pCas9	*cas9* and *tracr* vector; Chl^R	[56]
pKD20	*repA101*(ts)-*araBp-gam-bet-exo-ori101-bla*; Amp^R	[45]
pJZ002	*repA101(ts)-cas9-tracr-bla** (lacking BsaI site)*;* Amp^R; **[Addgene ID: 120231]**	This chapter
pCAP05-*ttc*	pCAP05 carrying *ttc* (*-1* through *+6*)	This chapter
pCB-*ttc*	pCB carrying *ttc*	This chapter
pACR12	pETDuet-1 carrying *ttcA-D* and *PL2TA16_02049* (PPTase) from *P. luteoviolacea* 2ta16	This chapter
pCB-*ttm*	pCB carrying *ttm* (*-1* through *+1*)	This chapter
pCB-*ttc*-int	pCB-*ttc* with IntB13	This chapter
pCB-*ttm*-int	pCB-*ttm* with IntB13	This chapter
pCB-*ttc*-int-Δ(-*1*)	pCB-*ttc*-int Δ(*ttc -1*)	This chapter
pCB-*ttc*-int-Δ(*+1*)	pCB-*ttc*-int Δ(*ttc +1*)	This chapter
pCB-*ttc*-int-Δ(*+2*)	pCB-*ttc*-int Δ(*ttc +2*)	This chapter
pCB-*ttc*-int-Δ(*+3+4*)	pCB-*ttc*-int Δ(*ttc +3*)(*ttc +4*)	This chapter
pCB-*ttc*-int-Δ(*ttcD*)	pCB-*ttc*-int Δ(*ttcD*)	This chapter
pTNS1	R6K replicon; encodes the *tnsABC+D* specific transposition pathway; Amp^R	[47]
pUC18R6K-mini-Tn7T-Gm	R6K replicon; mini-Tn7 vector; Amp^R, Gm^R	[47]
pTn7::*ttcD*	pUC18R6K-mini-Tn7T-Gm carrying *ttcD*	This chapter
pCB-*ttc-int*-Δ(*ttcC*)	pCB-*ttc*-int Δ(*ttcC*)	This chapter
pTn7::*ttcC*	pUC18R6K-mini-Tn7T-Gm carrying *ttcC*	This chapter
pTn7::*ttcC*-g702a	pUC18R6K-mini-Tn7T-Gm carrying *ttcC*-g702a (G234D)	This chapter
pJZ002::*vioX*	pJZ002 with *vioX* spacer	This chapter
pJZ002::C1a	pJZ002 with C1a spacer	This chapter

Table 4.5 Plasmids used in this chapter, continued.

Name	Description	Ref.
pCB-*ttc*-int-C1a 1	pCB-*ttc*-int with TtcA-H154A clone 1	This chapter
pCB-*ttc*-int-C1a 2	pCB-*ttc*-int with TtcA-H154A clone 2	This chapter
pJZ002::A1a(G)	pJZ002 with A1a(G) spacer	This chapter
pCB-*ttc*-int-A1a (G631D)	pCB-*ttc*-int with TtcA-G631D	This chapter
pJZ002:A1a(K)	pJZ002 with A1a(K) spacer	This chapter
pCB-*ttc*-int-A1a (K972A) 1	pCB-*ttc*-int with TtcA-K972A clone 1	This chapter
pCB-*ttc*-int-A1a (K972A) 2	pCB-*ttc*-int with TtcA-K972A clone 2	This chapter
pJZ002::T1a	pJZ002 with T1a spacer	This chapter
pCB-*ttc*-int-T1a	pCB-*ttc*-int with TtcA-S1041A	This chapter
pJZ002::KS1b	pJZ002 with KS1b spacer	This chapter
pCB-*ttc*-int-KS1b	pCB-*ttc*-int with TtcA-A1273G-C1274A	This chapter
pJZ002:AT1b	pJZ002 with AT1b spacer 1	This chapter
pJZ002:AT1b-2	pJZ002 with AT1b spacer 2	This chapter
pCB-*ttc*-int-AT1b	pCB-*ttc*-int with TtcA-S1728A	This chapter
pJZ002:A3	pJZ002 with A3 spacer 1	This chapter
pJZ002:A3-2	pJZ002 with A3 spacer 1	This chapter
pCB-*ttc*-int-A3	pCB-*ttc*-int with TtcB-K2045A	This chapter
pJZ002:T4a	pJZ002 with T4a spacer	This chapter
pCB-*ttc*-int-T4 A	pCB-*ttc*-int with TtcB-Δ3511-3596 version A	This chapter
pJZ002:T4b	pJZ002 with T4b spacer	This chapter
pCB-*ttc*-int-T4 B	pCB-*ttc*-int with TtcB-Δ3511-3596 version B	This chapter
pJZ002::C1	pJZ002 with C1 spacer 1	This chapter
pJZ002::C1-2	pJZ002 with C1 spacer 2	This chapter
pCB-*ttm*-int-C1 1	pCB-*ttm*-int with TtmA-H163A clone 1	This chapter
pCB-*ttm*-int-C1 2	pCB-*ttm*-int with TtmA-H163A clone 2	This chapter
pJZ002::C2	pJZ002 with C2 spacer 1	This chapter
pJZ002::C2-2	pJZ002 with C2 spacer 2	This chapter

Table 4.5 Plasmids used in this chapter, continued.

Name	Description	Ref.
pCB-*ttm*-int-C21	pCB-*ttm*-int with TtmA-H700A clone 1	This chapter
pCB-*ttm*-int-C22	pCB-*ttm*-int with TtmA-H700A clone 2	This chapter
pJZ002::MT6	pJZ002 with MT6 spacer	This chapter
pCB-*ttm*-int-MT6	pCB-*ttm*-int with TtmB-G230D	This chapter

Table 4.6 Strains used in this chapter.

Name	Description	Ref.
S. cerevisiae VL6-48N	*MATα trp1-Δ1 ura3-Δ1 ade2-101 his3-Δ200 lys2 met14 cir°*	[43]
E. coli TOP10	*mcrA, Δ(mrr-hsdRMS-mcrBC), Phi80lacZ(del)M15, ΔlacX74, deoR, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(SmR), endA1, nupG*	N/A
E. coli DH5α λpir	*sup E44, ΔlacU169 (ΦlacZΔM15), recA1, endA1, hsdR17, thi-1, gyrA96, relA1, λpir*	N/A
E. coli BW25113/ pIJ790	Strain: Δ(araD-araB)567, ΔlacZ4787(::rrnB-4), lacIp-4000(lacIQ), λ-, rpoS369(Am), rph-1, Δ(rhaD-rhaB)568, hsdR514 Plasmid: [*oriR101*], [*repA101*(ts)], *araBp-gam-be-exo*	[46]
E. coli HME68	W3110 *galKtyr145UAG ΔlacU169 [λ cI857 Δ(cro-bioA)] mutS<>cat*	[57]
P. putida EM383	KT2440 derivative; Δprophage1 Δprophage4 Δprophage3 Δprophage2 Δ*Tn7* Δ*endA-1* Δ*endA-2* Δ*hsdRMS* Δflagellum Δ*Tn4652* Δ*recA*	[15]

4.7.2 Supplementary Figures

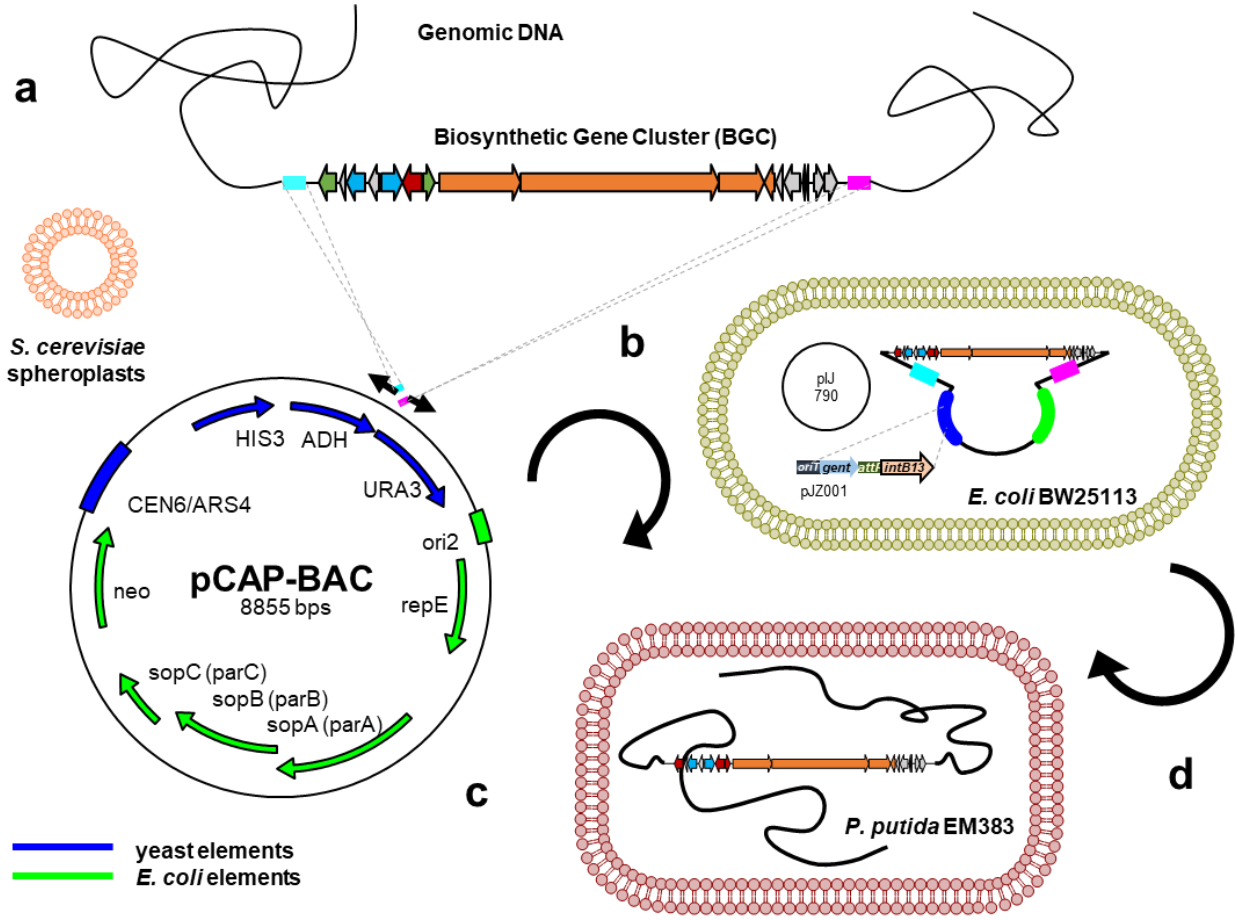


Figure 4.7 Platform for heterologous expression of thalassospiramide biosynthetic gene clusters. a) Direct cloning of gene clusters from genomic DNA was facilitated by transformation-associated recombination (TAR) cloning in *Saccharomyces cerevisiae* using new vector pCAP-BAC (pCB, see **Table 4.5**), which can be assembled into a cluster-specific capture vector by a one-step PCR reaction using long primers with homology arms flanking the genomic loci of interest. b) Individual clones that have been picked and screened are first transferred to *E. coli* TOP10 for maintenance and confirmation. Then, recombineering strain *E. coli* BW25113/pIJ790 is used to introduce the *intB13* integration cassette (from pJZ001, see **Table 4.5**) into the pCB vector backbone via Lambda Red recombination. c) Constructs retrofitted with IntB13 are transferred and integrated into the genome of the heterologous host *P. putida* EM383. d) Cloned constructs can be readily manipulated using *E. coli* recombineering tools and reintroduced to the heterologous host to rapidly connect genotype to chemical phenotype.

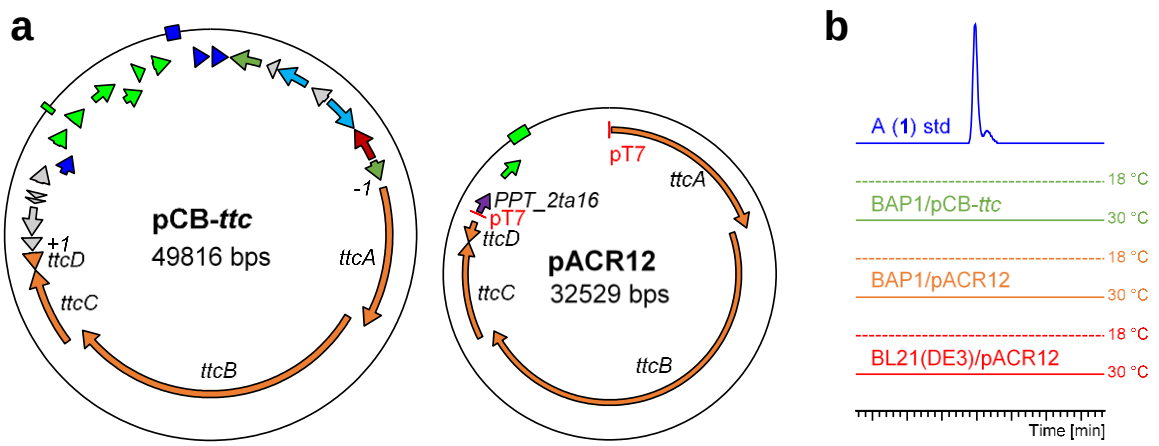


Figure 4.8 Failed heterologous expression in *E. coli*. a) Vector maps of pCB-*ttc* and pACR12, in which *ttcA-C* and PL2TA16_02049, a 4'-phosphopantetheinyl transferase (PPTase) from *P. luteoviolacea* 2ta16, have been cloned into pETDuet-1 and placed under control of the T7 promoter (pT7). b) Extracted ion chromatograms (EICs) of thalassospiramide A (**1**) from extracts of *E. coli* BL21(DE3) and *E. coli* BAP1 harboring pCB-*ttc* and pACR12. Cultures were grown at both 30 °C and 18 °C post induction with IPTG to try to stabilize protein expression. BAP1 possesses the Sfp PPTase from *Bacillus subtilis*. Production of (**1**) was not observed from heterologous expression of *ttc* in *E. coli*. Experiment was repeated 3 times independently with the same results.

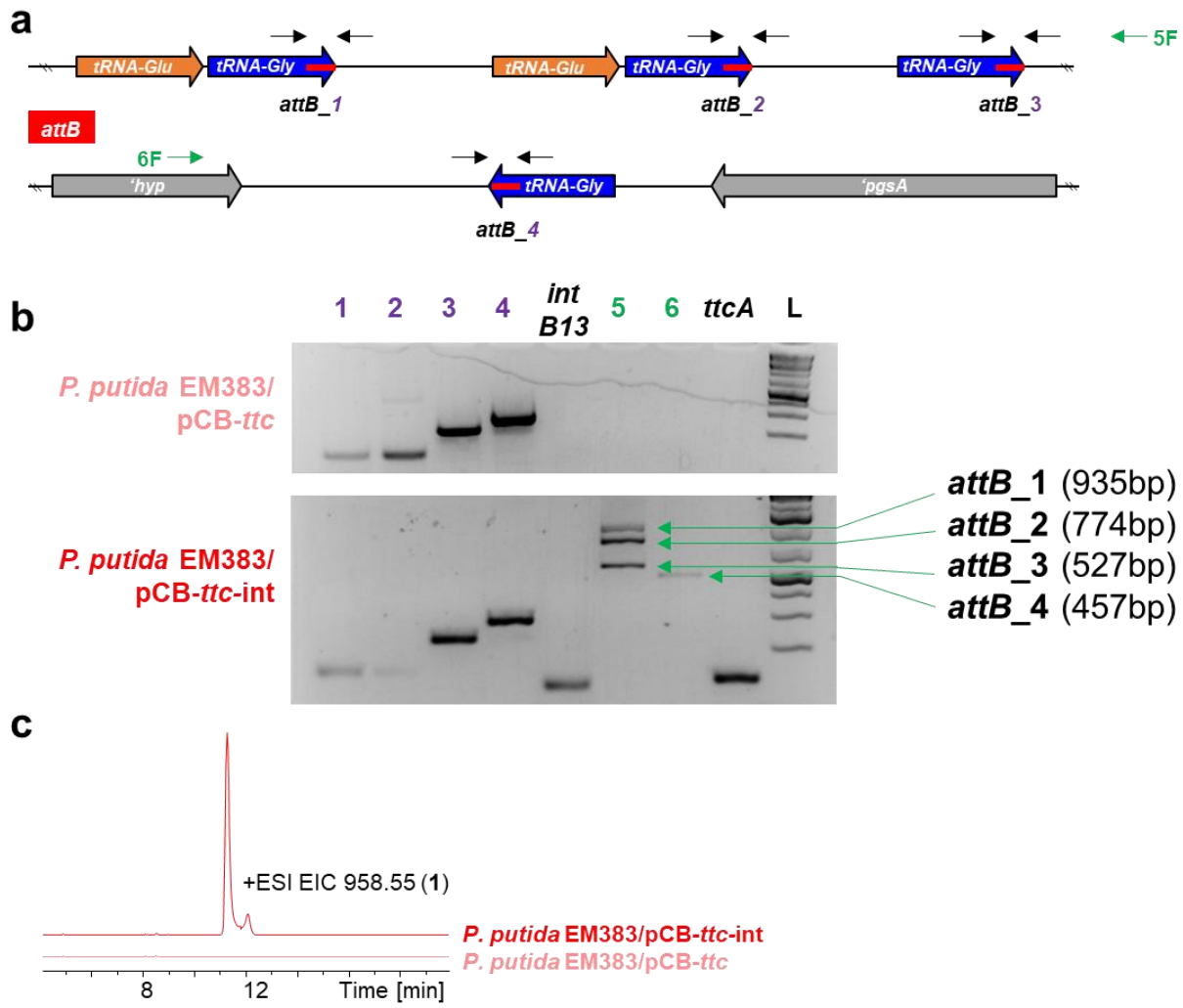


Figure 4.9 Characterization of IntB13-catalyzed integration of *ttc* into the genome of *P. putida* EM383. a) Schematic showing IntB13 attachment sites (*attB*, represented by red bars) in the *P. putida* genome, which sit at the 3' end of glycine tRNA genes. Sites 1-3 are clustered within a single locus, while site 4 is distantly located. Attachment sites were amplified across individually using primer pairs 1-4. Alternatively, primers 5F and 6F were paired with a reverse primer that anneals within *intB13* to determine heterogenous integration into multiple attachment sites within in a single clone. Integration into attachment sites 1-3 can be distinguished based on the size of the amplified PCR product. b) PCR analysis of individual clones of *P. putida* EM383 into which pCB-*ttc* and pCB-*ttc*-int were transferred. Kanamycin resistant *P. putida* EM383 clones can be obtained following transfer of pCB-*ttc*, but *ttc* is not integrated into the genome or stably maintained, as the *intB13* and *ttcA* genes cannot be amplified from genomic DNA. Conversely, IntB13 enables stable maintenance, as *ttcA* can be reliably amplified from clones harboring pCB-*ttc*-int. However, the construct does not integrate irreversibly into a single *attB* site, as integration into all sites can be detected from a single clone, suggesting the construct exists in equilibrium between an integrated and excised form. The experiment was repeated independently 3 times with similar results. L, ladder (GeneRuler 1kb Plus). c) Extracted ion chromatograms (EIC) of thalassospiramide A (**1**) obtained from LC-MS analysis, demonstrating that thalassospiramides are only produced by *P. putida* harboring pCB-*ttc*-int.

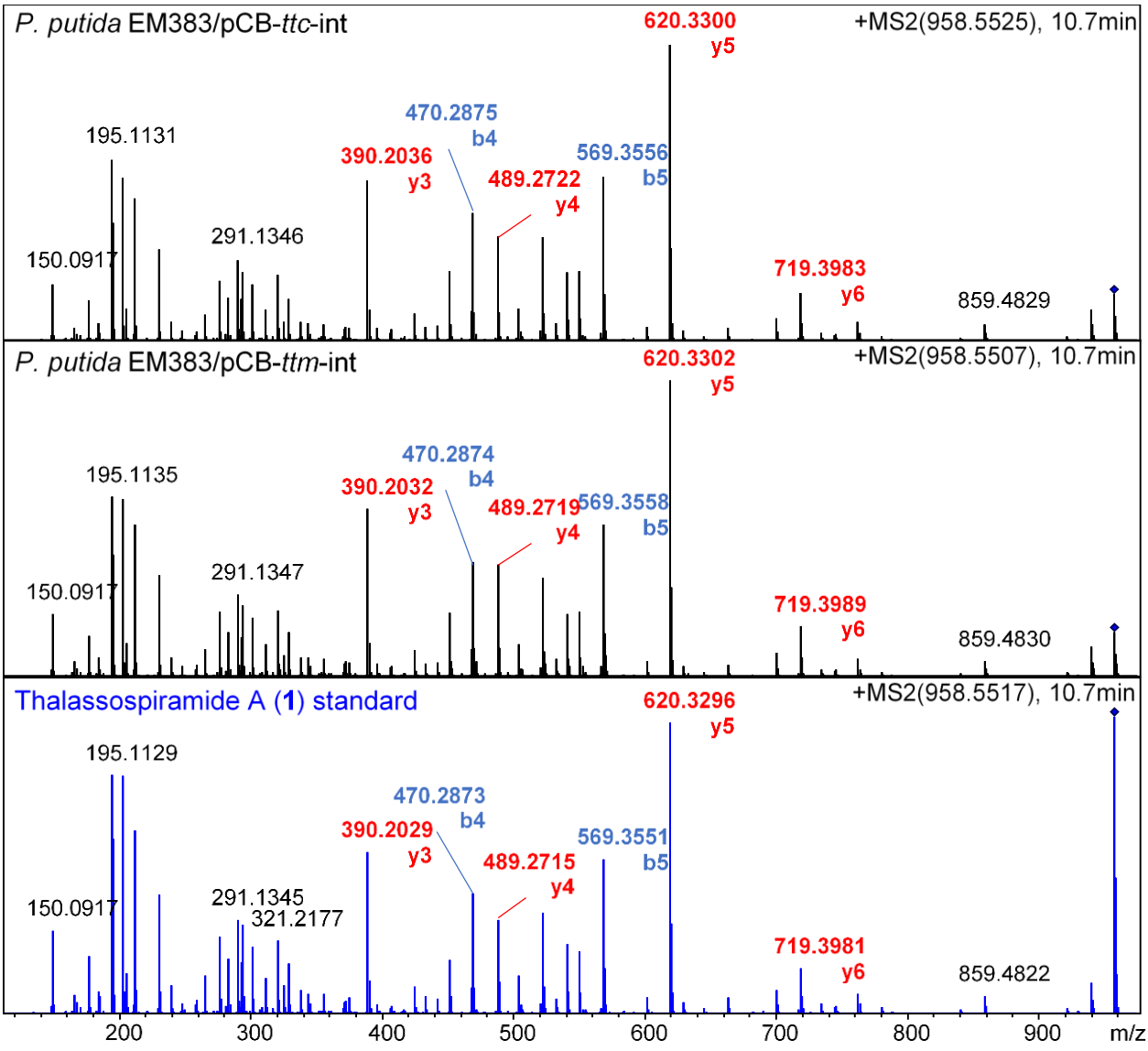


Figure 4.10 MS/MS spectra of thalassospiramide A (**1**). MS/MS spectra of m/z 958.5496 confirming heterologous thalassospiramide A (**1**) production from *P. putida* EM383 harboring the *ttc* and *ttm* biosynthetic gene clusters compared against a thalassospiramide A (**1**) standard. Similar MS/MS spectral comparisons were used to confirm all analogs reported. Experiment was repeated independently more than 10 times with similar results.

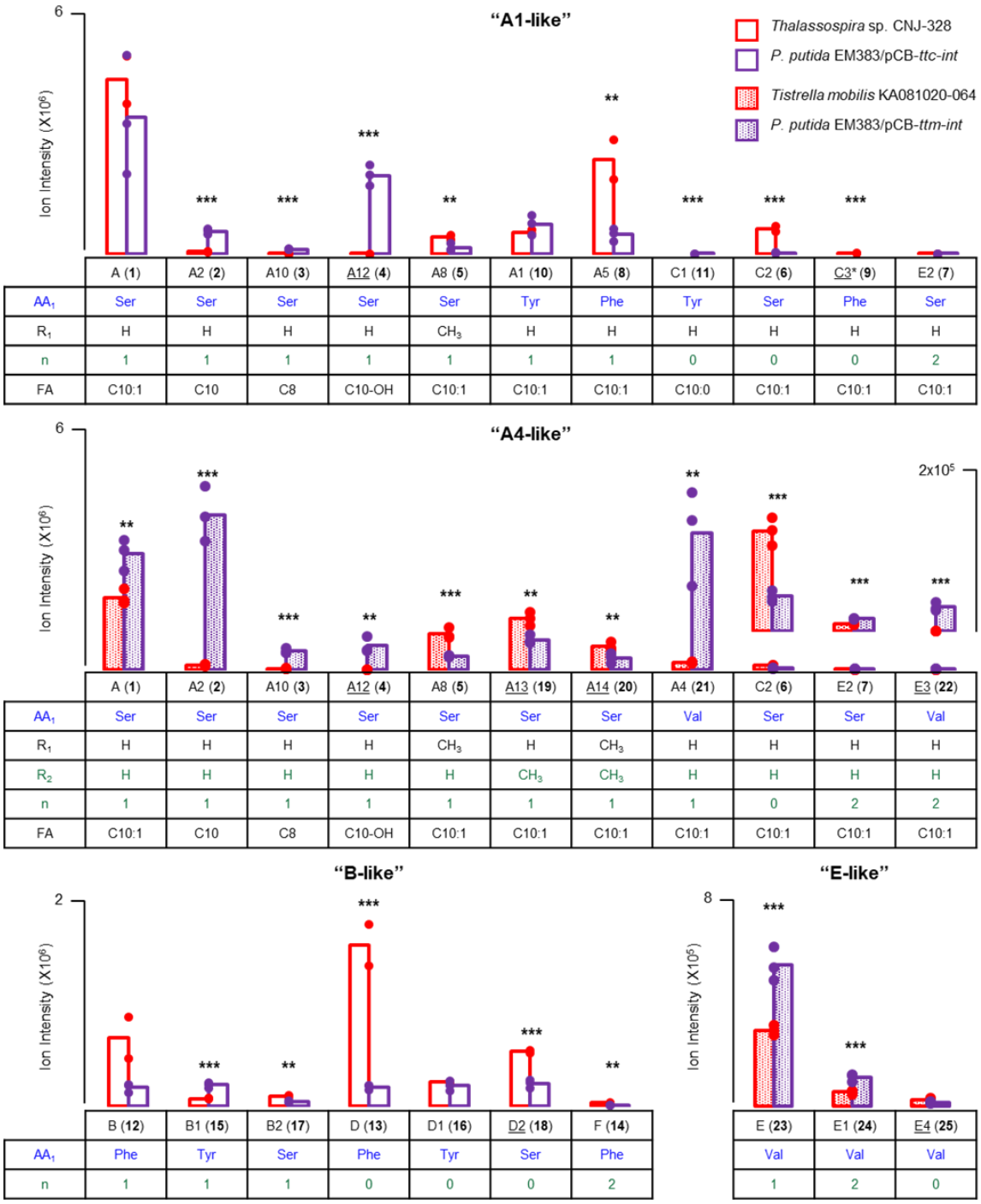


Figure 4.11 Heterologous production of thalassospiramide analogs from native producers and heterologous hosts. Extractions were made from triplicate 50 mL cultures and MS ion intensities were normalized by culture density at time of extraction. Significance was determined using a two-tailed Student's T test, n=3 biologically independent samples, **p<0.05, ***p<0.005. For full structures, see **Figure 4.3**. New analogs identified in this study are underlined. Asterix indicates analog was detected from the native producer but not the heterologous host. For HR-MS and MS/MS data, see **Table 4.3** and **Figures 4.21-4.34**.

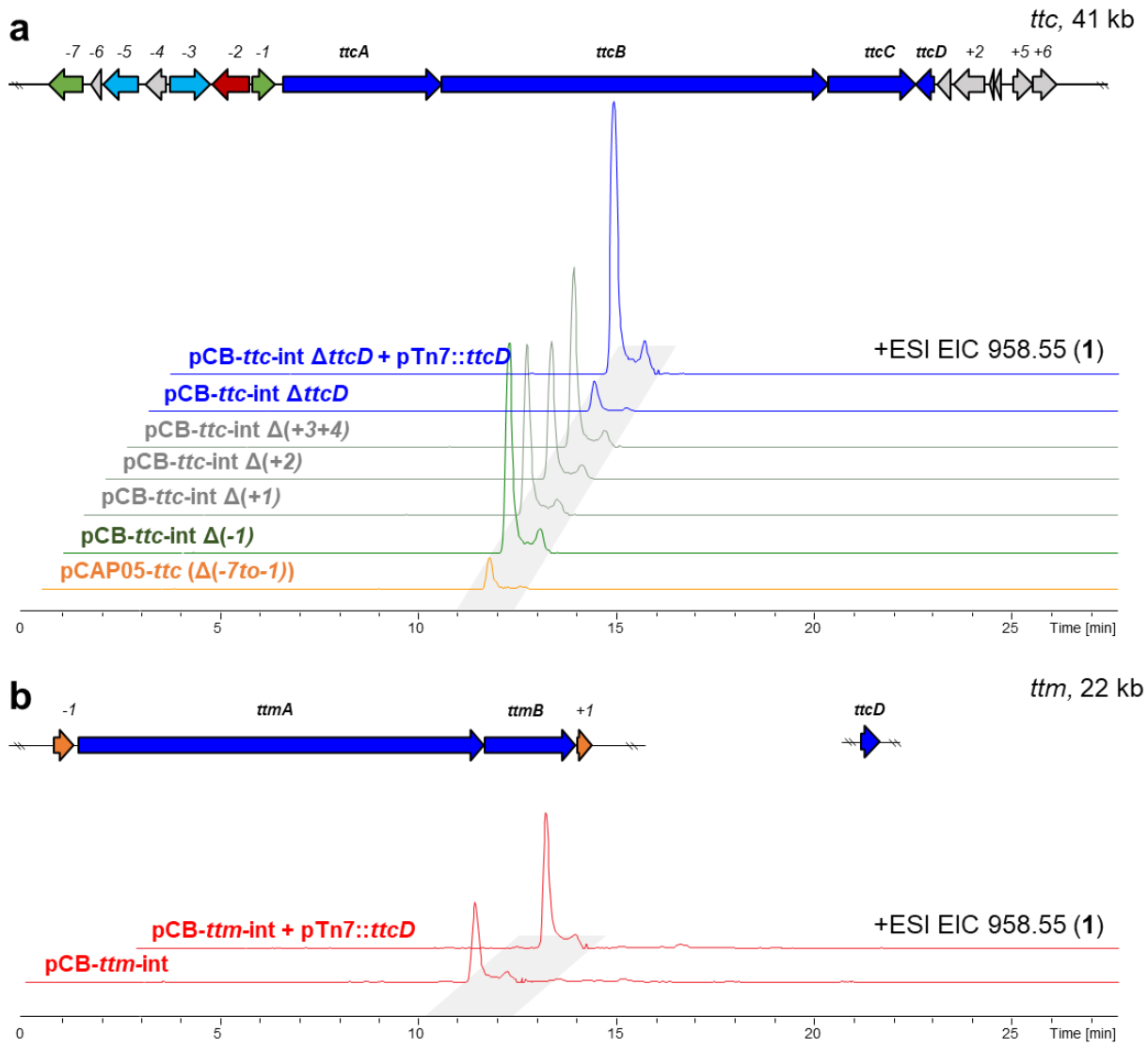


Figure 4.12 Deletion and complementation of non-assembly line genes. a) EICs of heterologously produced thalassospiramide A (**1**) following deletion or complementation of non-assembly line genes. Heterologous expression of *ttc* using pCAP05 (orange trace) did not include any upstream genes (*-7* through *-1*); although all thalassospiramides, including A (**1**), could be detected, the construct produced low yields and was found to be genetically unstable over time. Individual deletion of upstream (*-1*, green trace) or downstream (*+1* through *+4*, gray traces) genes in the stable pCB-*ttc*-int construct did not affect production of thalassospiramide A (**1**), although deletion of *ttcD* resulted in dramatic reduction which could be restored upon genetic complementation (blue traces). b) EIC traces showing co-expression of *ttcD* and *ttm* genes increases heterologous production of thalassospiramide A (**1**). All experiments were repeated independently 3 times with similar results.

a

ttc Experiment 1													
	"A1-like"						"B-like"						
	A (1)	A1 (10)	A5 (8)	C1 (11)	C2 (6)	E2 (7)	B (12)	B1 (15)	B2 (17)	D (13)	D1 (16)	D2 (18)	F (14)
AA_1	Ser	Tyr	Phe	Tyr	Ser	Ser	Phe	Tyr	Ser	Phe	Tyr	Ser	Phe
n	1	1	1	0	0	2	1	1	1	0	0	0	2
+*ttcD* / Δ*ttcD*	4.13 **	4.59	6.33 **	2.44 **	1.94 *	4.97 **	6.13 **	6.67 **	5.38 **	3.80 **	3.72 **	3.18 *	6.09 **
X+/A+ / XΔ/AΔ		*1.02*	*1.50*	*0.60* *	*0.46* *	*1.21*	*1.48*	*1.58* *	*1.31* *	*0.93*	*0.87*	*0.74*	*1.49*

ttc Experiment 2													
	"A1-like"						"B-like"						
	A (1)	A1 (10)	A5 (8)	C1 (11)	C2 (6)	E2 (7)	B (12)	B1 (15)	B2 (17)	D (13)	D1 (16)	D2 (18)	F (14)
AA_1	Ser	Tyr	Phe	Tyr	Ser	Ser	Phe	Tyr	Ser	Phe	Tyr	Ser	Phe
n	1	1	1	0	0	2	1	1	1	0	0	0	2
+*ttcD* / Δ*ttcD*	5.18 *	9.83	12.13 **	2.78 *	3.87 **	5.84 *	8.31 **	5.97 *	4.48 **	3.77 **	2.80 *	2.35 *	5.19 **
X+/A+ / XΔ/AΔ		*2.14*	*2.35* *	*0.55*	*0.79*	*1.11*	*1.62*	*1.19*	*0.95*	*0.76*	*0.60*	*0.50* *	*0.99*

b

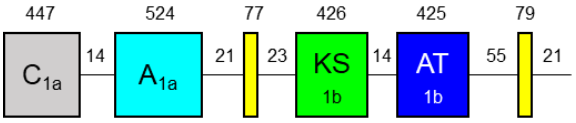
ttm Experiment 1								
	"A4-like"					"E-like"		
	A (1)	A4 (21)	C2 (6)	E2 (7)	E3 (22)	E (23)	E1 (24)	E4 (25)
AA_1	Ser	Val	Ser	Ser	Val	Val	Val	Val
n	1	1	0	2	2	1	2	0
+*ttcD* / -*ttcD*	1.51 *	2.08 *	0.60 *	10.3 *	12.7 *	1.09	4.13 **	0.68
X+/A+ / X–/A–		*1.42*	*0.40* **	*6.72* **	*8.42* **	*0.71*	*2.78* **	*0.47* *

ttm Experiment 2								
	"A4-like"					"E-like"		
	A (1)	A4 (21)	C2 (6)	E2 (7)	E3 (22)	E (23)	E1 (24)	E4 (25)
AA_1	Ser	Val	Ser	Ser	Val	Val	Val	Val
n	1	1	0	2	2	1	2	0
+*ttcD* / -*ttcD*	2.61 **	2.56 *	0.57 **	9.7 *	10.4 **	1.26	3.61 **	0.70
X+/A+ / X–/A–		*0.99*	*0.37* **	*5.95* **	*6.62* *	*0.79*	*2.36*	*0.44* *

Figure 4.13 Effects of *ttcD* deletion and complementation on thalassospiramide product distribution. a) Results of two independent experiments comparing host cultures expressing *ttc* with *ttcD* deleted or re-complemented. Extractions were made from triplicate 50 mL cultures and MS ion intensities were normalized by culture density at time of extraction. The top row of values represents the fold-change in ion intensity for a given analog upon *ttcD* complementation. For the bottom row, ion intensities were first normalized by levels of (**1**) from the same sample and then compared to represent a normalized fold-change value. b) Results of two independent experiments comparing host cultures expressing *ttm* with or without *ttcD* co-expression. Co-expression favorably enhances production of analogs with n≥1 and decreases production of analogs with n=0. Color intensity corresponds to numerical value. Significance was determined using a two-tailed Student's T test, n=3 biologically independent samples, *p<0.05, **p<0.005. For full chemical structures, see **Figure 4.3**.

TtcA

```
MEGQTLQKRTNTFAPGTERQGAPLTTPQSSIWLDWAKGDDCANYNICSVIRFDGDLDIPALHKAIVQTDQENDALR    76
LRFGSEGGNPFQFFADTCRETDFSVIDLSFADDPVLIAHNEVEEIRRRPLDPENGRHCRHRLLQLGKGRFWWVRVY   152
HHLVCDGYAGHLMAMRAAEIYSALKTATAVPESTFGSYLDFITADTTYPETPAFQRDLAYWQERLKDDHTITRFSS   228
DRPSGKLVQARLEQNLDGQELARLTQTAKACGTSPTAIVMAAYFILLGQITHIKHPTVTMPLLNRLGRAERQTPGA   304
FSYIVPFDADLGKHARFSDLARDIFARARRDVRHMRLGAPRMRAARLGARGLSGGGAFFNSLDGADPLSFAGLDTR   380
RVNIYNGPVTDLGLIYMMQALSPDRQEAEIIWQYDTACHDVKSVTRIADRFRHFLSRALDDPEQEIATIAAPPLDE   456
LAVIRKLETGPAFTRPVPELLIPDRIQHAISSRPGDIALITDDGREVTFATLGHMTNAVAHHLRQNGIAAGDFVGM   532
NVSPSAQQIAAVIGILKCGAVCVQLDPLQPVARNASMTRHLDCRLVLSDSTDAWCEIGNTTITVKDINAVPGIDGR   608
TVLEDVLADICGDDIAFVFHTSGSTGQPKPVPVHHQSLADKIDVAIAQFGIGSHENGNPETMGLFPSLGFDPWMHQ   684
FCVGLICGHRMWLMTHAILSDGHKFWNGVARHHVTHLMSVPSFYESVIDAIPADGNFKLRRMIIGGEVLQPTLSRR   760
LHSVFESADIWNCYGPTEATIHATTHLVPREIKTKNIPLGHVDKGAFVRILDEDGKRVTLGVPGEIYLGGTGLNRG   836
YINMPAVTAEKFVKDPTSETDETFYRSGDLASWGEDGLLYYHGRTDEQVKIRGQRIEIGEIEYQLSRVPGVGQGAV   912
LYVKNNHGGELIGYLLPDGSHSKRPDIATVRSELARHLSDAAVPTRLEWVETLPLLPSGKIDRKALAQFAQAPGGK   988
QPVAQETPPPASRKPDQMRIHQMAEKIAKIWMDLIETDEIEFDTNLFEAGAHSLLVPRAQFALSKLAGRNIASVEI  1064
FQHPTINSFALHLCADNQTVPETAKAEPVKTQRNVSENHDIAIIGMAMRLPGADDRDTFWSMLENGVDQIKDVDIS  1140
RLRTLGGVPSALEDADFVARHGLLDDIDMFDPAPFAMTAGETIETDPQQRLLLEVALGALEDGSCDPARDGPVGAY  1216
VGVGFPTYLVDSLRDRLRDRPDAIRYGMTLGNDKDFAATRLAYKLNLTGPAVASSTACSTGLVNIALAVEALRSGQ  1292
CRVALAGGAALGLSRPGGYYFTQGGIGSRSGVCRPFDSKADGVVGGSGAAIVLLKRLDDAIADRDTIHAVIKGVGM  1368
NNDGSAKAAFTAPSVHGQAKAIQTALDDAQIDPVTIRFVEGHGTATALGDPIEIAALNKAYAKTGTSPTHPIWLGS  1444
VKGNIGHLDSAAGIAGLIKAVLAIRHRAFPPTCHFNQPNENLGLDQGPFRVNSIVEPLSSGGDEPLRGGISSFGVG  1520
GTNAHIIIEETPRADKPASPTPKNDIPHVLPFSAASEAALQKLLTNTASWLDRNGSGISVSDVANSLSRRHAYRFR  1596
TAIVATNINTAISSLANPKSSARHQAVAQRNTPSITFLFPGQGSQRPGMAADLYEADQTSAAMIRNACQQVAKIGG  1672
PDDLFELLTVPDESAATSSRMARTEIAQPALFIYEYALASWLIGNGISPAALAGHSIGEYVAACIGGVMSFDDALR  1748
LVVTRGRLMGQTEPGSMFALSMTEAEVGDLLSAFGPDLSMAAINGPRQCVVAGSNECIDALEGHIKSQGKSGRRLV  1824
VSHAFHSRMMDPVLDEFRNVVAKIGLNVPTVPIQSNLTGNWLTPEDATNPDYWVQHLRNAVRFGDNIKGLLAGMPE  1900
TIMVECGFGNTASRLAIVNGTPTGNSIALQPAPGPDTARNANAPITGSNALAATIARLWVMGTAPDWDRVNQTKDA  1976
RRIPLPTYAFDRRRFWPDKAITTTADPKPFAVPAAPAEATMPTAQNAPNIQSDTQAQFTILAIWRDMFGDPAIGPD  2052
DDFFELGGDSLFAVRIATRLSEVLNAEIPAAVMFEGRTVSGVANLIASFDKNSAPTNTPPAPAAPPSTREQGVL*   2126
```



Domain	Start	Stop
C_{1a}	1	447
A_{1a}	462	985
T_{1a}	1007	1083
KS_{1b}	1107	1532
AT_{1b}	1547	1971
T_{1b}	2027	2105

Figure 4.14 Annotated sequence of TtcA. Domains highlighted, conserved motifs[22-24, 48-55] bolded, active site residues colored red, residues targeted for mutation underlined. Schematic not drawn to scale.

TtcB

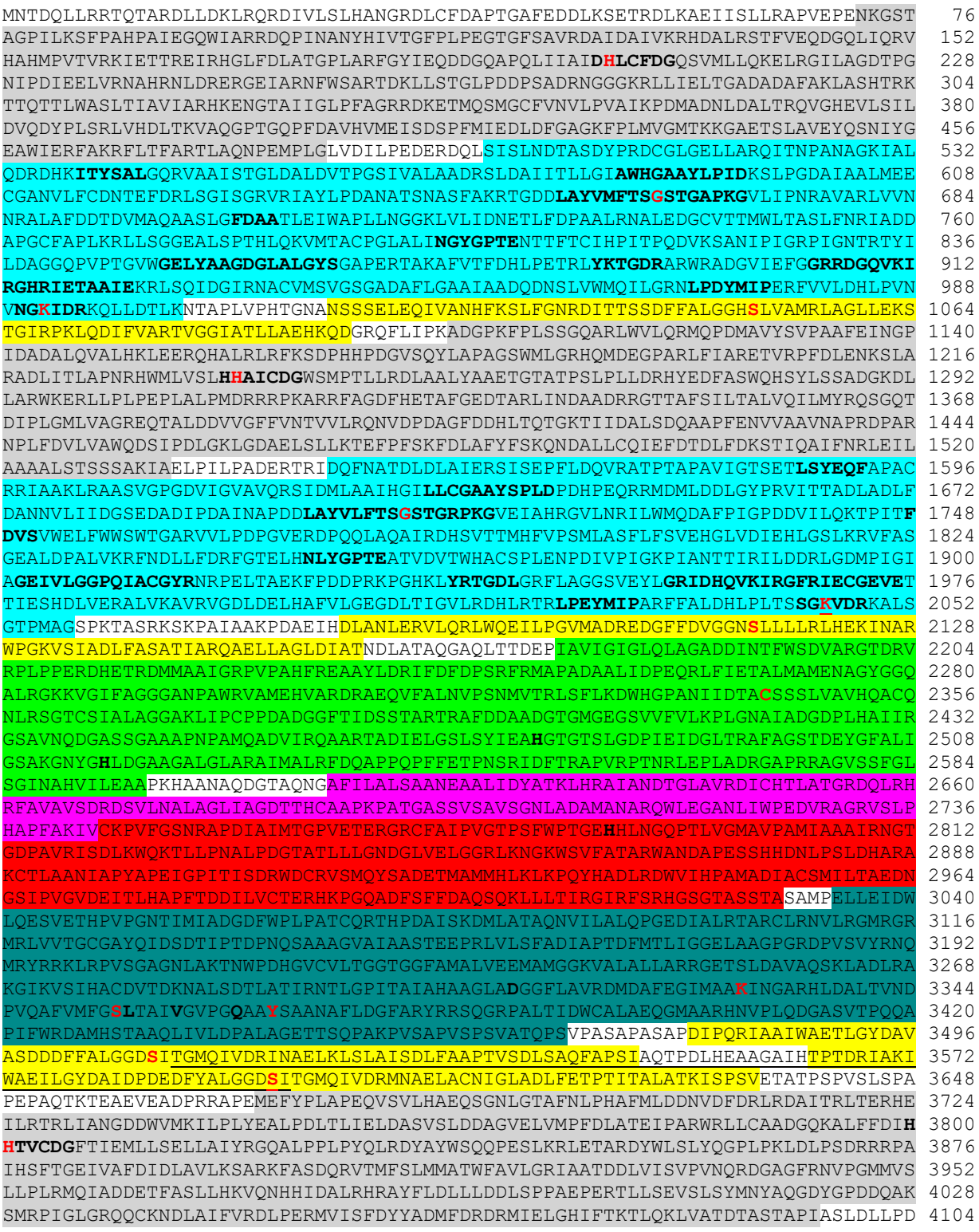
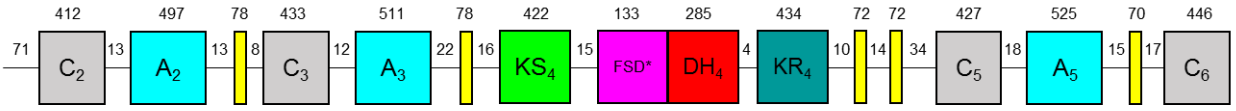


Figure 4.15 Annotated sequence of TtcB. Domains highlighted, conserved motifs[22-24, 48-55] bolded, active site residues colored red, residues targeted for mutation underlined. Schematic not drawn to scale. FSD*, flanking sub-domain.

EQKANIEIWEQGPEPDTRLDTDANAGIFAIFARQASITPDAPAIRDTNGT**WTFGEL**HANACAIAHCLQTAGIEPGD 4180
LVALHIERGRPAIAAILGI**TAIGAGYVPLD**PAYPAARNQFILADSGAKLVLVDDAGHDALDALGALSESGHPSPQR 4256
INIGGIAPAPTDFKLPELTNADSL**PAYIMYTSGSTGTPKG**VLIEQGAVRRLASGADYAKIVASDVVAQAGPLA**FDA** 4332
STFEIWATLLRGVEIAVIDRNDLLDPAKFGDALTQFGVTKMFMSVGLFNRQVDHDPQTLAGLDILMIGGDAISKQH 4408
VRNFINACPHVTVL**NGYGPTE**STTFAVVGPITENDLSGDTDATIIGKSIAHTKTLILDQNGQRAPVGVW**GELMIGG** 4484
SRLAREYWQRSDLTGERFIPDPDQPDQRL**YRTGDL**ARWTRDGRIEFG**GRHDNQIKLRGFRIELDEIE**QQLQSAPGI 4560
KNAVALFDVNAPDGGAIIGCIQTTQDQDIDIPALMTWLGTH**LPGYMIP**AKWHIVDTIPITA**NGKVDR**KALLETVRT 4636
QSNTMALDGGTDTPPANPAEELVCDIFTEVFAIPIRNRHASFTQLGGH**S**LMAIRIVNRLADRTGKRISMADFFADP 4712
TVAGLARRIENATLDIAATGNGQIPAAPAMEYYPASHAQKRLYLLSKMDGNSGAYGMLYVFRCTDDLRADILQSAL 4788
TRLVDRHETLRTGFAEHDGTITQKIHATTPPAVTFDDLADHADPAREALRLTRLEAATPIILDQPPLIRGRVIRVA 4864
ANESLMILMT**H**H**IVGDG**WSSRILINELGALYDAALHERNDALPPLPITYKDFAHWQSGQDWADAADYWRQQLANAP 4940
DQINLPTDRPVPEKQSFRGAHVHLALPDDVLSGLHQLASKNKTTLSSVGLSLFSAILYRLTRQRDMVIGMGVAGRE 5016
RIETEGLIGFFVNVLPIRMQLDDDTDLVDLIRATHVNIANALDRQDFPFDELVRAVAPKRKSNRQPLVNVVFEYQR 5092
FEAVSGLAQNGWAQSGLPLLAPDAAGILPSNMDAFVDNTTAKHDIILFLTEEAGQARFTLEYDTDLFDAETMQRWL 5168
AFLGKFAAAAAQNAQKDA* 5186



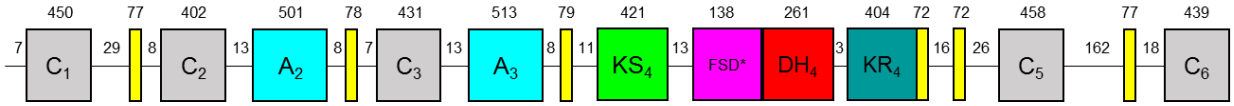
Domain	Start	Stop
C_2	72	483
A_2	497	1002
T_2	1016	1093
C_3	1102	1534
A_3	1548	2058
T_3	2081	2158
KS_4	2175	2596
FSD*	2612	2744
DH_4	2745	3029
KR_4	3034	3467
T_{4a}	3478	3549
T_{4b}	3564	3635
C_5	3670	4096
A_5	4115	4639
T_5	4655	4724
C_6	4742	5187

Figure 4.15 Continued.

TtmA

Figure 4.16 Annotated sequence of TtmA. Domains highlighted, conserved motifs[22-24, 48-55] bolded, active site residues colored red, residues targeted for mutation underlined. Schematic not drawn to scale. FSD*, flanking sub-domain.

```
LRAAVAALIRRHAALRTGIIGADHAQGGEGWPGDDDWYGGDDWYGGDDWRMVAIDAAIVTPSLAIEPIAHPVGTAP  4180
AAPIRPFSTDGGAMARWSLHRDPAGRLVLAFDIHHSMSDAQGVEILLTELAALYQGRDLSPVQVDLRDYGWWSAVR  4256
GDAAIQAARAWWDGIFPATAQGRDLPGRDLPRLDLPADRPRPPVHGHDAAATGFDLPPALSRDIRALAARRRTTPF  4332
TVLLTGWAILLSRLTGRDDITVALPVNTRAAAGFDGVVGMHAALVPVRMQVPADGSGAALLAHIRDRQAGALSHHA  4408
LSLGRMLAEFQAPAMPERSPLSEVSISYMNHAATAAGTLDAGGFRLRGLTRADGKNDLSVFIADKGEGFGLVLEYD  4484
TGLFDAARIEALGPMLQRILAGLTADPDQPVGRLPLLDDETRGLLAAREAAEGVSVLDDRGQPVPPGVWGRLHRLG  4560
DDGWQPTGDRARWRLDLTAIETGGRLDRLTLWQGHLLDPALIEAQLSGLPGVAALRILPRPGRLVAAITAAGPRPD  4636
PAHLRRAAGAILPPVLVPGAWHLLDALPEDDAALAAGVAATPSLEEAAEAATERLATVLAVVSDVLGRPATDPSAD  4712
FHALGGHSLLAMRVVNRLEAATGIRIPMRAFFADASMAGIARSLGEAASEATRAAAEDVIPPGAGPGAHPASHGQQ  4788
RLYLLQQMQPESPAWLMLFVLPAGAVDARALRLALDRLSARHATLRTGFATTDDGIVQLVAAEAPPPLVVDDLRHR  4864
TDGMDEALRLARHEAATPFDLAAAPLFRARLIRVGDHESLLLLVLHHIVGDGWSSRILLAELATLYHAATEGRGDP  4940
LPPLPASYADYARWQQGRDWQAALDGWRDRLAGAPDAIALPTDRPPRPDASGRGATVRRELPAEVLAGLSRVARRH  5016
NTGLAAVGLALFAAVLYRLSRQDDMVIGMGVAGRDRAETEGLIGFFVNVLPIRLQLDDETGLGQLIDLTRDRLAEA  5092
LDRRELPFDRLVRALAPGRAGGRQPLANVIFEYQRFDDVKAATTEAQLPPPPADGGRLGSAVAGLVDASTAKHDLI  5168
LFFEDGGDRARFSAEYDTDLFDAATIEAWTGYLLRIATAVAASPDEI*                               5215
```

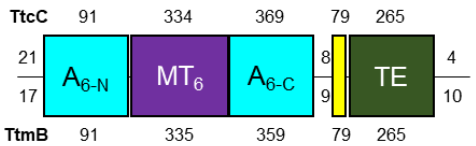


Domain	Start	Stop
C_1	8	457
T_1	487	563
C_2	572	973
A_2	987	1487
T_2	1496	1573
C_3	1581	2011
A_3	2025	2537
T_3	2546	2624
KS_4	2636	3056
FSD*	3070	3207
DH_4	3209	3469
KR_4	3473	3876
T_{4a}	3877	3948
T_{4b}	3965	4036
C_5	4063	4520
T_5	4683	4759
C_6	4778	5216

Figure 4.16 Continued.

TtcC and TtmB

```
TtcC    MKIADFDFDAL-DTPTPGMNGTPDVPPSATTGSIIDRFASIAASYPHRIATRDETGEISYTDLDRQSNR   68
TtmB    MKIGDLLLDDIAQKPRLAEAAAP-----AVDGTVLDAILRRAAAHPDRMAVQDEDGAVSYAGLVAEARA   64

TtcC    IAHFIAGLALQTGSLVGVMTGRNRHYIAATLGVLKAGLAFVPLDPNVPLIRRTTLVRNAALPLIISEAT  137
TtmB    IAGVIAAAGLGDEARVGVMTGRSRQYIAAVLGVMMARAAFVPLDPAMPLPRRQMIARAAGLAMIIADRD  133

TtcC    HIGDLHQLQWRCKTVTNILCLDRDDIDTLTENHGSLMSVELWDHLAGDDADNILAGGWKSAFTGQPIAQ  206
TtmB    RMGDARRLAWSCPDIAHLVCPDLDDPDSPTEAHGARMSAELWDHLAGSAADDVAAAGWRSPFTGQPIAE  202

TtcC    SAMDAFGANARNKVAGLVSRDARVLEIGCASGFTMKHVAPVVGTYVATDLSRRNVERTEALAINLGMTH  275
TtmB    AVMAEFGQAAAGKLSGLVTPASRVLEIGCASGFTLRALAPLAGPYLATDISRRAVERAGGVARRLGLDR  271

TtcC    VIGRQLGAHDIDVYEPGSFDLIIMNSVIENFPGYGYLRTVLTKAKTLLAPGGAIYLGNVWDADRRDDYL  344
TtmB    VDLKPMAAHDLDLLAGRSFDLIVLNSVIESFPGFGYLADVLTRAAGLLAPGGAIFLGNIRDLDRRDAFL  340

TtcC    ADLARFARETVG-KGYNTRLDFLEDLFVPQCFLEDWAREQGGFAANRSAIDADGFDPAPYTFDFVLTSD  412
TtmB    DDLVGFARSHAGDPEIGTRLDAADDLFVPRRFFEDWAARHGGFRAEASALDVPGFDLAPYVYDLVLRRD  409

TtcC    --AAKHPVAPKKHRHDLAAL-NAANDQRPATVTSANDLAYVLFTSGTTGTPKGVMIEHGSVVNLAQSVL  479
TtmB    EAAPAHP--PRHSHHGRAALAKAASGPLPAAPT-PDMLAYVVFTSGTTGTPKGAMIEHASLANLAMATA  475

TtcC    DTQLKQIGADDPDRTLNLTCVAPFAFDGSIVQIFAAILNGHCLNITGSATRSDPEALHDLMVTNAIDQL  547
TtmB    QVFYAPLHPSD-TRPLDLCCIAPFAFDASIIQTLPVLANGHRLHLPDDDTRRDPALLDAFLRARNIAMI  543

TtcC    DATPSLFSLLLDYWEESRLSCPARMVVLGGEAVSTRLVERFFALPGQTDRKLVNAYGPTECCVSSAQYV  616
TtmB    DLTPSQFTLLLDHWKTSGTHAPVLQVVLGGEPVTTALVERIYADPAHAGLRLINAYGPTECCVGAAAHA  612

TtcC    MTAHNHKQILPPPIGKSIQGVYLEICDDIGRPLPDGVPGEIRIGGAGVARGYLNDQILTAQKFVTAPNG  685
TtmB    MTAHGWREILPPPIGRPFPGVVIEIRDAAGRPVPPGVAGEIVIGGRGVGRGYLG-AAGAATRFETDAAG  680

TtcC    NRVYRTGDMGRRRSDGEIIFLGREDGQIKVRGNRIELGEVEHALLTHPHIRRAAVVAHDRHQNGNVVLV  754
TtmB    RRWYRTGDMGRSLAGGAIAFIGREDGQVKIRGNRIELGEVESALLAHPFIRAAVATLADG------ALL  743

TtcC    AYVVTDDDAHAGFDISACRAKLEQALPNYMVPSHLITMDDLPLTVNGKIDLSRL--PVPENTDQVTGVT  821
TtmB    AHIV----PRPGFDEARCRADLEARLPAVMIPSRILVIDEIPLTVNGKVDTARLPRPAPAGTNRPTAAA  808

TtcC    -AKPLGSDHEKTVARLMGEVLDCTIDDATADFFELGGHSVLAVQFLSKINKLTNVKIPLGDLFTCASVE  889
TtmB    PLRPLADETERRVARIMGEVLEVAVDDAAADFFRLGGHSVLAVQLVDRLRRAFGVKLPLSDLFAHATVE  877

TtcC    KLAARISERENARHVKSPIVTVNASGTKPPMVCFHPVGGNVLCYQNLAREFGPDQPLYMVEAQGLEEGQ  958
TtmB    GLARRLARQ--ARTRRDRLVVVNAAGHLPPIVCFHPVGGNVLCYQALAAGLGPDQPVAMVEAAGLEADD  944

TtcC    NLQPTVEEMVAAYMPAIRQTVPDGPIHVIGWSFGGLLAFEAAFRMQQAGIDVRNIILLDAIAVPDAVRA 1027
TtmB    TPEPSVEDMVAATLDDITRFAAGRPLHLAGWSFGGLLATEAARRLDQAGSPTGRVLLFDAVASPDPIRD 1013

TtcC    ILRKGEAEYLADIFAELGLVTVDELRPLTPEERIDLLLERGKNCDLLPDEADRGTMRRLLAVFQNNALA 1096
TtmB    MLAKDDADYLATLFAGMGIADAATFRALDPEGRLDLLIDRGKTALGLPDGVDRDAMRRLLAVFQNNAVA 1082

TtcC    AVHYVPPKIDGIDVLLLRPTTETRAAPSIEGDQYSGWQRVAGDALTLEWVGGTHGTMMFDPHVIEVAER 1165
TtmB    AIRYRVPRLDRLSALLVRPKTASAQAPGLPGDPWNGWRDRFGAGVTLAWMDGDHASMMMPPHVEVLAGV 1151

TtcC    IRRHLEKSK*------           1174
TtmB    VRTYLGGAPEAGSAA*           1166
```



	TtcC		TtmB	
Domain	Start	Stop	Start	Stop
A_6 – N	22	112	18	108
MT_6	113	446	109	443
A_6 – C	447	815	444	802
T_6	824	902	812	890
TE	906	1170	892	1156

Figure 4.17 Annotated and aligned sequences of TtcC and TtmB. Domains highlighted, conserved motifs[22-24, 48-55] bolded, active site residues colored red, residues targeted for mutation underlined. Schematic not drawn to scale.

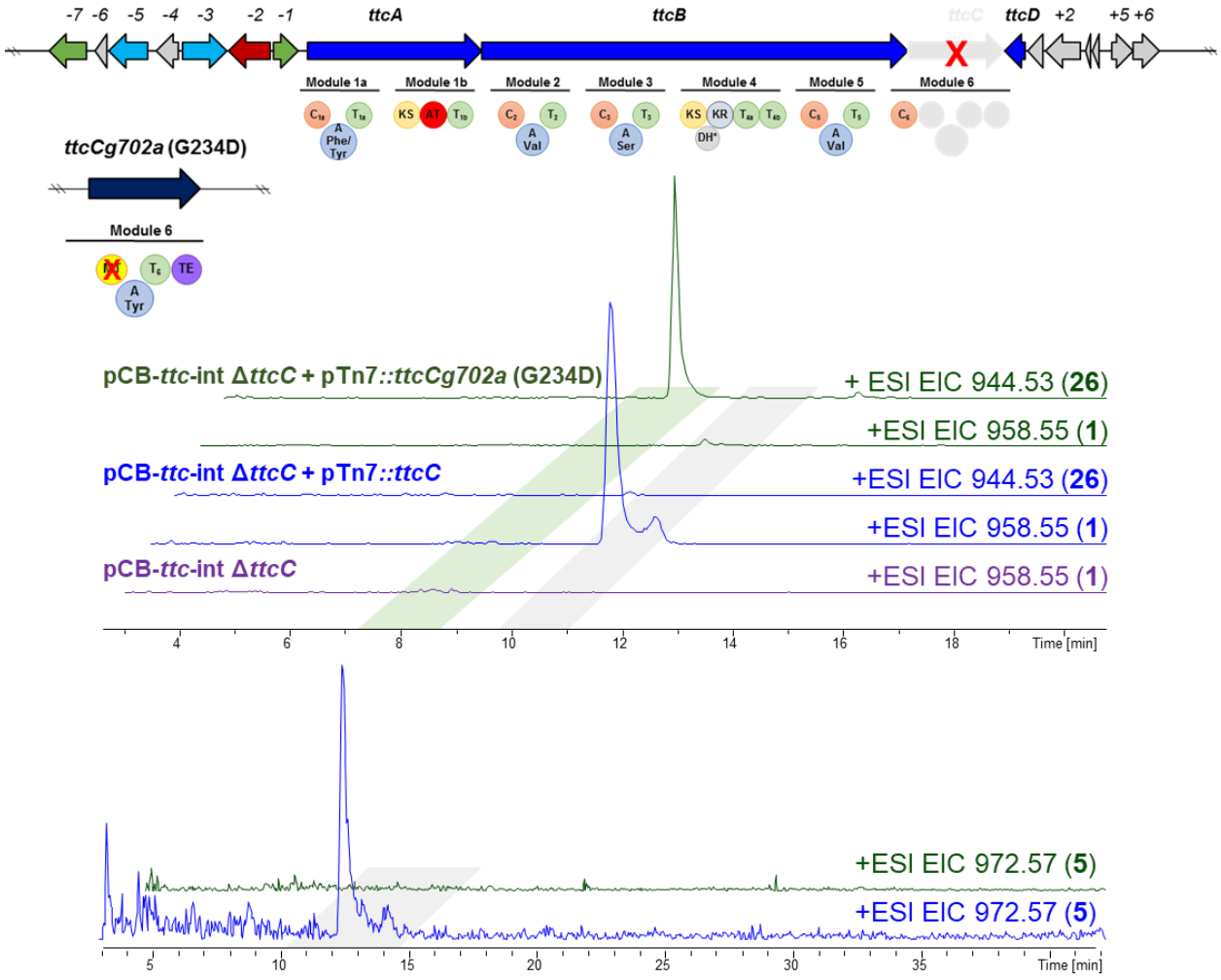


Figure 4.18 MT_6 inactivation via gene deletion and complementation of *ttcC*. Schematic for gene deletion and complementation strategy for precise inactivation of MT_6 encoded on TtcC. EICs for thalassospiramide A (**1**), A15 (**26**), and A8 (**5**) from deletion (purple) and complementation (wild-type, blue; MT_6 mutant, green) strains. MT_6 inactivation results in loss of N-methylated products A (**1**) and A8 (**5**) and concomitant formation of a desmethyl analog A15 (**26**). EICs of A8 (**5**) are shown at different scale for clarity. Experiment was repeated independently 3 times with similar results.

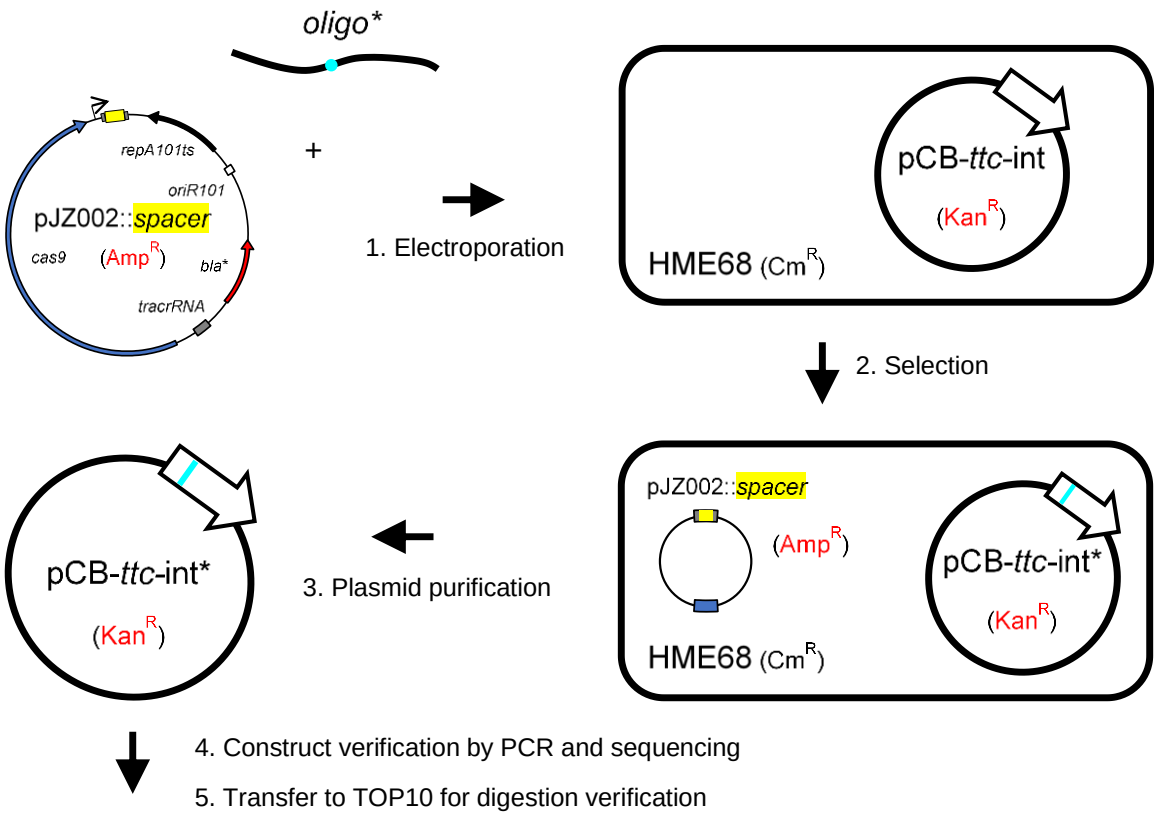
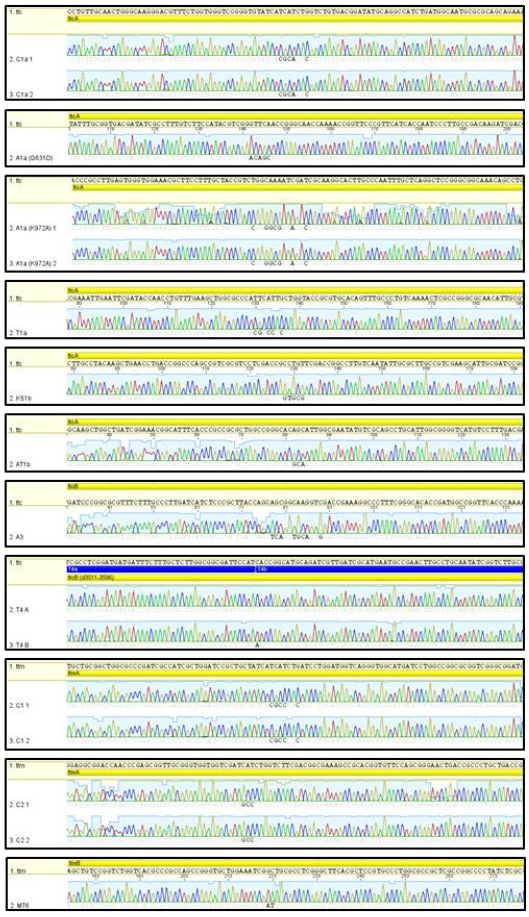


Figure 4.19 Schematic for oligo recombineering/CRISPR-Cas9 method. A spacer sequence (yellow) is cloned into targeting vector pJZ002 following previously established protocols[56], and an oligo (*oligo**) is designed with the desired mutation (turquoise). Targeting vector and oligo are mixed and introduced to *E. coli* HME68/pCB-*ttc*-int by electroporation following previously established protocols[57]. Selection with appropriate antibiotics (Amp, ampicillin; Kan, kanamycin) should select against unedited pCB-*ttc*-int constructs and for oligo recombineering, which results in generation of the desired mutant construct that is no longer recognized by the targeting vector. Edited constructs are purified from up to 4 individual clones and screened by PCR and sequencing. In order to confirm construct integrity (no unwanted deletions or rearrangements), plasmids must be passaged through *E. coli* TOP10 to obtain DNA that can be clearly analyzed by restriction digestion. Constructs in this schematic are not shown to scale.

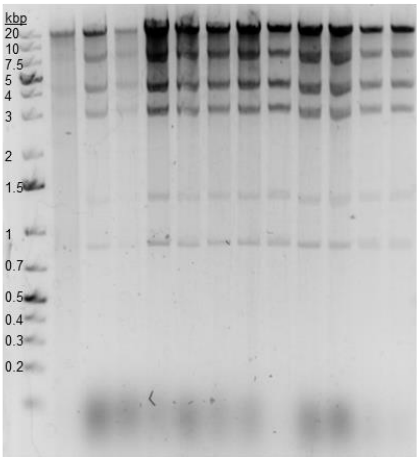
a

	Name		Name
1	pCB-*ttc*-int	10	pCB-*ttc*-int-A3
2	pCB-*ttc*-int-C1a 1	11	pCB-*ttc*-int-T4 A
3	pCB-*ttc*-int-C1a 2	12	pCB-*ttc*-int-T4 B
4	pCB-*ttc*-int-A1a (G631D)	13	pCB-*ttm*-int
5	pCB-*ttc*-int-A1a (K972A) 1	14	pCB-*ttm*-int-C1 1
6	pCB-*ttc*-int-A1a (K972A) 2	15	pCB-*ttm*-int-C1 2
7	pCB-*ttc*-int-T1a	16	pCB-*ttm*-int-C2 1
8	pCB-*ttc*-int-KS1b	17	pCB-*ttm*-int-C2 2
9	pCB-*ttc*-int-AT1b	18	pCB-*ttm*-int-MT6

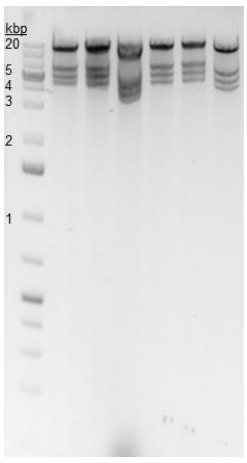
b



c

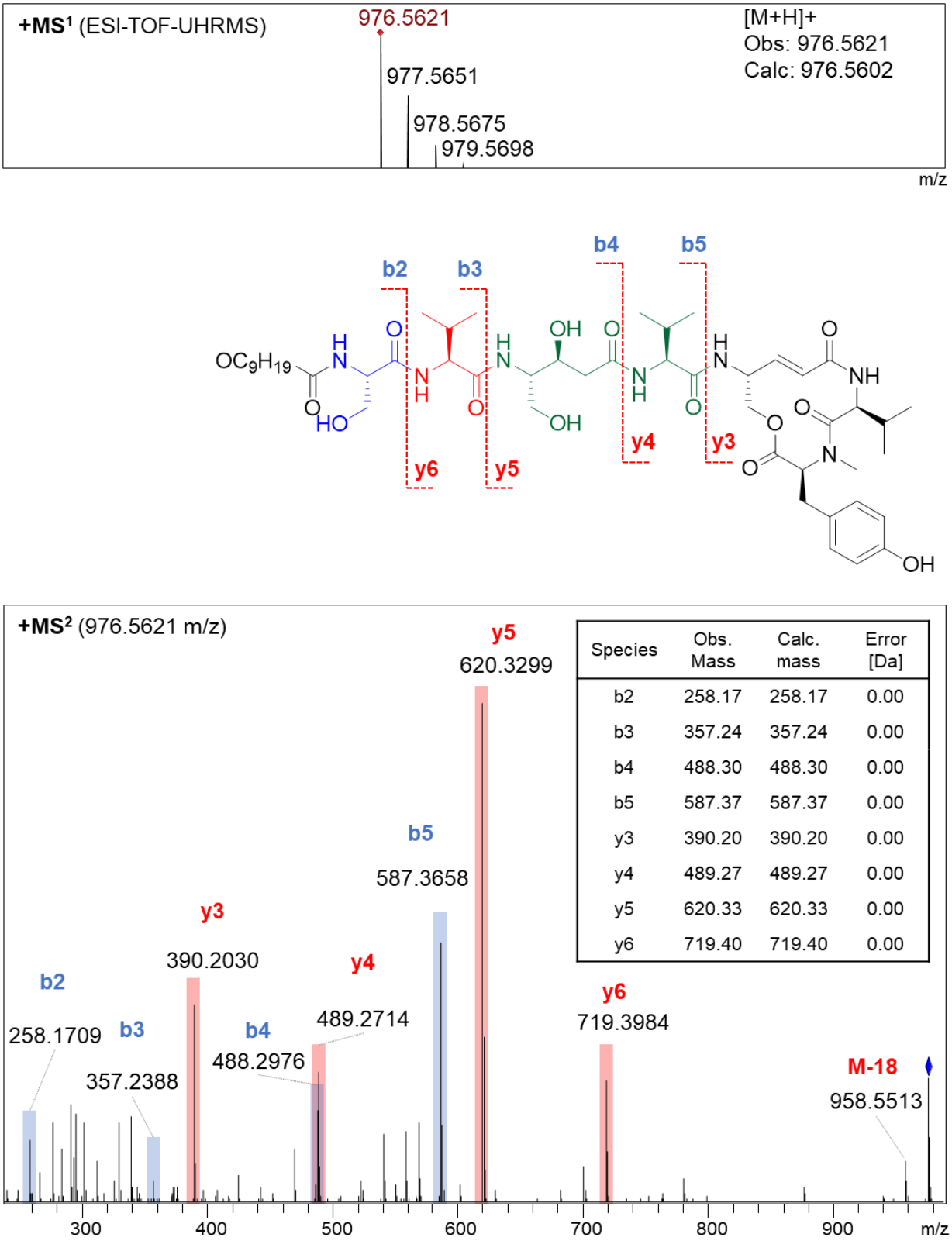


pCB-*ttc*-int / NdeI (1-12)
25019 bp
8200
4796
4561
3271
3233
1383
922
908



pCB-*ttm*-int / NdeI (13-18)
16512 bp
6567
5220
4561

Figure 4.20 Verification of edited constructs. a) List of constructs generated. b) Sanger sequencing results confirming mutations. c) Restriction digestion verification. L, ladder (GeneRuler 1kb Plus).



Species	Obs. Mass	Calc. mass	Error [Da]
b2	258.17	258.17	0.00
b3	357.24	357.24	0.00
b4	488.30	488.30	0.00
b5	587.37	587.37	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00
y5	620.33	620.33	0.00
y6	719.40	719.40	0.00

Figure 4.21 Characterization of thalassospiramide A12 (**4**). MS^n analysis.

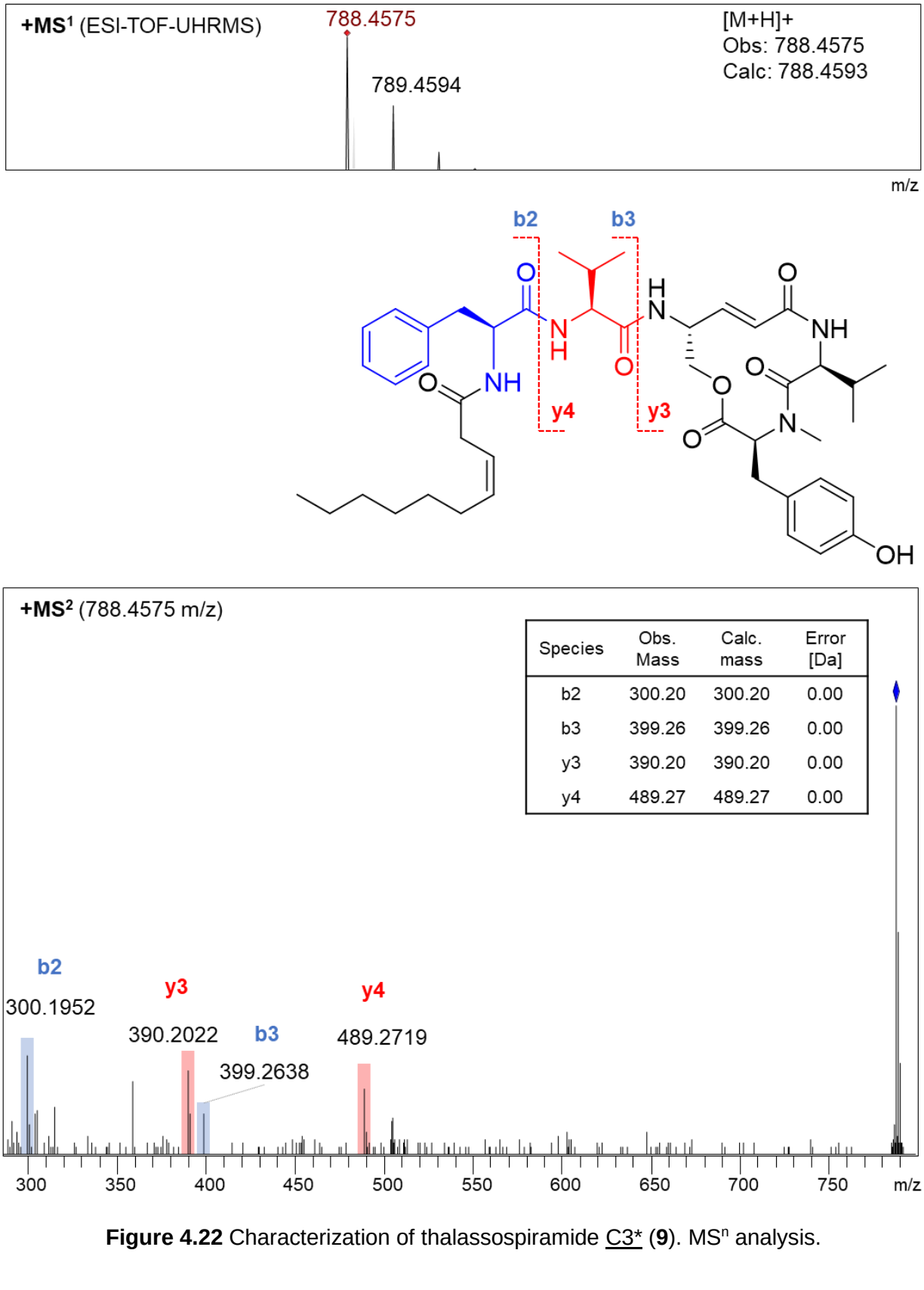
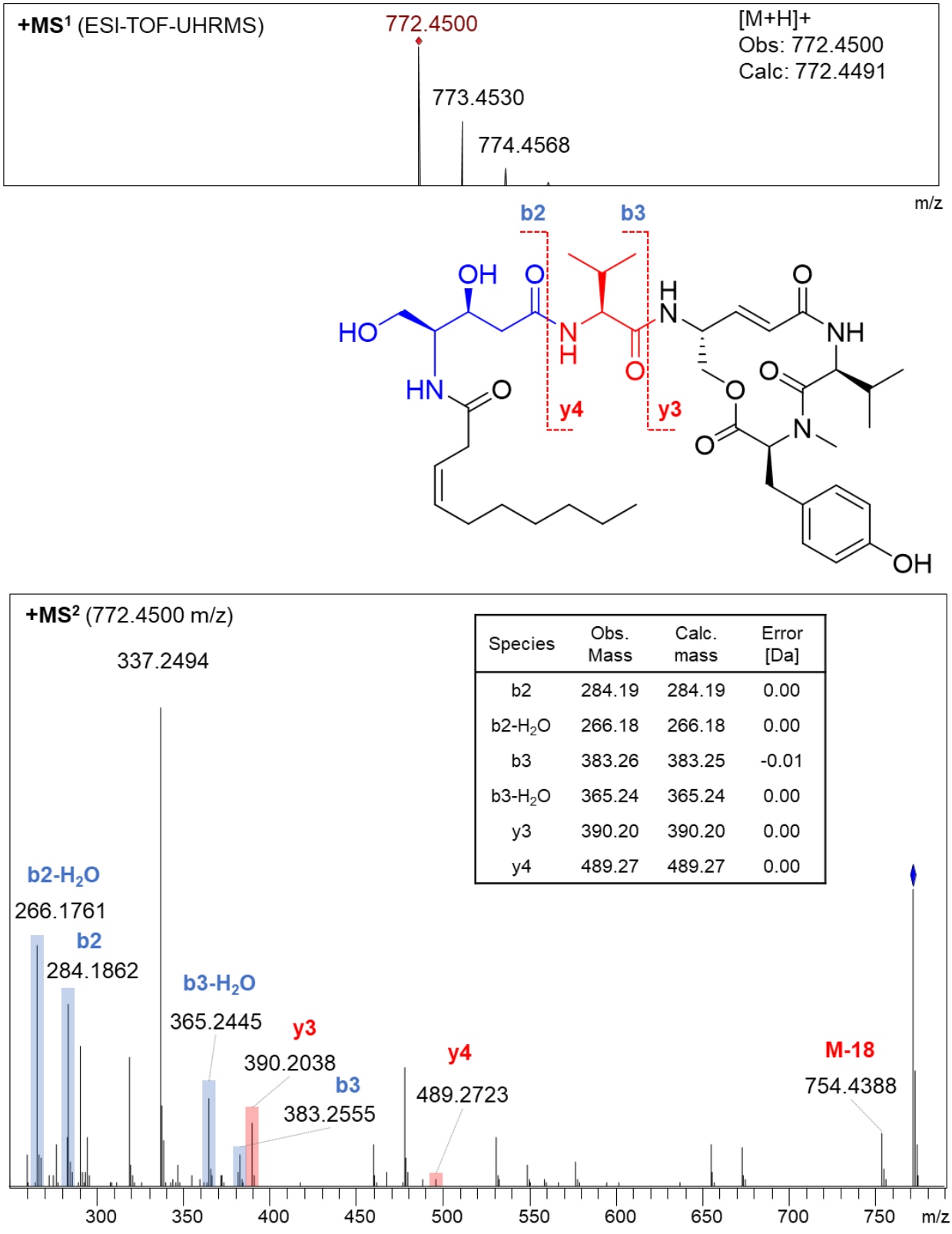
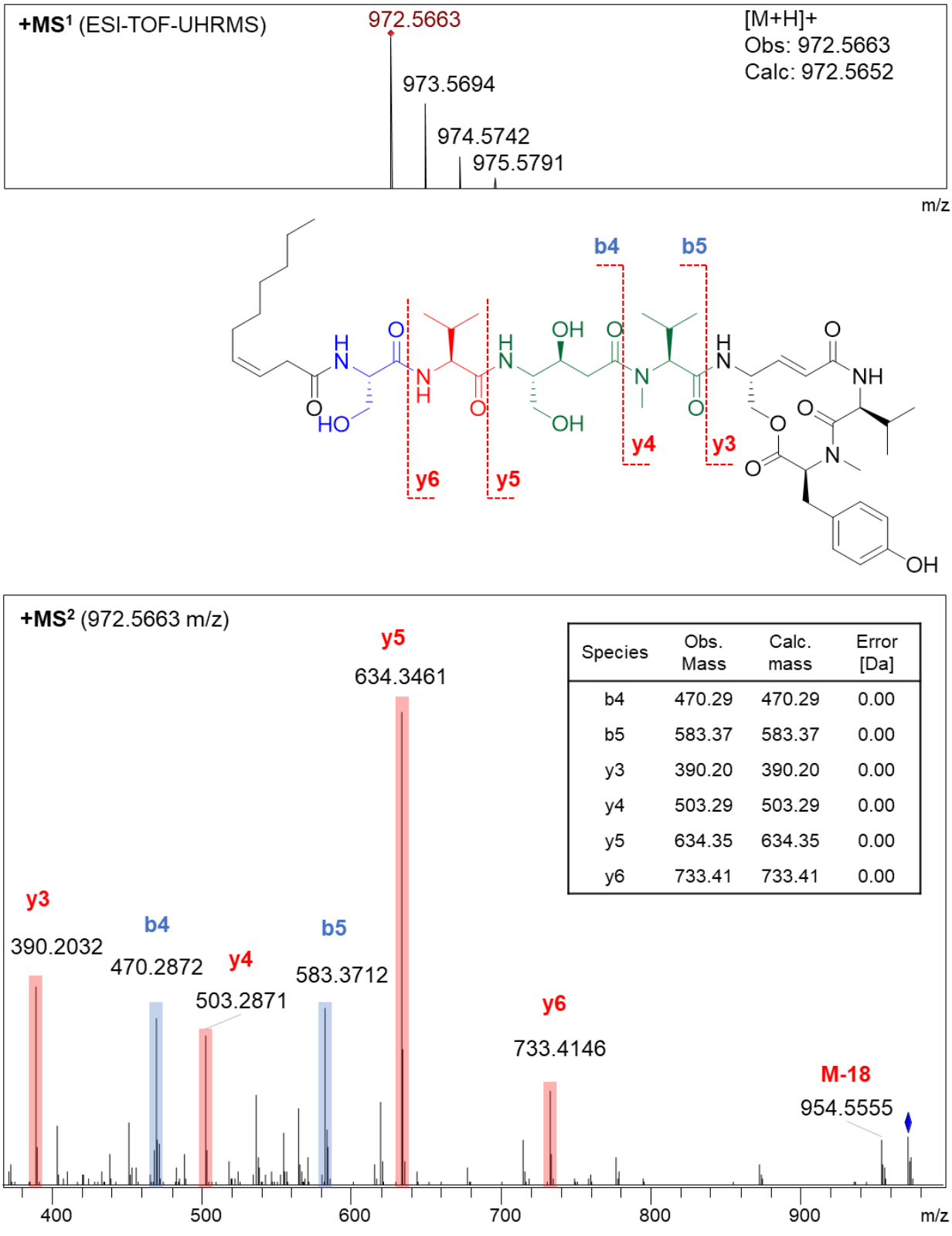


Figure 4.22 Characterization of thalassospiramide C3* (**9**). MS^n analysis.



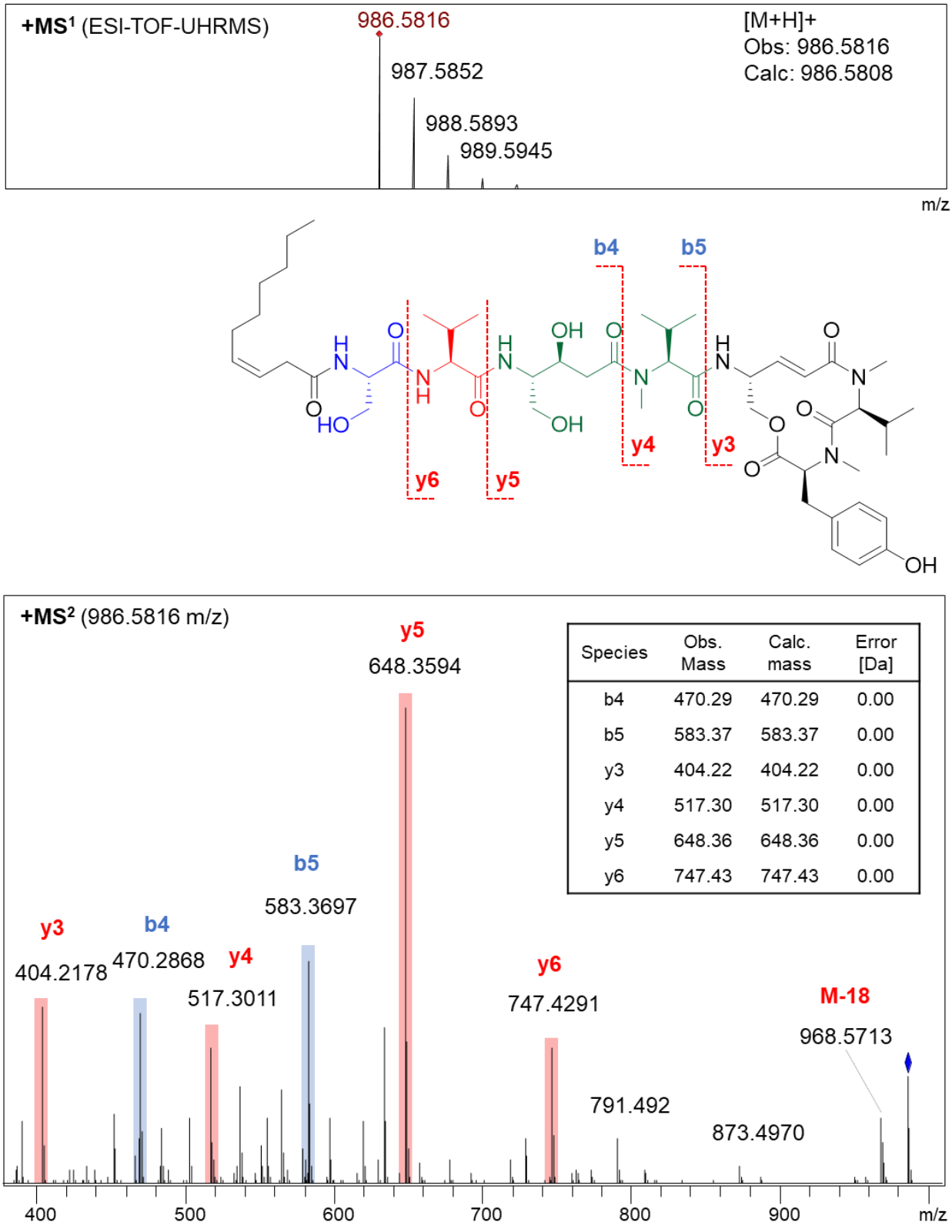
Species	Obs. Mass	Calc. mass	Error [Da]
b2	284.19	284.19	0.00
b2-H_2O	266.18	266.18	0.00
b3	383.26	383.25	-0.01
b3-H_2O	365.24	365.24	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00

Figure 4.23 Characterization of thalassospiramide D2 (**18**). MS^n analysis.



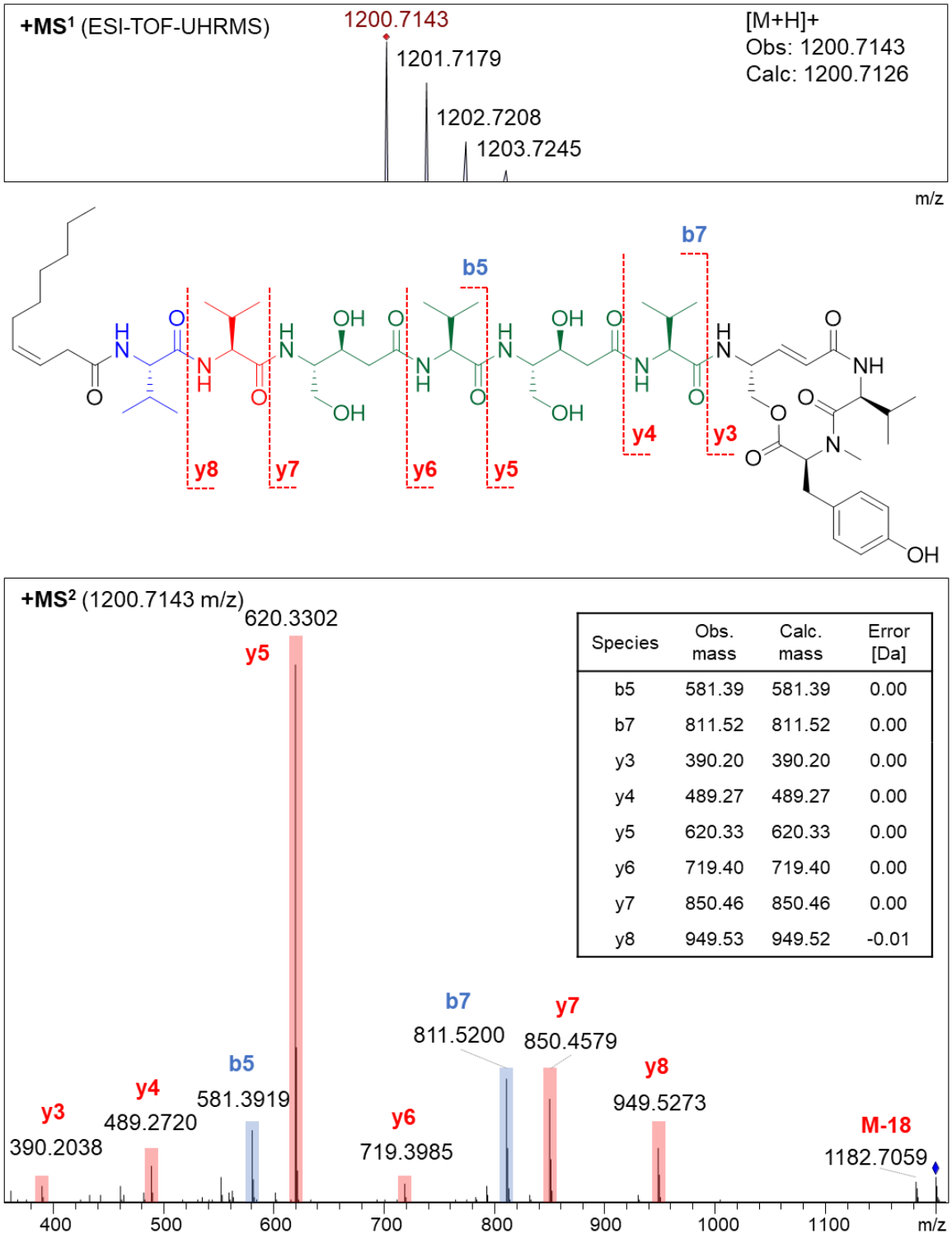
Species	Obs. Mass	Calc. mass	Error [Da]
b4	470.29	470.29	0.00
b5	583.37	583.37	0.00
y3	390.20	390.20	0.00
y4	503.29	503.29	0.00
y5	634.35	634.35	0.00
y6	733.41	733.41	0.00

Figure 4.24 Characterization of thalassospiramide A13 (**19**). MS^n analysis.



Species	Obs. Mass	Calc. mass	Error [Da]
b4	470.29	470.29	0.00
b5	583.37	583.37	0.00
y3	404.22	404.22	0.00
y4	517.30	517.30	0.00
y5	648.36	648.36	0.00
y6	747.43	747.43	0.00

Figure 4.25 Characterization of thalassospiramide A14 (**20**). MS^n analysis.



Species	Obs. mass	Calc. mass	Error [Da]
b5	581.39	581.39	0.00
b7	811.52	811.52	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00
y5	620.33	620.33	0.00
y6	719.40	719.40	0.00
y7	850.46	850.46	0.00
y8	949.53	949.52	-0.01

Figure 4.26 Characterization of thalassospiramide E3 (**22**). MS^n analysis.

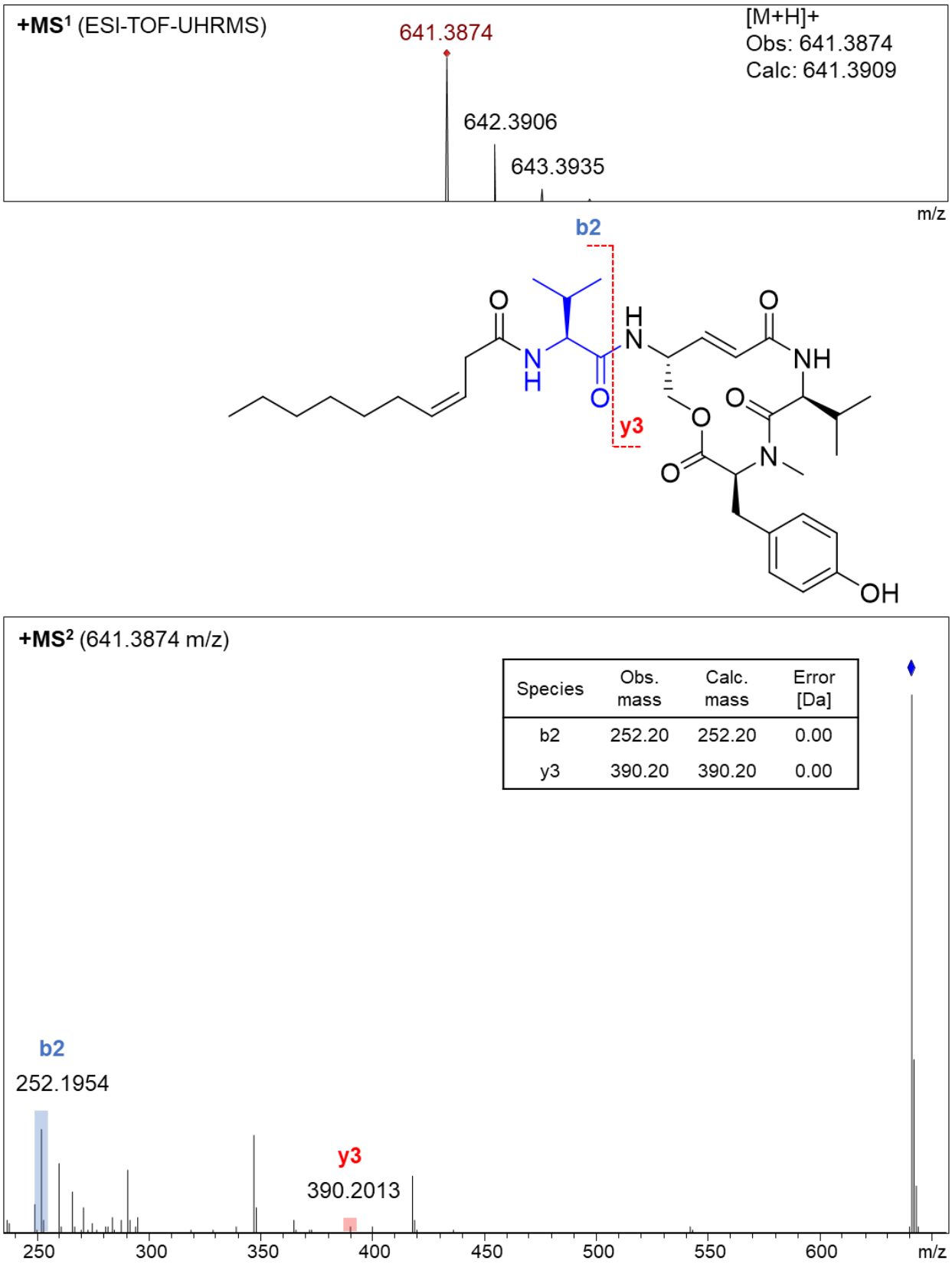
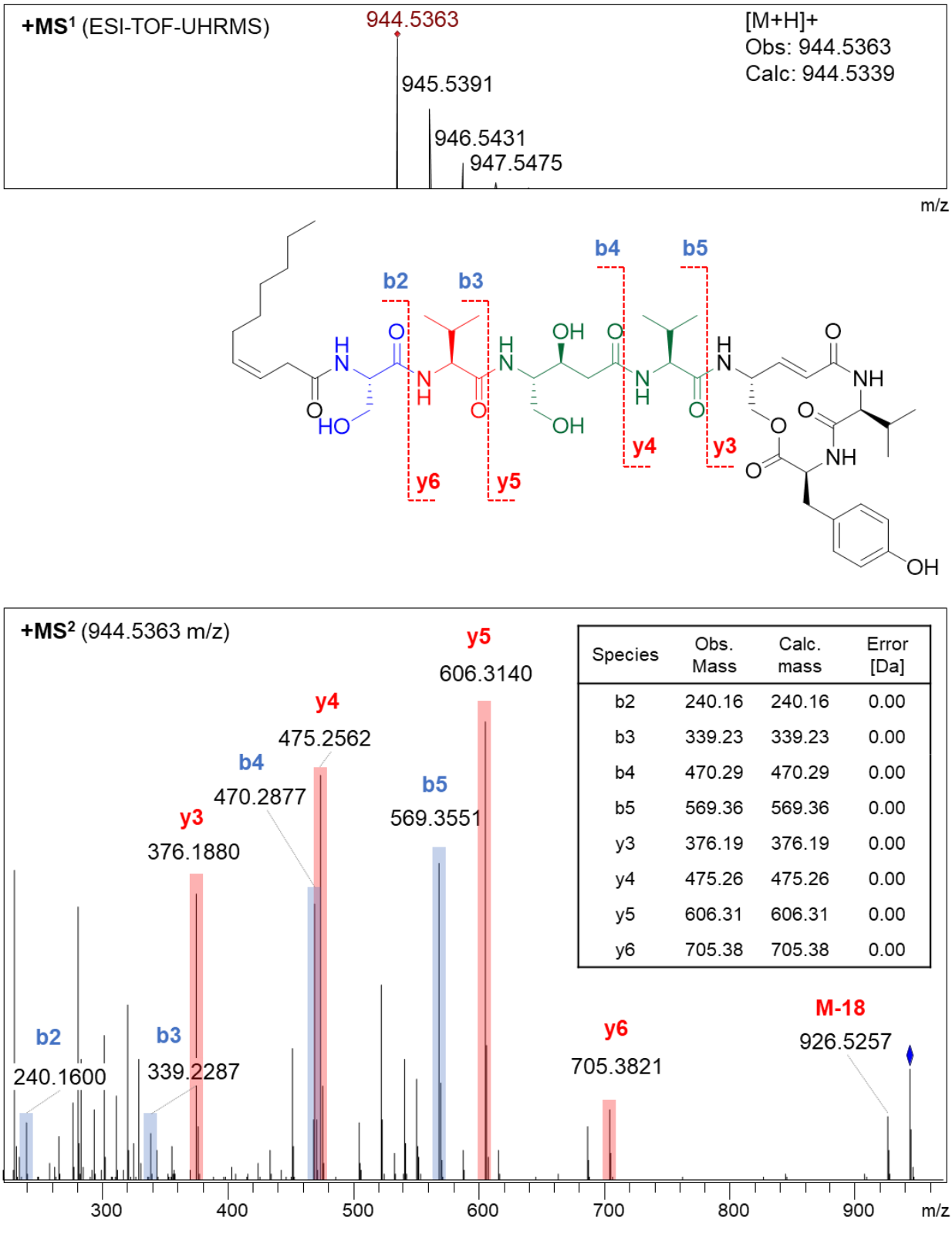


Figure 4.27 Characterization of thalassospiramide E4 (**25**). MS^n analysis.



Species	Obs. Mass	Calc. mass	Error [Da]
b2	240.16	240.16	0.00
b3	339.23	339.23	0.00
b4	470.29	470.29	0.00
b5	569.36	569.36	0.00
y3	376.19	376.19	0.00
y4	475.26	475.26	0.00
y5	606.31	606.31	0.00
y6	705.38	705.38	0.00

Figure 4.28 Characterization of thalassospiramide A15 (**26**). MS^n analysis.

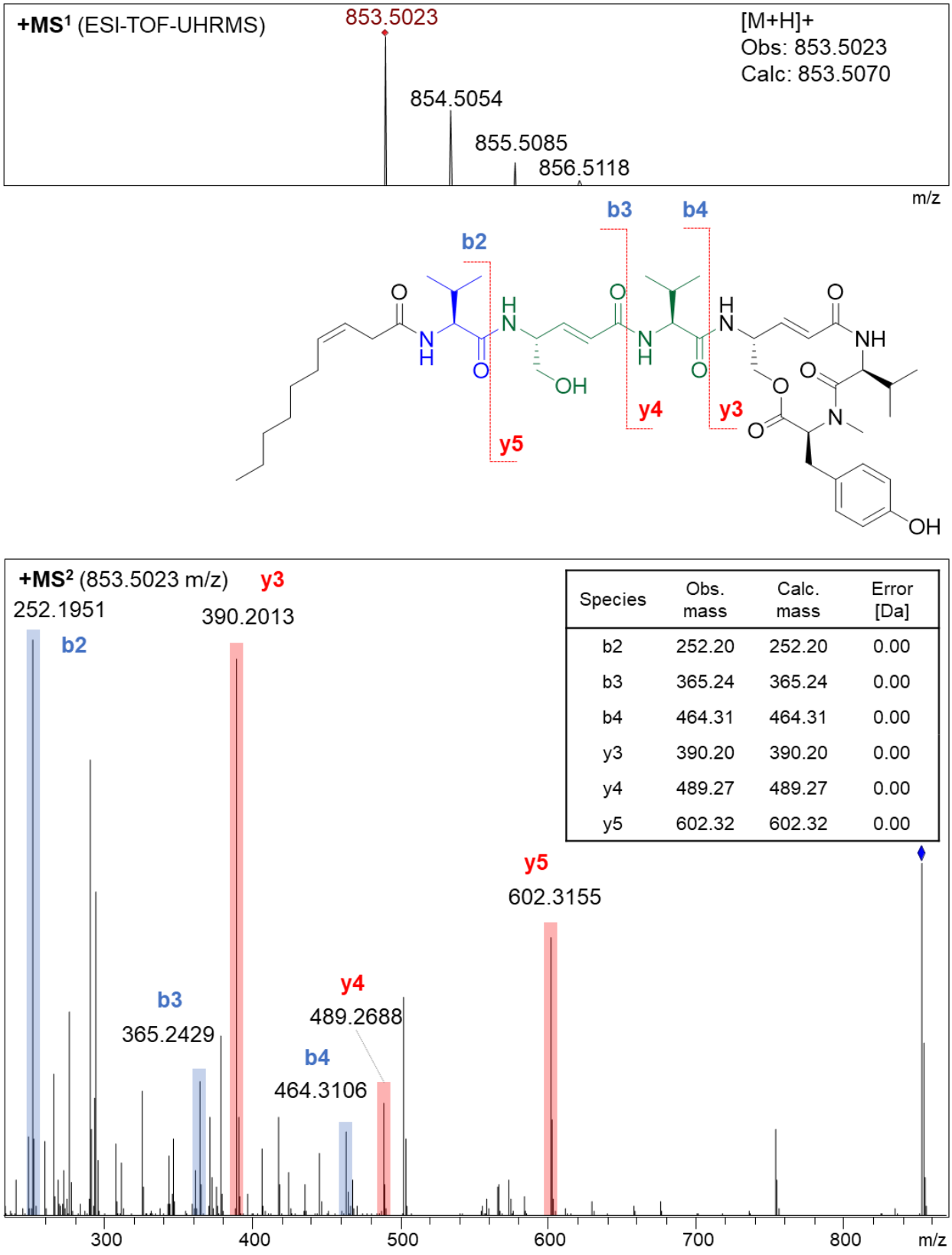
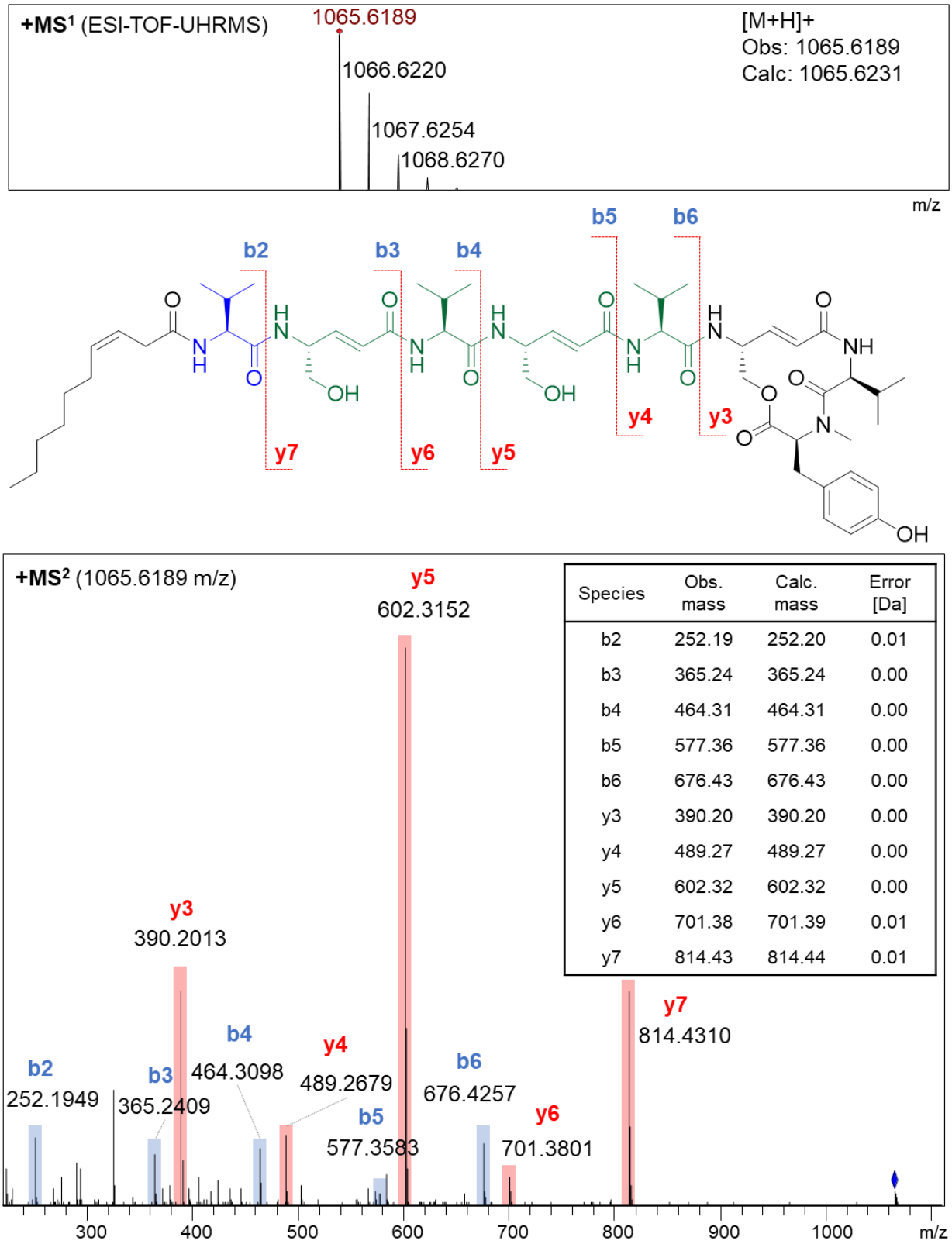
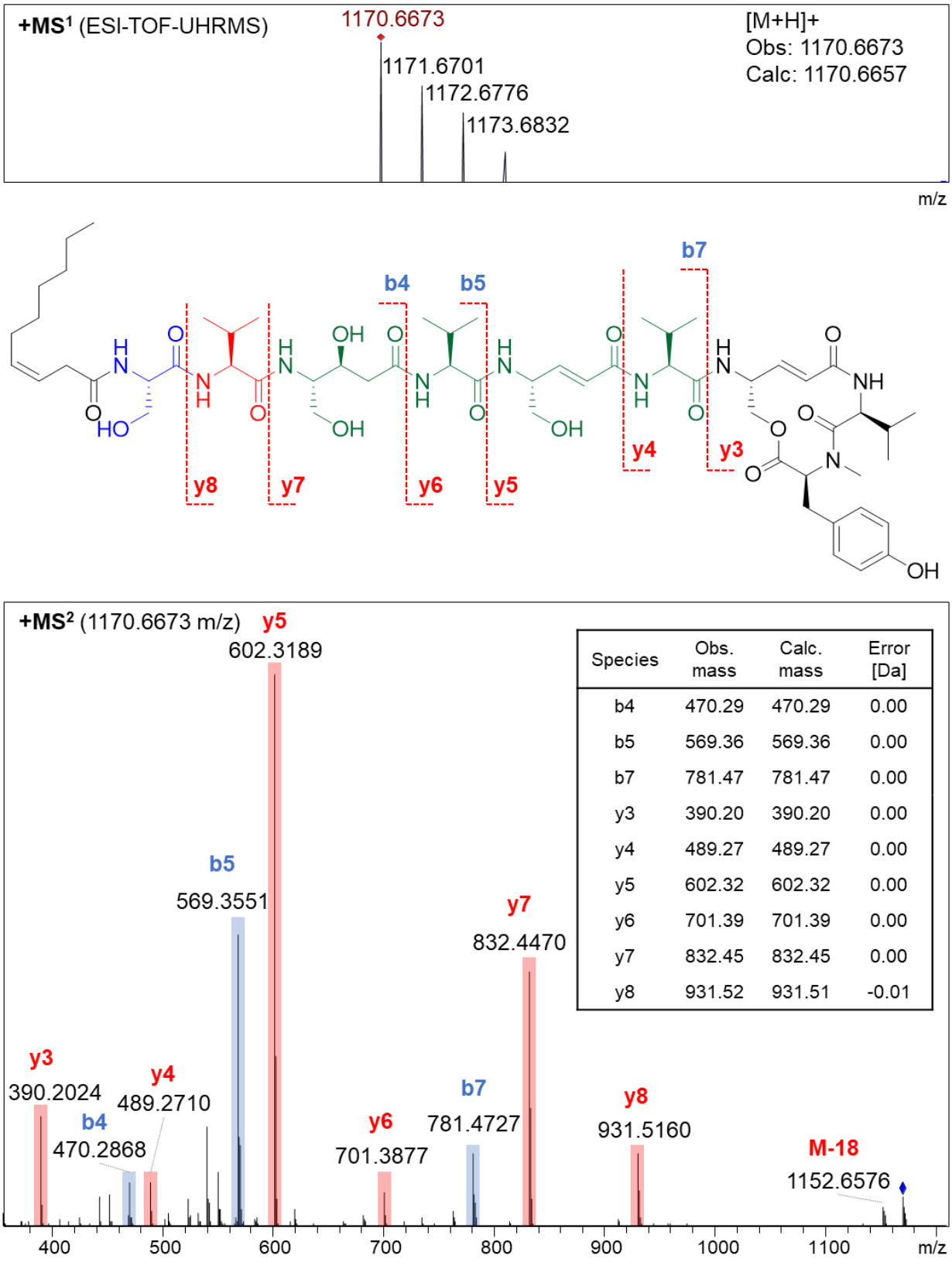


Figure 4.29 Characterization of thalassospiramide E5 (**27**). MS^n analysis.



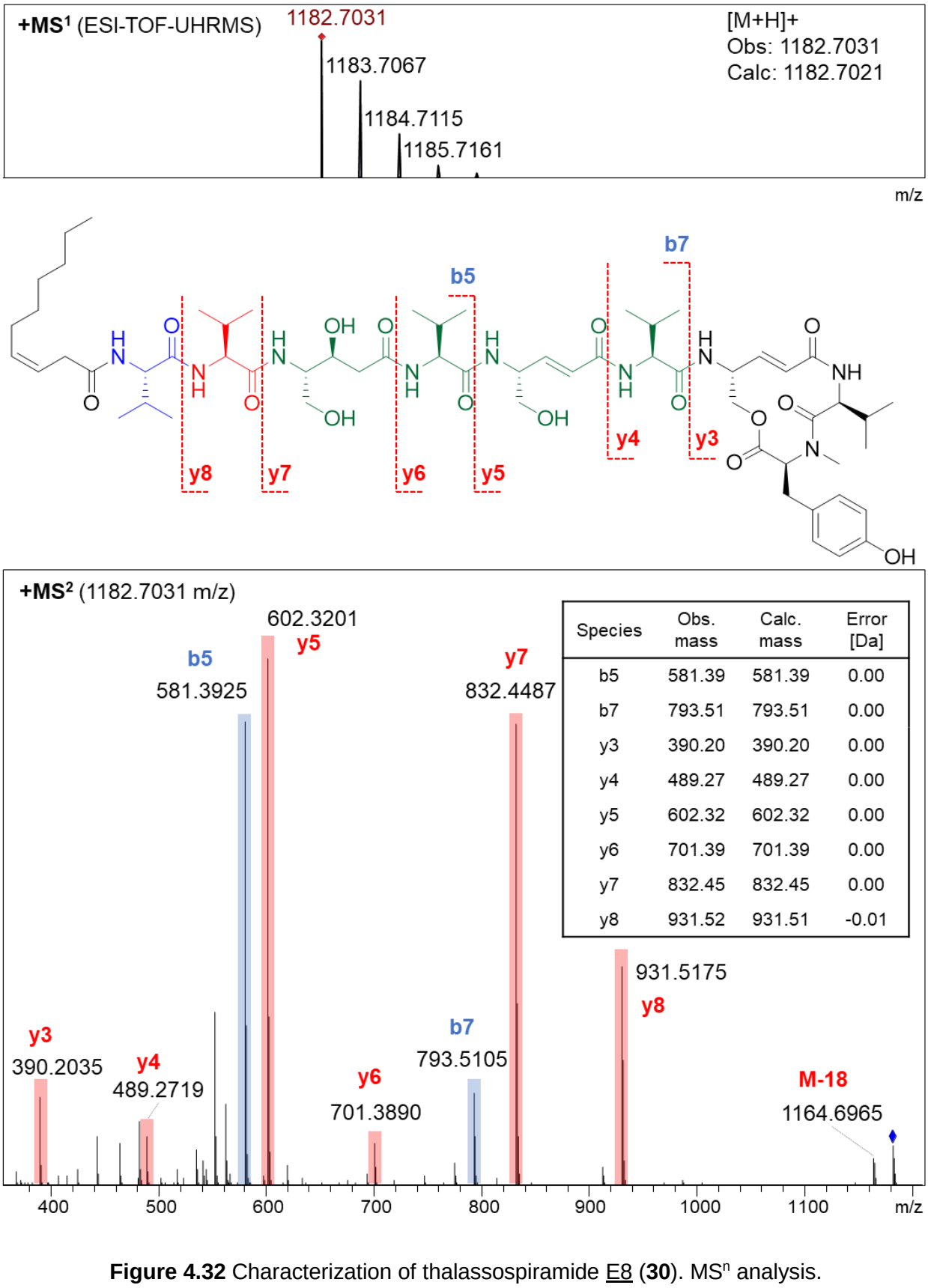
Species	Obs. mass	Calc. mass	Error [Da]
b2	252.19	252.20	0.01
b3	365.24	365.24	0.00
b4	464.31	464.31	0.00
b5	577.36	577.36	0.00
b6	676.43	676.43	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00
y5	602.32	602.32	0.00
y6	701.38	701.39	0.01
y7	814.43	814.44	0.01

Figure 4.30 Characterization of thalassospiramide E6 (**28**). MS^n analysis.



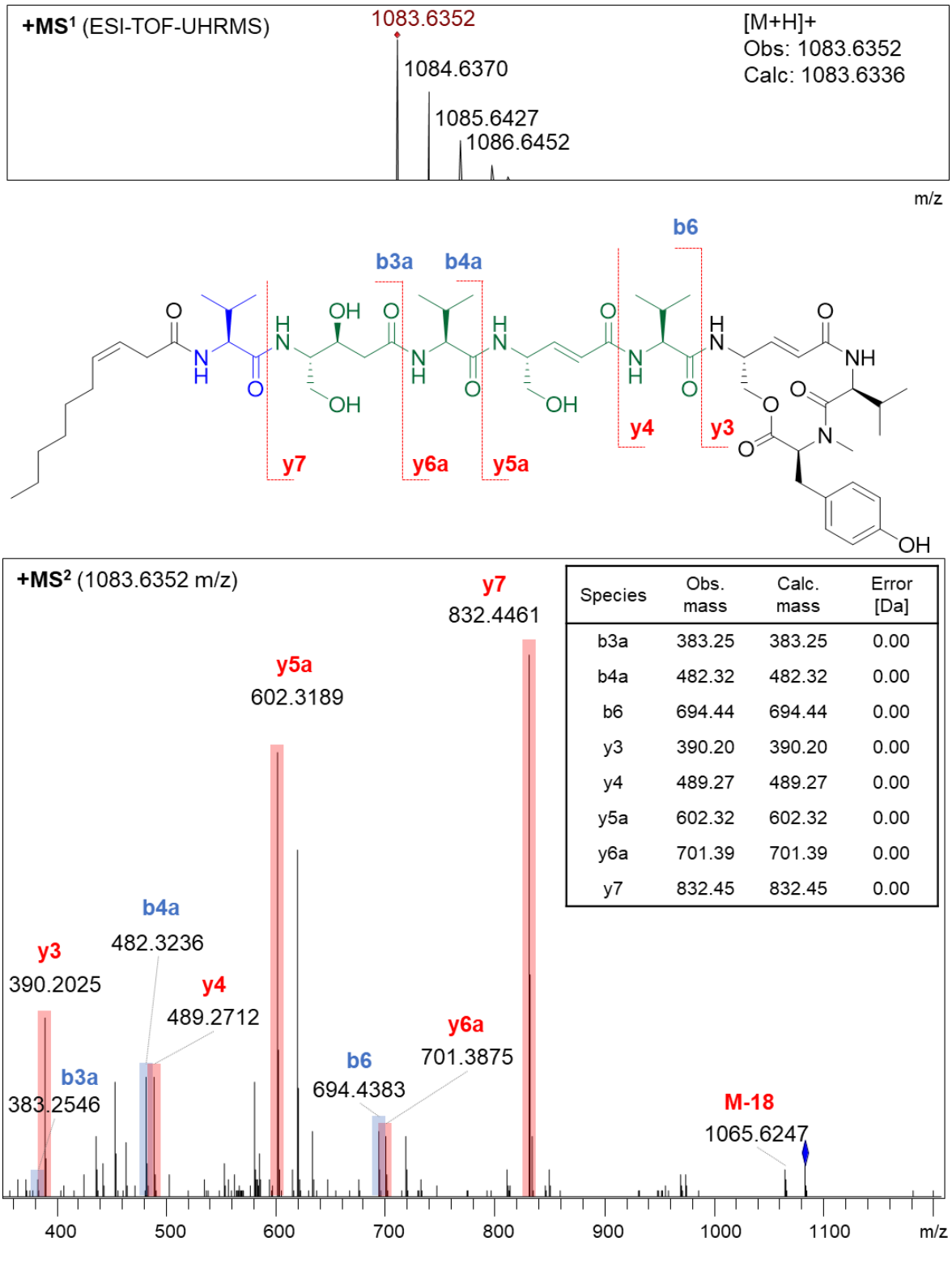
Species	Obs. mass	Calc. mass	Error [Da]
b4	470.29	470.29	0.00
b5	569.36	569.36	0.00
b7	781.47	781.47	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00
y5	602.32	602.32	0.00
y6	701.39	701.39	0.00
y7	832.45	832.45	0.00
y8	931.52	931.51	-0.01

Figure 4.31 Characterization of thalassospiramide E7 (**29**). MS^n analysis.



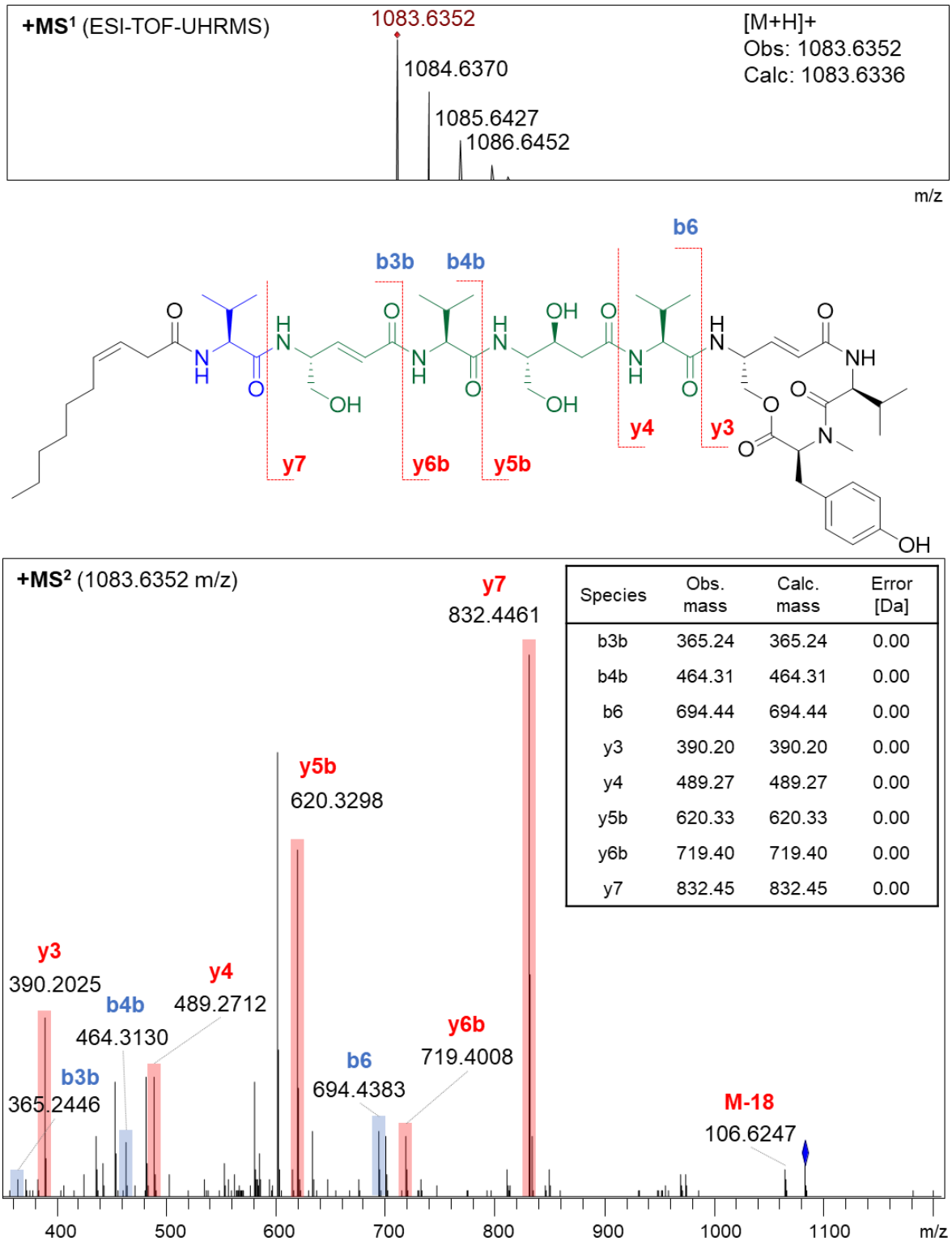
Species	Obs. mass	Calc. mass	Error [Da]
b5	581.39	581.39	0.00
b7	793.51	793.51	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00
y5	602.32	602.32	0.00
y6	701.39	701.39	0.00
y7	832.45	832.45	0.00
y8	931.52	931.51	-0.01

Figure 4.32 Characterization of thalassospiramide E8 (**30**). MS^n analysis.



Species	Obs. mass	Calc. mass	Error [Da]
b3a	383.25	383.25	0.00
b4a	482.32	482.32	0.00
b6	694.44	694.44	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00
y5a	602.32	602.32	0.00
y6a	701.39	701.39	0.00
y7	832.45	832.45	0.00

Figure 4.33 Characterization of thalassospiramide E9a (**31a**). MS^n analysis. (NOTE: same spectra as E9b)



Species	Obs. mass	Calc. mass	Error [Da]
b3b	365.24	365.24	0.00
b4b	464.31	464.31	0.00
b6	694.44	694.44	0.00
y3	390.20	390.20	0.00
y4	489.27	489.27	0.00
y5b	620.33	620.33	0.00
y6b	719.40	719.40	0.00
y7	832.45	832.45	0.00

Figure 4.34 Characterization of thalassospiramide E9b (**31b**). MS^n analysis. (NOTE: same spectra as E9a)

4.8 Acknowledgements

I thank J.R. van der Meer (University of Lausanne) for providing plasmid pRMR6K-Gm, V. de Lorenzo (National Center for Biotechnology-CSIC) for providing strain *P. putida* EM383, and D.L. Court (National Cancer Institute, NIH) for providing strain *E. coli* HME68. I am grateful to A. Edlund (J. Craig Venter Institute), P.Y. Qian (Hong Kong University of Science and Technology), H. Xia (Shanghai Institutes for Biological Sciences, CAS), and W. Fenical and P.R. Jensen (Scripps Institution of Oceanography, UCSD) for facilitating access to equipment, chemical standards, and bacterial strains. I also thank Y. Kudo, P.A. Jordan, J.R. Chekan, L.T. Hoang, and S. Carreto for helpful discussion and technical assistance.

Chapter 4, in part, has been submitted for publication of the material as it may appear in *Nature Chemical Biology*, Zhang, J. J.; Tang, X.; Huan, T.; Ross, A. C.; Moore, B. S., 2019. The dissertation author was the primary investigator and author of this paper.

4.9 References

1 Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11,** 660-670, doi:10.1038/nchembio.1883 (2015).

2 Cane, D. E. Programming of erythromycin biosynthesis by a modular polyketide synthase. *J. Biol. Chem.* **285,** 27517-27523, doi:10.1074/jbc.R110.144618 (2010).

3 Robbel, L. & Marahiel, M. A. Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J. Biol. Chem.* **285,** 27501-27508, doi:10.1074/jbc.R110.128181 (2010).

4 Helfrich, E. J. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33,** 231-316, doi:10.1039/c5np00125k (2016).

5 Sussmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem. Int. Ed. Engl.* **56,** 3770-3821, doi:10.1002/anie.201609079 (2017).

6 Magarvey, N. A., Haltli, B., He, M., Greenstein, M. & Hucul, J. A. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrob. Agents Chemother.* **50,** 2167-2177, doi:10.1128/AAC.01545-05 (2006).

7 Felnagle, E. A., Rondon, M. R., Berti, A. D., Crosby, H. A. & Thomas, M. G. Identification of the biosynthetic gene cluster and an additional gene for resistance to the

antituberculosis drug capreomycin. *Appl. Environ. Microbiol.* **73,** 4162-4170, doi:10.1128/AEM.00485-07 (2007).

8 Thomas, M. G., Chan, Y. A. & Ozanick, S. G. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob. Agents Chemother.* **47,** 2823-2830 (2003).

9 Du, L., Sanchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7,** 623-642 (2000).

10 Gehring, A. M., DeMoll, E., Fetherston, J. D., Mori, I., Mayhew, G. F., Blattner, F. R., Walsh, C. T. & Perry, R. D. Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5,** 573-586 (1998).

11 Oh, D. C., Strangman, W. K., Kauffman, C. A., Jensen, P. R. & Fenical, W. Thalassospiramides A and B, immunosuppressive peptides from the marine bacterium *Thalassospira* sp. *Org. Lett.* **9,** 1525-1528, doi:10.1021/ol070294u (2007).

12 Ross, A. C., Xu, Y., Lu, L., Kersten, R. D., Shao, Z., Al-Suwailem, A. M., Dorrestein, P. C., Qian, P. Y. & Moore, B. S. Biosynthetic multitasking facilitates thalassospiramide structural diversity in marine bacteria. *J. Am. Chem. Soc.* **135,** 1155-1162, doi:10.1021/ja3119674 (2013).

13 Zhang, W., Lu, L., Lai, Q., Zhu, B., Li, Z., Xu, Y., Shao, Z., Herrup, K., Moore, B. S., Ross, A. C. & Qian, P. Y. Family-wide structural characterization and genomic comparisons decode the diversity-oriented biosynthesis of thalassospiramides by marine Proteobacteria. *J. Biol. Chem* **291,** 27228-27238, doi:10.1074/jbc.M116.756858 (2016).

14 Miyazaki, R. & van der Meer, J. R. A new large-DNA-fragment delivery system based on integrase activity from an integrative and conjugative element. *Appl. Environ. Microbiol.* **79,** 4440-4447, doi:10.1128/AEM.00711-13 (2013).

15 Martinez-Garcia, E., Nikel, P. I., Aparicio, T. & de Lorenzo, V. *Pseudomonas* 2.0: genetic upgrading of *P. putida* KT2440 as an enhanced host for heterologous gene expression. *Microb. Cell Fact.* **13,** 159, doi:10.1186/s12934-014-0159-3 (2014).

16 Zhang, J. J., Tang, X., Zhang, M., Nguyen, D. & Moore, B. S. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. *MBio* **8,** e01291, doi:10.1128/mBio.01291-17 (2017).

17 Labby, K. J., Watsula, S. G. & Garneau-Tsodikova, S. Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **32,** 641-653, doi:10.1039/c4np00120f (2015).

18 Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7,** 78, doi:10.1186/1471-2148-7-78 (2007).

19 Crosby, J. & Crump, M. P. The structural role of the carrier protein--active controller or passive carrier. *Nat. Prod. Rep.* **29,** 1111-1137, doi:10.1039/c2np20062g (2012).

20 Calderone, C. T., Bumpus, S. B., Kelleher, N. L., Walsh, C. T. & Magarvey, N. A. A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. *Proc. Natl. Acad. Sci. U S A* **105,** 12809-12814, doi:10.1073/pnas.0806305105 (2008).

21 Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., Narayan, A. R., Hakansson, K., Sherman, D. H., Smith, J. L. & Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510,** 560-564, doi:10.1038/nature13409 (2014).

22 Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Allen, C. L., Skiniotis, G., Aldrich, C. C. & Gulick, A. M. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529,** 235-238, doi:10.1038/nature16163 (2016).

23 Reimer, J. M., Aloise, M. N., Harrison, P. M. & Schmeing, T. M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529,** 239-242, doi:10.1038/nature16503 (2016).

24 Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C. & Gulick, A. M. Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.* **291,** 22559-22571, doi:10.1074/jbc.M116.746297 (2016).

25 Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33,** 136-140, doi:10.1039/c5np00082c (2016).

26 Tarry, M. J., Haque, A. S., Bui, K. H. & Schmeing, T. M. X-ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure* **25,** 783-793 e784, doi:10.1016/j.str.2017.03.014 (2017).

27 Bozhuyuk, K. A. J., Fleischhacker, F., Linck, A., Wesche, F., Tietze, A., Niesert, C. P. & Bode, H. B. *De novo* design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10,** 275-281, doi:10.1038/nchem.2890 (2018).

28 Dunn, B. J., Watts, K. R., Robbins, T., Cane, D. E. & Khosla, C. Comparative analysis of the substrate specificity of trans- versus cis-acyltransferases of assembly line polyketide synthases. *Biochemistry* **53,** 3796-3806, doi:10.1021/bi5004316 (2014).

29 Awakawa, T., Crusemann, M., Munguia, J., Ziemert, N., Nizet, V., Fenical, W. & Moore, B. S. Salinipyrone and pacificanone are biosynthetic by-products of the rosamicin polyketide synthase. *Chembiochem* **16,** 1443-1447, doi:10.1002/cbic.201500177 (2015).

30 Moss, S. J., Martin, C. J. & Wilkinson, B. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat. Prod. Rep.* **21,** 575-593, doi:10.1039/b315020h (2004).

31 Thomas, I., Martin, C. J., Wilkinson, C. J., Staunton, J. & Leadlay, P. F. Skipping in a hybrid polyketide synthase. Evidence for ACP-to-ACP chain transfer. *Chem. Biol.* **9,** 781-787 (2002).

32 Wenzel, S. C., Meiser, P., Binz, T. M., Mahmud, T. & Muller, R. Nonribosomal peptide biosynthesis: point mutations and module skipping lead to chemical diversity. *Angew. Chem. Int. Ed. Engl.* **45,** 2296-2301, doi:10.1002/anie.200503737 (2006).

33 Mootz, H. D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D. & Marahiel, M. A. Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.* **124,** 10980-10981 (2002).

34 Gao, L., Guo, J., Fan, Y., Ma, Z., Lu, Z., Zhang, C., Zhao, H. & Bie, X. Module and individual domain deletions of NRPS to produce plipastatin derivatives in *Bacillus subtilis*. *Microb. Cell Fact.* **17,** 84, doi:10.1186/s12934-018-0929-4 (2018).

35 Yu, D., Xu, F., Zhang, S. & Zhan, J. Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.* **8,** 15349, doi:10.1038/ncomms15349 (2017).

36 Linne, U. & Marahiel, M. A. Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* **39,** 10439-10447 (2000).

37 Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284,** 486-489 (1999).

38 Lu, L., Meehan, M. J., Gu, S., Chen, Z., Zhang, W., Zhang, G., Liu, L., Huang, X., Dorrestein, P. C., Xu, Y., Moore, B. S. & Qian, P. Y. Mechanism of action of thalassospiramides, a new class of calpain inhibitors. *Sci. Rep.* **5,** 8783, doi:10.1038/srep08783 (2015).

39 Fischbach, M. A. & Clardy, J. One pathway, many products. *Nat. Chem. Biol.* **3,** 353-355, doi:10.1038/nchembio0707-353 (2007).

40 Firn, R. D. & Jones, C. G. Natural products--a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20,** 382-391 (2003).

41 Davidsen, J. M. & Townsend, C. A. In vivo characterization of nonribosomal peptide synthetases NocA and NocB in the biosynthesis of nocardicin A. *Chem. Biol.* **19,** 297-306, doi:10.1016/j.chembiol.2011.10.020 (2012).

42 Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A. W., Smith, C. P. & Micklefield, J. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem. Int. Ed. Engl.* **51,** 7181-7184, doi:10.1002/anie.201202043 (2012).

43 Zhang, J. J., Yamanaka, K., Tang, X. & Moore, B. S. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol.* **621,** 87-110, doi:10.1016/bs.mie.2019.02.026 (2019).

44 Zhang, H., Fang, L., Osburne, M. S. & Pfeifer, B. A. The Continuing Development of E. coli as a Heterologous Host for Complex Natural Product Biosynthesis. *Methods Mol. Biol.* **1401,** 121-134, doi:10.1007/978-1-4939-3375-4_8 (2016).

45 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U S A* **97,** 6640-6645, doi:10.1073/pnas.120163297 (2000).

46 Tang, X., Li, J., Millan-Aguinaga, N., Zhang, J. J., O'Neill, E. C., Ugalde, J. A., Jensen, P. R., Mantovani, S. M. & Moore, B. S. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10,** 2841-2849, doi:10.1021/acschembio.5b00658 (2015).

47 Choi, K. H. & Schweizer, H. P. mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. *Nat. Protoc.* **1,** 153-161, doi:10.1038/nprot.2006.24 (2006).

48 Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321,** 659-663, doi:10.1126/science.1159850 (2008).

49 Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A. & Weissman, K. J. Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chemical Science* **5,** 3081-3095, doi:10.1039/c3sc53511h (2014).

50 Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R., Hakansson, K., Sherman, D. H., Smith, J. L. & Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **510,** 512-517, doi:10.1038/nature13423 (2014).

51 Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J. Mol. Biol.* **426,** 2229-2245, doi:10.1016/j.jmb.2014.03.015 (2014).

52 Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384,** 941-953, doi:10.1016/j.jmb.2008.09.084 (2008).

53 Akey, D. L., Razelun, J. R., Tehranisa, J., Sherman, D. H., Gerwick, W. H. & Smith, J. L. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18,** 94-105, doi:10.1016/j.str.2009.10.018 (2010).

54 Al-Mestarihi, A. H., Villamizar, G., Fernandez, J., Zolova, O. E., Lombo, F. & Garneau-Tsodikova, S. Adenylation and S-methylation of cysteine by the bifunctional enzyme TioN in thiocoraline biosynthesis. *J. Am. Chem. Soc.* **136,** 17350-17354, doi:10.1021/ja510489j (2014).

55 Miller, B. R., Sundlov, J. A., Drake, E. J., Makin, T. A. & Gulick, A. M. Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins* **82,** 2691-2702, doi:10.1002/prot.24635 (2014).

56 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233-239, doi:10.1038/nbt.2508 (2013).

57 Sawitzke, J. A., Thomason, L. C., Bubunenko, M., Li, X., Costantino, N. & Court, D. L. Recombineering: highly efficient in vivo genetic engineering using single-strand oligos. *Methods Enzymol.* **533,** 157-177, doi:10.1016/B978-0-12-420067-8.00010-6 (2013).