BIOSYNTHETIC GENE CLUSTERS GUIDE ANTIBIOTIC DISCOVERY

BIOSYNTHETIC GENE CLUSTERS GUIDE RATIONAL ANTIBIOTIC DISCOVERY FROM ACTINOMYCETES

By ELIZABETH J. CULP, B.Sc.

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy

DESCRIPTIVE NOTE

McMaster University DOCTOR OF PHILOSOPHY (2020) Hamilton, Ontario (Biochemistry and Biomedical Sciences)

TITLE: Biosynthetic gene clusters guide rational antibiotic discovery from Actinomycetes

AUTHOR: Elizabeth J. Culp, B.Sc.

SUPERVISOR: Gerard D. Wright, Ph.D.

NUMBER OF PAGES: xv; 231

Foreword

LAY ABSTRACT

Antibiotics are essential for treating life-threatening infections, but the rise of antibiotic resistance renders them ineffective. To treat these drug-resistant infections, new antibiotics that work in new ways are required. A family of bacteria commonly isolated from soil called Actinomycetes produce most antibiotics we use today, but it has become increasingly difficult to find new antibiotics from this source. My work describes three techniques that can be applied to actinomycetes to help overcome the challenges associated with antibiotic discovery. Specifically, these techniques guide discovery efforts by making use of regions in actinomycete genomes called biosynthetic gene clusters that often encode antibiotics. In doing so, I describe ways to uncover rare antibiotics from actinomycete strains that produce common and uninteresting antibiotics, use antibiotic family trees to discover antibiotics that work in new ways, and apply antibiotic resistance to identify biosynthetic gene clusters likely to act on a certain bacterial target.

ABSTRACT

As the spread of antibiotic resistance threatens our ability to treat infections, avoiding the return of a pre-antibiotic era urgently requires the discovery of novel antibiotics. Actinomycetes, a family of bacteria commonly isolated from soil, are a proven source of clinically useful antibiotics. However, easily identifiable metabolites have been exhausted and the rediscovery of common antibiotics thwarts searches for rarer molecules. Sequencing of actinomycete genomes reveals that they contain far more biosynthetic gene clusters with the potential to encode antibiotics than whose products can be readily observed in the laboratory. The work presented in this thesis revolves around developing approaches to mine these previously inaccessible metabolites as a source of new antibiotics.

First, I describe how inactivation of biosynthetic gene clusters for common antibiotics can uncover rare antibiotics otherwise masked in these strains. By applying CRISPR-Cas9 to knockout genes encoding nuisance antibiotics, I develop a simple strategy to reveal the hidden biosynthetic potential of actinomycete strains that can be used to discover rare or novel antibiotics.

Second, I describe the use of the evolutionary history of biosynthetic gene clusters to prioritize divergent members of an antibiotic family, the glycopeptide antibiotics, that are likely to possess new biological activities. Using these predictions, I uncover a novel functional class of glycopeptide antibiotics that blocks the action of autolysins, essential peptidoglycan hydrolases required for remodelling the cell wall during growth.

Finally, I apply target-directed genome mining, which makes use of target duplication as a predicted resistance mechanism within an antibiotic's biosynthetic gene cluster. Using this approach, I discover the association of a family of gene clusters with the housekeeping protease ClpP and characterize the produced metabolite's effect on ClpP function.

These three research projects mine previously inaccessible chemical matter from a proven source of antibiotics, actinomycetes. The techniques and antibiotics described are required now more than ever to develop life-saving antibiotics capable of combatting multidrug-resistant pathogens.

ACKNOWLEDGEMENTS

To Gerry – I can't thank you enough for the opportunities, support and advice you've provided. Your encouragement and excitement through the hills and valleys of science always kept me going. Your dedication to excellence inspires the type of scientist I want to be, and your belief in and motivation of your students is the type of mentor I will model myself after.

To Marie and Brian, my committee members – thank you for making my committee meetings productive and motivational, and taking the time for thoughtful contemplation and helpful suggestions about my science.

To the Wright lab members, past and present – I have you to thank for everything in this thesis, but also for making the journey so much fun. Thank you all. Especially, and in no particular order…Nick – most of my projects wouldn't exist without your insight and work. Wen – thank you for teaching me everything I know about natural product purification, purifying countless compounds for me, and also endless unsolicited advice about real estate. Min – I trust no one more than you for advice about *Streptomyces*. I owe you many favours. Mike – I'll miss our walks to get coffee; a lot of great science ideas came from those. Also a lot of puns. Linda – Thank you for keeping the lab together, from teaching us the ropes to running things smoothly. Kalinka – thanks for everything you do for the lab and being my early morning buddy. Grace – my CRISPR partner in crime. Thanks for spearheading that project, and also for imparting the way you say "OK" and "Oh boy" onto me. Dave – your enthusiasm for science is infectious. Thanks for all the helpful discussions and always having a willingness to lend expertise. Allison, Drew, Haley, Matt, Adam, Caitlyn, Andy, Christian, Emily, Vishwas – you guys are such great friends and made it all a blast. Thanks for all the ridiculous shenanigans.

To my spin crew – Thanks for the sweat puddles, heavy beats, and always giving me somewhere to retreat to.

To my awesome family – Thanks for listening to my complaints, fixing my broken stuff and 9+ years of visiting me in Hamilton. I could not have done this without your endless love and support.

To so many others who made this work possible – Every time I thought I was finished writing this acknowledgments section I thought of more people who helped me along the way. Thank you.

Table of Contents

LIST OF FIGURES

Chapter I

Chapter II

LIST OF TABLES

LIST OF ABBREVIATIONS

ADEP ..... acyldepsipeptide antibiotic
AZC ..... azetidine-2-carboxylic acid
BGC ..... biosynthetic gene cluster
CATCH ..... Cas9-assisted targeting of chromosomes
CORASON ..... CORe analysis of Syntenic Orthologs to prioritize ..... Natural product biosynthetic gene clusters
DiPaC ..... direct pathway cloning
FDAA ..... fluorescent D-amino acids
FMO ..... flavin-containing monooxygenase
GBL ..... γ-butyrolactone
GNPS ..... global natural products social
GPA ..... glycopeptide antibiotics
HADA ..... HCC-amino-D-alanine
HALA ..... HCC-amino-L-alanine
HR ..... homologous recombination
LB ..... Lennox broth
LC ..... liquid chromatography
LLHR ..... linear-plus-linear homologous recombination
MHB ..... Mueller Hinton broth
MIC ..... minimum inhibitory concentration
MRSA ..... methicillin-resistant *Staphylococcus aureus*
MS ..... mass spectrometry
NHEJ ..... non-homologous end joining
NMR ..... nuclear magnetic resonance
NP ..... natural product
NRPS ..... non-ribosomal peptide synthetase
PCA ..... principle component analysis
PG ..... peptidoglycan
PKS ..... polyketide synthetase
RiPP ..... ribosomally synthesized and post-translationally modified peptide
RM ..... restriction modification
RT-PCR ..... reverse-transcriptase polymerase chain reaction
SAM ..... S-adenosylmethionine
TAR ..... transformation associated recombination
TSB ..... tryptone-soya broth
TEM ..... transmission electron microscopy
UDP-MurNAc-PP ..... uridine 5'-diphosphate-*N*-acetylmuramic acid-pentapeptide
VRE ..... vancomycin resistant *Enterococcus*
VISA ..... vancomycin intermediate *Staphyloccocus aureus*

DECLARATION OF ACADEMIC ACHIEVEMENT

I have performed all the research in this body of work except where indicated in the preface of each chapter

Chapter I – Introduction

NATURAL PRODUCTS AS A SOURCE OF ANTIBIOTICS

Antibiotics are a cornerstone of modern healthcare that are vital for treating bacterial infections in humans, animals, and crops[1]. Inextricably linked to the use of antibiotics is the evolution of resistance in disease-causing bacteria, and the rise of this resistance increasingly results in multidrug resistant pathogens for which there are few, if any, treatment options. The health, economic and social impacts of antibiotic resistance are widespread, requiring multipronged solutions to tackle the many aspects of this growing crisis[1]. Chief among these solutions is the discovery of novel antibiotics capable of treating multidrug-resistant pathogens.

The "Golden Age" of antibiotic discovery was first sparked by Alexander Fleming's discovery of penicillin in 1929 and officially commenced with its widespread use in 1942[2]. Over the next 20 years, the Waksman platform, whereby natural product (NP) extracts are screened for growth inhibitory activity against an indicator strain, would uncover new antibiotics with relative ease[3]. However, as rediscovery of known compounds became an increasing burden, in the 1960s the field moved away from NPs. Instead focusing on medicinal chemistry of known scaffolds or target-based drug discovery based on *in vitro* screens of synthetic chemical libraries, no new classes of antibiotics were brought to the clinic during this period[3,4]. The landscape for antibiotic discovery has changed since the Golden Age, with few pharmaceutical companies remaining in the R&D space and the responsibility shifting to academic labs and small companies. Armed with innovative ideas, modern technology and the ability to take on high-risk approaches, NPs once again are being tapped as a source of novel antibiotics.

NPs are a tremendously diverse family of specialized metabolites gifted with high chemical complexity and biological activity, including antibacterial, antifungal, antihelmintic and anticancer properties (Fig. 1). Interest in their medicinal use has driven the identification of >33,000 bioactive NPs, close to a third of which are produced by Actinobacteria, including the champion genus, *Streptomyces*[5]. Impressively, 76% of all NP scaffolds used as antimicrobial therapeutics are derived from actinomycetes[6], making them a rich source for future discovery.

BIOSYNTHETIC GENE CLUSTERS

The machinery required for the biosynthesis of specialized metabolites by bacteria are encoded together in genomic regions termed biosynthetic gene clusters (BGC). BGCs encode all the necessary biosynthetic, transport and regulatory genes required for the production of a specialized metabolite, and in the case of antibiotics, also include resistance determinants. Several bioinformatic tools have been developed over the past decade for the prediction and characterization of BGCs (reviewed previously [7,8]). High confidence BGC discovery is achieved by making use of query proteins and domains as signatures of a NP class, allowing for rapid detection of BGCs belonging to known families[7]. This strategy is employed by the accessible and widely-used antiSMASH platform[9], allowing researchers without computational expertise to quickly annotate BGCs in a given genome. Adding to this foundational tool, additional computational techniques are available with various applications in targeted research. For example, new BGCs without similarity to known classes can be predicted (e.g. ClusterFinder and EvoMining algorithms[10,11]), BGCs can be grouped into families in order to assess

novelty[12], and curated repositories of BGCs (e.g. MiBIG [13]) can fuel research on the discovery, biosynthesis, ecology and chemistry of these specialized metabolites.

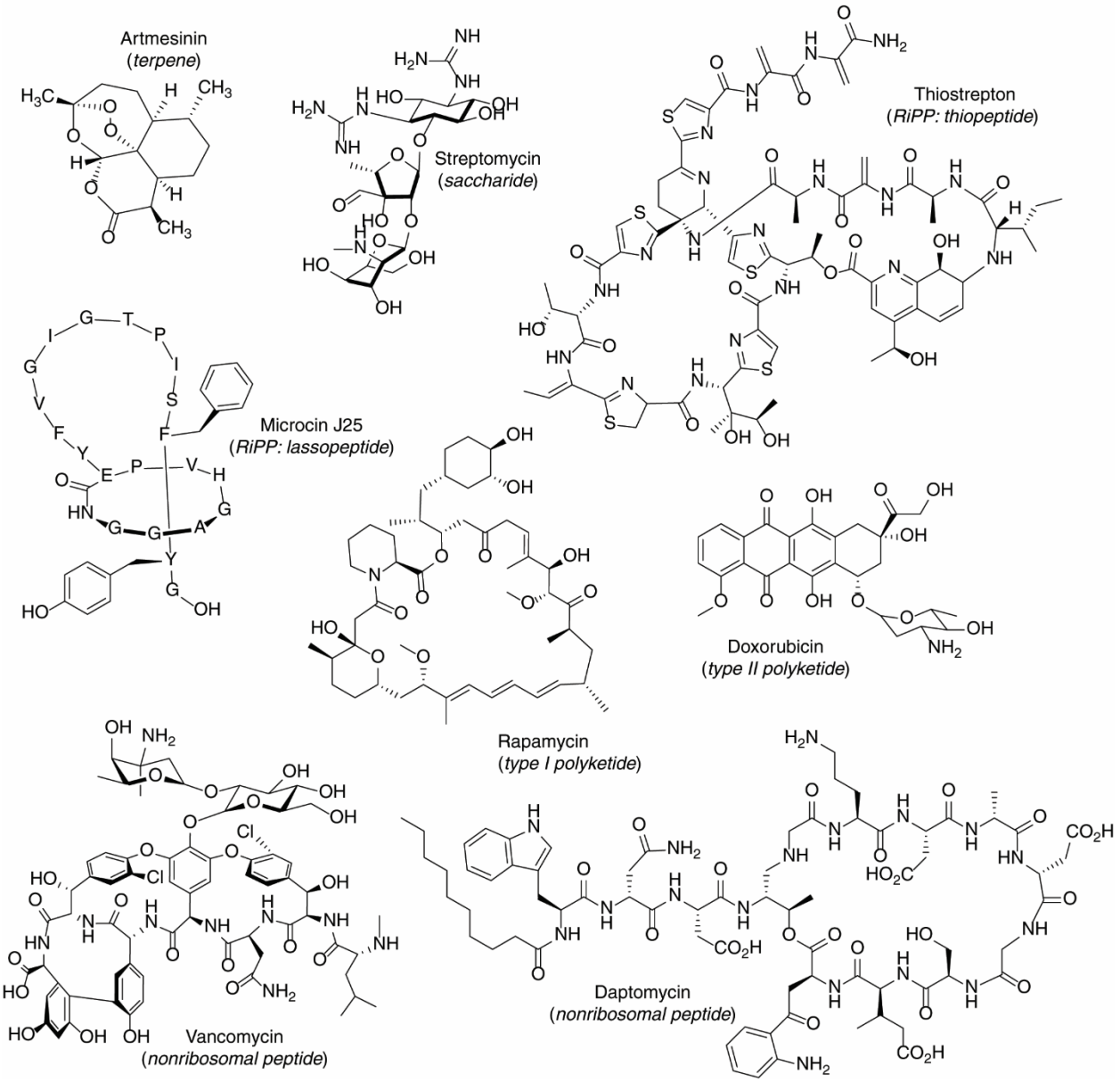


Figure 1. Representative members of major BGC classes
Major natural product classes shown include clinically relevant members such as antibiotics (streptomycin, vancomycin, daptomycin), antimalarials (artmesinin), anticancers (doxorubicin) and immunosuppressants (rapamycin).

Key natural product classes

There are many different classes of NPs, each including clinically relevant compounds of diverse functions. Some of the most common classes of NPs are terpenoids, derived from isoprenoid units and including the antimalarial drug artmesinin[14], saccharide encoding BGCs, for example aminoglycoside antibiotics, and ribosomally synthesized and post-translationally modified peptides (RiPPs), which include many subfamilies depending on their structure and biological activity (e.g. microcins, lantipeptides, thiopeptides)(Fig. 1). Furthermore, a large proportion of BGCs are hybrid, containing biosynthetic signatures from multiple classes[10].

Some of the most successful NPs and highly studied biosynthetic classes are non-ribosomal peptide synthetases (NRPS) and polyketide synthetases (PKS). These assembly-line biosynthetic machines are capable of producing remarkable chemical diversity from simple building blocks (reviewed previously[15]). NRPS use amino acid monomers, both proteinogenic as well as non-proteinogenic, that are iteratively condensed into linear oligomers in a modular fashion. Each module consists of an adenylation domain responsible for selecting and activating the amino acid, which becomes covalently tethered to the phosphopantetheinyl arm of an adjacent thiolation domain (or peptidyl carrier protein). Condensation domains catalyze amide bond formation between the downstream aminoacyl amine and upstream peptidyl thioester, and so biosynthesis proceeds in a colinear fashion. The terminal thioester domain cleaves the final product from the NRPS, sometimes with concomitant cyclization. By predicting the

amino acid specificity of each adenylation domain using bioinformatic platforms (e.g. NRPSpredictor [16]), it is possible to anticipate the structure of the encoded peptide.

In the case of PKS systems, building block monomers are acyl-CoAs, typically malonyl or methylmalonyl-CoA[15]. The biosynthetic logic resembles that of NRPS, consisting of acyltransferase domains that select monomers, ketosynthase domains responsible for catalyzing C-C bond formation, thiolation domains tethered to biosynthetic intermediates (also called acyl carrier proteins), and thioesterase domains for product release. PKSs also optionally include ketoreductase, dehydratase and enoylreductase for partial to complete reduction of carbonyl groups. There are three types of PKS, where type I domains are found in *cis*, type II in *trans*, and type III lacking carrier T domains[15].

Even with extensive knowledge of the modular biosynthesis by NRPS and PKS mega-enzymes, exceptions to the rules and extensive post-processing still make the structure of resultant metabolites difficult to predict. Some bioinformatic tools attempt to link BGCs to mass spectrometry (MS) data by combinatorially predicting all possible mass profiles from a given BGC (e.g. PRISM[17]), but these predictions remain challenging. Barring a closely related and characterized BGC, the structures of highly diverse NP classes such as terpenes, alkaloids, saccharides, and hybrid BGCs cannot be predicted with accuracy at this time. Continuing to curate BGC repositories, allowing for comparison to known BGCs, and investigate BGCs of unknown classes will be important to improve our ability to make BGC to NP predictions.

BARRIERS TO DISCOVERY: DEREPLICATION

Despite the rich potential of actinomycetes as a source of clinically useful compounds, there exist several challenges. One of the biggest challenges is the rediscovery of known NPs. With over 75 years worth of easily discoverable antibiotics already identified, novel compounds are increasingly rare. Highlighting the magnitude of this burden, it is estimated >10 million strains would need to be screened using the Waksman platform to find a novel, clinically useful antibiotic[18]. If one factors into account that ~25% of actinomycetes are antibiotic producers, this means that 2.5 million of these 10 million produce a known antibiotic[18]. Identifying and discarding these uninteresting compounds, a process termed dereplication, is prohibitively laborious, unproductive, and clearly infeasible as a discovery platform.

Easing the dereplication burden

A number of creative approaches have been developed in order to circumvent or ease the burden of dereplication. Firstly, platforms that can identify known compounds in crude extracts can allow for dereplication without needing to perform activity-guided purification. These approaches are based either on molecular features determined by analytical spectrometry or activity-based phenotypic fingerprints. Approaches relying on molecular features usually make use of liquid chromatography (LC) hyphenated with MS or nuclear magnetic resonance (NMR) (e.g. LC-MS, tandem LC-MS/MS, LC-NMR, and LC-NMR-MS[19–22]). By comparing the acquired spectra to databases of known compounds, such as the Dictionary of Natural Products[23], MarinLit[24], Antibase[25], ChemSpider[26], PubChem[27], and METLIN[28], known compounds in crude mixtures can be

predicted. One of the largest collaborative and ongoing efforts in this field is the Global Natural Products Social (GNPS)[29], built on LC-MS/MS based molecular networking. The GNPS database brings together community generated mass spectra of NPs, while making use of fragmentation data to provide more detailed fingerprints than molecular mass alone[29]. GNPS is integrated with platforms for dereplication through molecular networking and spectra comparison: VarQuest[30] and DEREPLICATOR+[31] for peptidic or general NPs, respectively. It is also integrated with RiPPquest for comparison of MS/MS spectra to whole genome sequences, allowing for RiPP NP mining[32]. However, these approaches are limited by the quality of the reference database and cannot recognize if a new compound, and therefore new mass fingerprint, belongs to an already well-established class.

In contrast to structurally motivated techniques, phenotype-guided approaches to dereplication have the advantage of identifying the compound in a crude extract that is conclusively responsible for the observed bioactivity. For example, using the activity profile against a range of bacterial pathogens, the BioMAP platform is able to predict the drug class of a given unknown, thereby discarding novel but uninteresting variants of known antibiotics[33]. The Antibiotic Resistance Platform provides even greater precision by individually expressing specific resistance proteins in an isogenic *Escherichia coli* background[34,35]. By arraying the strain collection against crude extracts, common antibiotics/classes can be easily identified. While these phenotype-based approaches are useful for dereplicating the most common classes of antibiotics, in comparison to

analytical approaches, they are limited in their throughput, detection sensitivity, and the number of NP classes included in their platforms.

Exploring unchartered territory for new NPs

Apart from dereplication itself, another strategy for increasing the chances of finding new antibiotics is to likewise look in new places. Indeed, microbes derived from exotic locations including Antarctic or Artic regions[36] or remote caves[37] are often considered as potential producers of new compounds. While the biogeographical distribution of microbes and their gene pools is largely dispersed, specific ecological niches (e.g. sea water, host associated, airborne) do appear to select for distinct organisms and biosynthetic potential[38]. Marine actinomycetes are one such group of organisms that have been empirically shown to contain a high density of unique BGCs, and therefore an underexplored source of NPs[39]. Among the exclusively marine-derived actinomycete genera is *Salinispora*, a proven source of novel NPs including the anticancer agent salinosporamide A, a 20S proteasome inhibitor[40]. NP discovery from marine cyanobacteria and sponges are also fruitful areas of active research[41]. Apart from unique environmental microbes, members of the human microbiome are also beginning to be tapped as a source of bioactive NPs, with intrigue added by hypotheses about their role in microbe-microbe and host-microbe interactions[42]. A variety of novel NPs with various activities, including the antibiotics lugdunin[43], humimycin[44], and ruminococcin C[45], have been discovered from members of the human microbiota.

As another underexplored area of biosynthetic potential, it has long been appreciated that the vast majority of environmental bacteria are not captured by culture-

based methods[46]. Accessing these organisms requires either improved culturing technologies or culture-independent sequence-based approaches. A primary example of the former is the iChip, consisting of miniature diffusion chambers that are inoculated with single environmental cells and incubated *in situ*[47]. In this way, naturally occurring growth factors present in the environment are thought to promote growth, while isolating single cells prevents overgrowth by fast growers. iChip technology was used to cultivate and purify a novel antibiotic, teixobactin, from a previously unculturable organism, *Eleftheria terrae*[48]. Alternatively, circumventing the need to culture bacteria, by constructing cosmid libraries of environmental DNA, intact BGCs can be reconstructed and used for functional heterologous expression (discussed in the next section). Sean Brady and coworkers have been especially successful in this area, combining environmental DNA heterologous expression with various other strategies to enable novel NP discovery[49–52].

Finally, while searching for broad-spectrum antibiotics derived from NPs may yield diminishing returns, by performing clever screens, new compounds and/or their undiscovered activities can be identified. One strategy is to prioritize strains possessing narrow-spectrum activity against a pathogen of interest, and therefore avoid the discovery of more common compounds. Recent examples of novel discovery in this way include promysalin, a Pseudomonad secondary metabolite with selective activity against certain other *Pseudomonas* spp.[53], ibomycin, a *Streptomyces* derived macrocycle with selective activity against *Cryptococcus neoformans*[54], and a number of anti-tubercular compounds targeting uniquely essential targets such as ClpC1[55–57]. Alternatively, compounds can be

identified that do not inhibit growth *in vitro*, but block pathways related to quorum sensing, biofilm formation, and virulence[58,59], or nutrient biosynthesis essential for growth *in vivo*[60]. Screening for antibiotic adjuvants that potentiate the activity of an antibiotic against resistant bacteria can also reveal previously unknown activities even of known NPs. For example, aspergillomarasmine A, a fungal metabolite originally described in the early 1960s, was discovered as a potent inhibitor of clinically relevant metallo-β-lactamases in 2014[61]. Clearly, given a clever approach, phenotype-based screens still have their place in modern discovery of clinically useful compounds.

BARRIERS TO DISCOVERY: CRYPTIC BIOSYNTHETIC GENE CLUSTERS

Another challenge in NP discovery is accessing actinomycetes' full biosynthetic capability. The realization that actinomycetes hold much greater biosynthetic potential than previously appreciated came with sequencing of the first *Streptomyces* genome in 2002, *Streptomyces coelicolor* M145[62]. Known at the time to produce only six specialized metabolites, it came as a surprise when 22 BGCs were identified in its genome[63]. Today, it is commonplace to identify 20-40 BGCs per actinomycete genome, with only a handful of the resulting specialized metabolites detectable through standard laboratory culturing. Genome mining these cryptic BGCs, that is BGCs whose products are unknown, offers a treasure trove for NP discovery.

Pleiotropic approaches to BGC activation

BGCs are often cryptic because they are transcriptionally silent under a given culturing condition. There exist a number of pleiotropic approaches that aim to trigger

global changes in secondary metabolism and activate these silent BGCs (Fig. 2a). Starting with the most empirical of these techniques, altering growth conditions such as nutrient content, temperature, or pH is often enough to elicit production of new metabolites. This framework is sometimes referred to as the "one strain many compounds" (OSMAC) approach[64]. Co-culture of actinomycetes with other bacteria or fungi can also activate BGCs, and appeals to the idea of mimicking interactions that occur naturally in environment[65–69]. A change in nutrient composition or exposure to cell wall fragments (e.g. mycolic acid from mycolata species, *N*-acetylglucosamine from fungi) have been shown as triggers for eliciting production in these scenarios[67–69]. However, few studies decipher the responsible signals involved during co-culture.

In contrast to the ill-defined signals for activating BGCs through OSMAC or co-culture, known chemical elicitors of antibiotic production can be added directly to cultures (Fig. 2a). One such class of dedicated signalling molecules are γ-butyrolactones (GBLs), a widespread family of autoregulators in *Streptomyces* that interact directly with cognate GBL receptors to activate morphological development and secondary metabolism[70]**.** However, while exogenous addition of GBLs sometimes works, the strain specific nature of GBLs and barriers in purifying or synthesizing sufficient stereo-specific compound impede their use[71,72]. Alternatively, when applied at concentrations below their minimum inhibitory concentration (MIC), several antibiotics can activate antibiotic production through altering gene expression profiles, targeting the ribosome, or liberating precursors[70]. As a prime example of the latter, the ARC2 series of synthetic compounds were discovered in a high-throughput screen for inducers of pigment production in *S.*

coelicolor[73]. Resembling the FabI enoyl reductase inhibitor triclosan, ARC2 compounds enhance polyketide production by diverting resources from fatty acid primary metabolism, and have been useful for subsequent discovery of rare secondary metabolites from cryptic BGCs[74].

General activation of secondary metabolism can also be achieved by directly engineering regulatory machinery (Fig. 2a). One broadly used technique, developed by Ochi and coworkers, is to isolate spontaneous *rpsL* (S12 ribosomal protein) or *rpoB* (β-subunit of RNA polymerase) mutants resistant to streptomycin or rifampicin, respectively. Certain *rpsL* mutations stabilize the ribosome to enhance protein synthesis during late growth phase, while *rpoB* mutations in the rifampicin binding domain are thought to mimic the conformation of ppGpp bound RNA polymerase, which often triggers transcription of BGCs during secondary metabolism[75].

Finally, BGCs can be activated by targeting global regulators through deleting repressors (e.g. ScbR), overexpressing activators (e.g. AdpA, AfsR, Crp), or manipulating two component systems (e.g. PhoRP, AbsA1A2, AfsQ1Q2) and extracytoplasmic sigma factors[76–78]. It is these global regulators that integrate the environmental cues set by media conditions, co-culturing and chemical elicitors (Fig. 2a). Completing the regulatory cascade, these signalling pathways ultimately act on cluster-situated regulators that directly control a BGC of interest. An unprecedented proportion of actinomycete genomes are dedicated to regulatory genes, for example 2.03% represented by sigma factors, extracytoplasmic sigma factors, and two component systems in *S. coelicolor* compared to 0.73% in *E. coli*[62]. This complexity demonstrates

their importance to actinomycetes for adapting to different environmental conditions. Directly manipulating transcriptional regulators begins to incorporate our accrued knowledge of actinomycete regulatory circuits into hypothesis-driven approaches for BGC activation.

Pleiotropic approaches to BGC activation have the advantage of requiring little knowledge about the BGC being targeted and so are amenable to mid to high throughput applications. For example, many of these techniques were first developed using screens for pigment production (actinorhodin and prodigiosins) by *S. coelicolor* or *Streptomyces lividans* as an easily observable readout. More recent tools for assessing BGC activation involve engineering the cluster of interest with a fluorescent reporter (e.g. the HiTES platform[79]) or comparative metabolomics[80].

Targeted approaches to BGC activation

It is often the case that researchers have a specific cryptic BGC that they wish to activate. While a pleiotropic approach could be effective, these are inherently empirical and so the successful condition resulting in the BGC's activation is difficult to predict. In these cases, pathway specific approaches offer a targeted route to BGC activation.

Many targeted approaches to BGC activation are based on genetically manipulating the cluster. However, environmental strains containing a BGC of interest are often genetically intractable, slow growing, or produce many secondary metabolites that make purification challenging. In these cases, heterologous expression of the BGC offers a powerful route for cluster manipulation and characterization.

The first step to heterologous expression is mobilization of the cluster into a shuttle vector (Fig. 2b), which has traditionally been achieved by screening genomic libraries of cosmid/fosmids (up to 45kb inserts), bacterial artificial chromosomes (up to 490kb inserts) or P1-phage artificial chromosomes (up to 200kb insert)[81]. BGCs captured in these libraries are often split among multiple constructs, requiring additional stitching to complete the cluster. More recently, this bottleneck was overcome by targeted approaches that directly capture a BGC of interest in a single step (Fig. 2b). Targeted approaches are based on recombination *in vivo* using BGC specific homology arms, such as transformation associated recombination (TAR)[82], Cas9-Assisted Targeting of Chromosomes (CATCH)[83], and linear-plus-linear homologous recombination (LLHR)[84], or *in vitro* using Direct Pathway Cloning (DiPaC) facilitated by Gibson assembly[85,86].

Once in a vector, clusters are shuttled through *E. coli* for conjugation to a suitable heterologous host for expression. Optimal heterologous hosts are easy to manipulate, have compatible transcriptional/translational machinery (e.g. sigma factors and codon usage), and encode the essential components for common biosynthetic machinery (e.g. phosphopantetheinyl transferase and MbtH homologs for NRPS). While more closely related actinomycete strains to the BGC source are often more successful heterologous hosts, this is not a universal rule and often multiple strains are tested empirically[87,88]. Superhosts have also been engineered to have clean metabolic backgrounds, enhanced precursor supply and express pleiotropic positive regulators (e.g. *afsRS, crp*)[89–93]. For example, *S. coelicolor* (M1145/M1152/M1154) has been intensely developed as a heterologous host by deleting four major antibiotic gene clusters (Δ*act* Δ*red* Δ*cpk* Δ*cda*)

and introducing *rpoB* and *rpsL* mutations[91], while *S. lividans* (Sbt5/Sbt18)[94], *Streptomyces avermitilis* (SUKA5/SUKA17/SUKA22)[92,93] and *Streptomyces albus*[89,90] are also common superhosts.

Heterologous expression of BGCs opens the door to a myriad of engineering strategies for cluster activation and titer improvement that are not possible in genetically intractable native producers. Cluster-situated positive regulators can be overexpressed, repressors deleted, or pathway-specific precursor supply engineered, and there are numerous successful examples for each of these strategies (reviewed in [95–97]). The cluster can also be entirely refactored by promoter exchange. To this end, a number of libraries of *Streptomyces* strong constitutive promoters have been built[98,99], and various genetic strategies for promoter swapping have been developed (Fig. 2b), usually performed in *Saccharomyces cerevisiae* using yeast assembly[100], recombination with/without assistance by CRISPR-Cas9[101,102], or *E. coli* using Red/ET recombineering[103].

Unlike pleiotropic approaches to BGC activation that can be performed in relatively high throughput, targeted approaches require a good deal of time and concerted effort by the researcher. Therefore, it becomes critical to prioritize BGCs that are likely to yield novel compounds with interesting biological activity. Directing attention towards novel chemical matter, networks of BGCs based on measures of shared gene or domain content and sequence identity have been built in order to identify unexplored BGCs lacking known relatives[10,12,104,105]. Further correlating MS and genetic data can link the presence of a metabolite to a given BGC[12,104], which can be particularly helpful to avoid pursuing known compounds with unknown BGCs. Recently, the tool CONKAT-seq was

developed to perform such analysis on environmental DNA so that when combined with strategies for capture and heterologous expression, can be used to access low frequency BGCs in the soil microbiome[105]. Alternatively, building phylogenies of key biosynthetic sequence tags can point towards divergent members of known BGC families (see Chapter 3). Finally, by using putative resistance genes as a marker for activity, BGCs likely to encode antibiotics can be prioritized and can inform hypotheses about mechanism of action (see Chapter 4). As databases of actinomycete whole genome sequences continue to expand (>2700 actinobacterial genomes in GenBank as of 09/2020), bioinformatic tools such as these will become increasingly important to sort through data for meaningful signal.

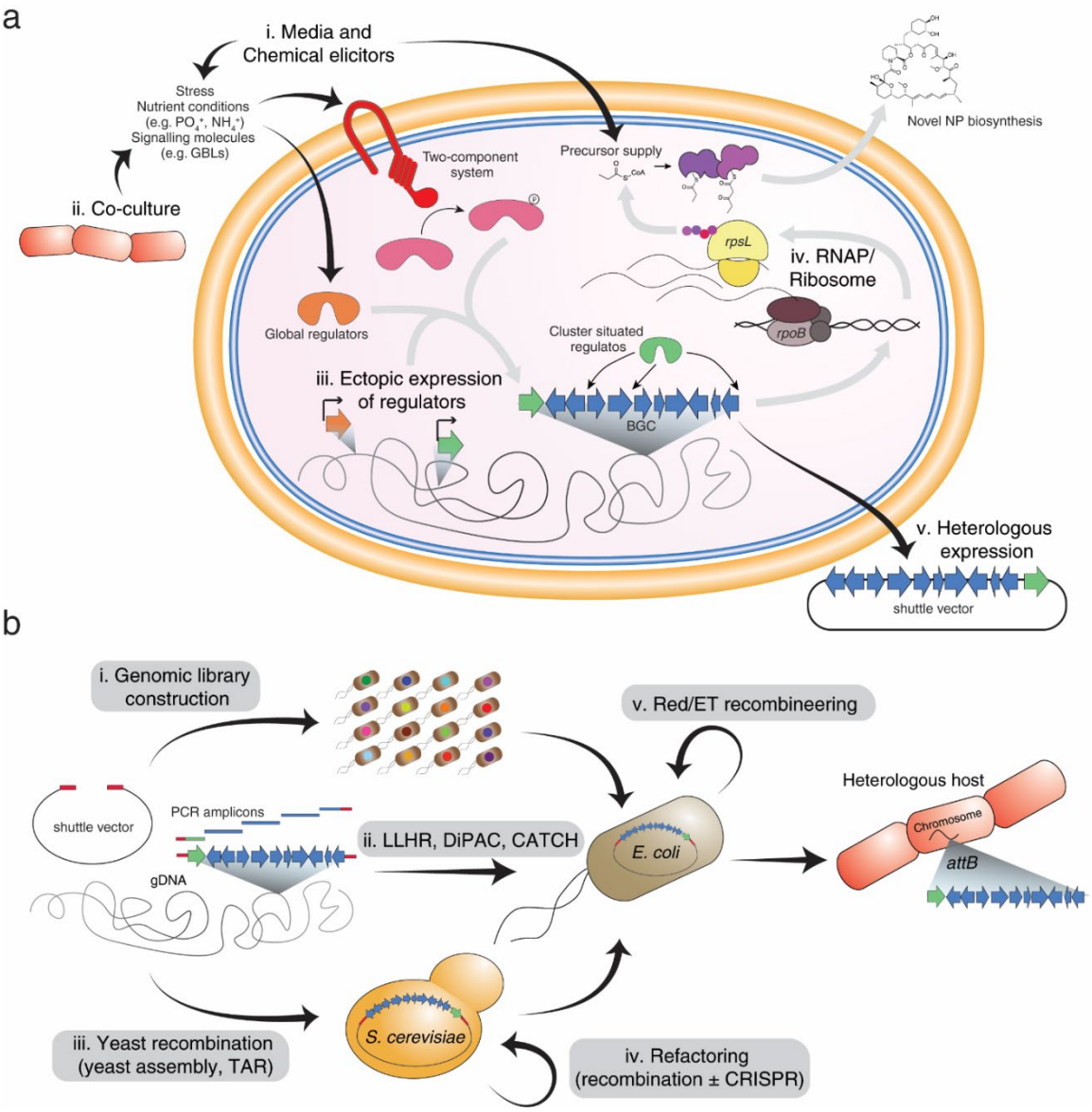


Figure 2. Strategies for cryptic BGC activation
(A) A combination of pleiotropic and targeted approaches can be used to transcriptionally activate and improve production of novel NPs. i-ii) Varying culturing conditions, applying chemical elicitors and using co-culture can produce signals that are recognized by global transcriptional regulators, two-component systems and extracytoplasmic sigma factors (not shown) to ultimately control transcription of BGCs, usually via cluster situated regulators. Media conditions and chemical elicitors can also affect precursor supply for NP biosynthesis. iii) Overexpressing global or cluster situated activators can directly manipulate transcriptional networks. Alternatively, transcriptional repressors can be deleted (not shown). iv) RNA polymerase (*rpoB*) and ribosome (*rpsL*) engineering can increase transcription and translation during secondary metabolism. v) Heterologous

expression allows all of these approaches (i-iv) to be applied in an easily manipulatable surrogate host. (B) Strategies for capturing a BGC into a shuttle vector usually starts with gDNA containing the BGC of interest or PCR amplification of the cluster in parts. i) Untargeted methods rely on ligation to build genomic libraries (e.g. cosmid, fosmid, BAC, PAC) which are subsequently screened for the BGC of interest. ii) Direct capture into *E. coli* can be achieved using LLHR, featuring RecET enhanced recombination[84], or DiPAC, featuring PCR amplification followed by Gibson assembly[85]. CATCH allows for Gibson assembly directly from gDNA by introducing specific double strand breaks *in vitro* using CRISPR-Cas9[83]. iii) Methods that rely on yeast homologous recombination shuttle the construct through *S. cerevisiae* before transforming into *E. coli.* BGCs can be refactored directly by inserting promoters during yeast assembly, or (iv) after capture using traditional yeast recombination[101]. Additional application of CRISPR-Cas9 technology allows for multiple promoter replacements to take place at once[102]. v) Refactoring can also be performed in *E. coli* using Red/ET recombineering. Final constructs are shuttled to a heterologous host of choice and typically integrated into the chromosome at *attB* sites.

PURPOSE AND GOALS OF THIS THESIS

The work described in this thesis focuses on expanding the toolbox of techniques that can be used to mine actinomycetes of their full biosynthetic potential, and subsequently characterize novel NPs discovered in the process. In chapter 2, I address the dereplication problem by developing a CRISPR-Cas9 based platform for inactivating common antibiotic BGCs that pollute actinomycete strain collections. By removing these nuisance compounds, I show the enrichment for discovery of rare antibiotics that are otherwise masked in wildtype strains. In chapter 3, I describe an improved BGC prioritization technique that integrates phylogeny with the presence of antibiotic resistance genes to uncover a novel functional family of glycopeptide antibiotics. As my major contribution to this study, the mechanism of action of these antibiotics was determined to be via blocking the remodelling of peptidoglycan by a family of essential enzymes, autolysins. Chapter 4 takes an alternative approach to BGC prioritization by focusing on clusters associated with the bacterial protease ClpP and demonstrates the utility of heterologous expression and promoter refactoring for cryptic BGC activation. The common thread that ties these chapters together, BGC-guided discovery of antibiotics from actinomycetes, is contrasted by the three orthogonal approaches taken, be it working with native producers versus engineered surrogate strains and focusing on known versus unknown BGCs as starting points for discovery. The conclusion to this thesis, chapter 5, will highlight common themes throughout and illuminate future prospects for antibiotic discovery in the modern scientific era.

Chapter II – Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics

PREFACE

The work presented in this chapter was previously published in:

Culp EJ, Yim G, Waglechner N, Wang W, Pawlowski AC, Wright GD. Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nat Biotechnol*. 2019;37(10):1149-1154. doi:10.1038/s41587-019-0241-9

E.J.C. and G.Y. contributed equally to this work. E.J.C., G.Y. and G.D.W. conceived the study and designed the experiments. W.W. performed bioactivity guided purification. N.W. assembled whole genome sequences and performed phylogenetic and BGC content analysis. A.C.P. designed the computer script for sgRNA identification. E.J.C. and G.Y. performed all other experiments. E.J.C., G.Y. and G.D.W wrote the manuscript.

ABSTRACT

Actinobacteria, which are one of the largest bacterial phyla and comprise between 13-30% of the soil microbiota, are the main source of antibiotic classes in clinical use[1]. During screens for antimicrobials, as many as 50% of actinomycete strains are discarded because they produce a known antibiotic (Supplementary Fig. 1)[2]. Despite each strain likely having the capacity to produce many compounds, strains are abandoned because the already-characterized antibiotic could interfere with screening for, or purification of, newly discovered compounds[3]. We applied CRISPR/Cas9-genome engineering to knock out genes encoding two of the most frequently rediscovered antibiotics, streptothricin or streptomycin, in eleven actinomycete strains. We report that this simple approach led to production of different antibiotics that were otherwise masked. We were able to rapidly discover rare and previously unknown variants of antibiotics including thiolactomycin, amicetin, phenanthroviridin and 5-chloro-3-formylindole. This strategy could be applied to existing strain collections to realize their biosynthetic potential.

MAIN

Antibiotics have been used to treat bacterial infections for the past 70 years but the emergence of multidrug-resistant pathogenic bacteria for which there are few, and in some cases no treatment options, is an urgent and pressing global problem[4]. No new classes of antibiotics have come through the clinical pipeline in the past 30 years[4] and refilling this pipeline is crucial if we are to address the challenge of multidrug-resistant pathogens.

Actinomycetes are a rich source of antibiotics with each genome containing 20-40 distinct biosynthetic gene clusters (BGCs) encoding specialized metabolites, only a minority of which have been chemically explored. Despite their potential, the canonical Waksman platform for antibiotic discovery[5], where actinobacterial extracts are screened against susceptible organisms for antimicrobial activity, has largely failed to yield novel drug scaffolds[3]. Chief among the platform's drawbacks is the rediscovery of known compounds and the need to identify these in complex mixtures, a process known as dereplication. Baltz has estimated that one antibiotic (streptothricin) is found in one in ten isolates and a handful of others at frequencies ~10^{-2} to 10^{-3} (streptomycin, tetracycline, and actinomycin D)[6]. Using the Waksman platform, libraries of tens of millions of isolates must be screened to find new antibiotics. To ease the consequent discovery burden, versatile and efficient platforms have been developed that can dereplicate common antibiotics (e.g. ref [2,7]). However, dereplication alone does not address the fact that strain collections are polluted with isolates producing common antibiotics. Triaging these strains based on the production of a known molecule may result in the collateral loss of new or rare compounds that are also encoded by the same strains but whose expression is masked by these common antibiotics.

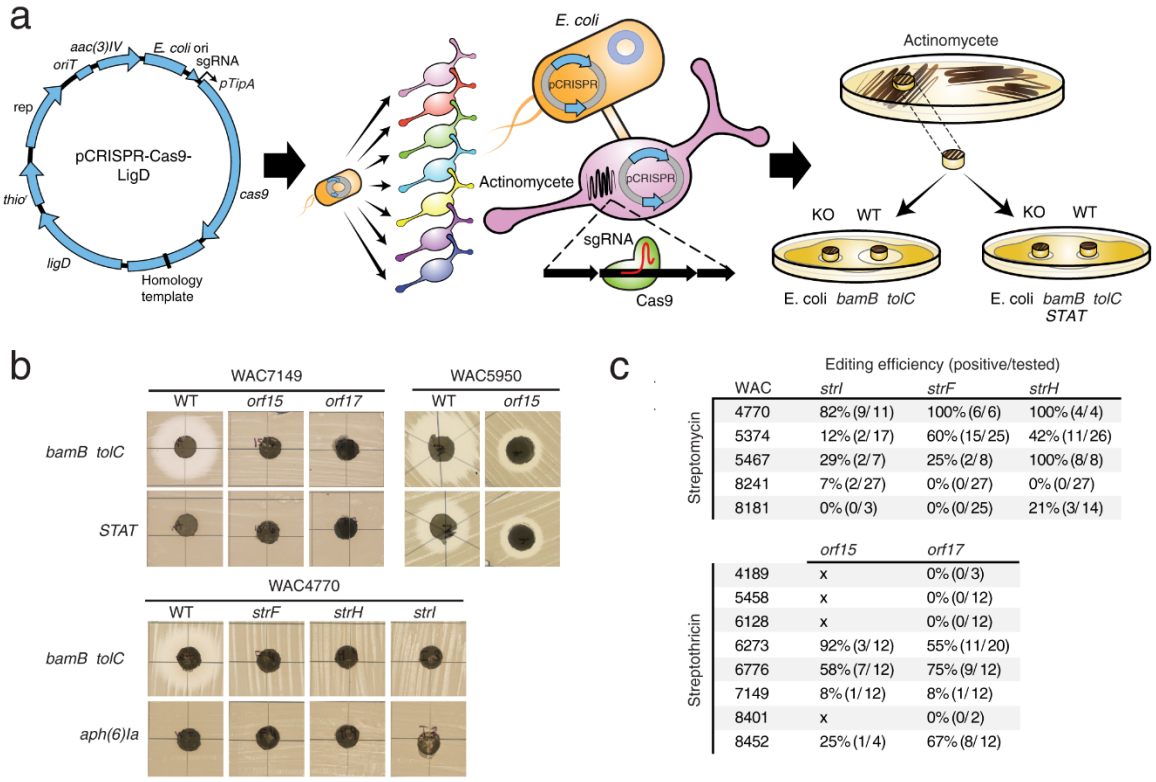
In order to exploit actinomycetes to their full biosynthetic potential, a toolbox of methods has been developed. These include activation of a specific BGC of interest by promoter refactoring or manipulation of pathway specific transcriptional regulators[8]. These targeted approaches require genetic manipulation and specialized constructs for each BGC of interest. Alternatively, methods for untargeted BGC activation include

introduction of pleiotropic activators such as AfsQ1[9], chemical elicitors[10] or transcription factor decoys[11]. However, these methods still require detection of any compound of interest alongside the production of large quantities of other active metabolites.

We hypothesized that it might be possible to re-investigate discarded strains producing common antibiotics by disrupting the production of those common antibiotics, thereby providing a straightforward, untargeted strategy to re-screen the 'tailings' of historic strain collections (Fig. 1a). We reasoned that disruption of conserved biosynthetic genes of the main antibiotic produced by a strain might facilitate detection of metabolites whose activity was otherwise overlooked. It is feasible that disrupting this production could also alter BGC regulatory circuits or liberate precursors, thereby enabling the increased production of metabolites produced in low levels (or not at all) in wildtype strains, as has been observed previously[12]. We report a generalizable CRISPR-Cas9 based system capable of inactivating production of antibiotics of interest by disrupting key common biosynthetic genes.

First we devised a pipeline to identify highly conserved sgRNA target sequences in a given BGC, enabling application of the same CRISPR-based targeted construct in multiple strains without prior knowledge of specific BGC sequences. Targeting the most commonly produced antibiotic, streptothricin (Supplementary Fig. 1), we mined 29 streptothricin BGCs from GenBank and our in-house genome sequences (Supplementary Table 1) and identified conserved target sites using a custom python script (Supplementary Fig. 1); sgRNA sites found in all 29 clusters were further refined (see

Methods), leaving us with two targets in the streptothricin BGC, *orf15* and *orf17* (Supplementary Fig. 2, Supplementary Table 2)[13].



	WAC	strI	strF	strH
Streptomycin	4770	82% (9/11)	100% (6/6)	100% (4/4)
	5374	12% (2/17)	60% (15/25)	42% (11/26)
	5467	29% (2/7)	25% (2/8)	100% (8/8)
	8241	7% (2/27)	0% (0/27)	0% (0/27)
	8181	0% (0/3)	0% (0/25)	21% (3/14)

		orf15	orf17
Streptothricin	4189	x	0% (0/3)
	5458	x	0% (0/12)
	6128	x	0% (0/12)
	6273	92% (3/12)	55% (11/20)
	6776	58% (7/12)	75% (9/12)
	7149	8% (1/12)	8% (1/12)
	8401	x	0% (0/2)
	8452	25% (1/4)	67% (8/12)

Figure 2. Application of CRISPR-Cas9 to inactivate streptothricin and streptomycin production.
(A) Summary of CRISPR-Cas9 strategy for the inactivation of commonly found antibiotics. First, the pCRISPR-Cas9-LigD plasmid, containing both *ligD* and a homology repair template, is engineered to target a highly conserved protospacer sequence in a BGC of interest. Next, the construct is delivered via conjugation to a wide range of actinomycete strains allowing for efficient genome editing. Successful gene inactivations resulting in loss of antibiotic production are identified by performing agar plug assays against susceptible and resistant *E. coli* Δ*bamB* Δ*tolC*. (B) Comparative bioassays are shown for representative streptothricin producers, WAC7149 and WAC5950, and streptomycin producer WAC4770. Two independent fermentations and bioassays gave similar results. (C) Summary of editing efficiencies based on bioassay results using the pCRISPR-Cas9 system. "X" indicates that conjugations were unsuccessful.

At the time of this study, two CRISPR-Cas9 systems for actinomycetes had been published: pCRISPomyces-2[14] and pCRISPR-Cas9[15] (Fig. 1a, Supplementary Fig. 1). The systems are similar but differ in promoters and DNA repair pathways employed. While both carry a repair template for homologous recombination (HR), we modified the pCRISPR-Cas9 plasmid to promote non-homologous end joining (NHEJ) as well. We tested both systems on eight streptothricin producers identified by our Antibiotic Resistance Platform[2]. Gene inactivation was confirmed by bioassay (Fig. 1b and Supplementary Fig. 3), LC-MS (Supplementary Fig. 4) and Sanger sequencing across the predicted lesion site (Supplementary Fig. 3). We disrupted streptothricin in five of the eight strains tested at efficiencies ranging from 8% to 75% (Supplementary Table 3, Fig. 1c). In those strains that failed, it was largely due to an inability to obtain exconjugants. We also targeted streptomycin, another highly abundant antibiotic, for inactivation, targeting three biosynthetic genes: *strF, strH* and *strI* (Supplementary Fig. 2 and Supplementary Table 2)[2,6]. Using the pCRISPR-Cas9 system, inactivation of streptomycin biosynthesis was verified in all five strains (Fig. 1, Supplementary Fig. 3, Supplementary Fig. 4 and Supplementary Table 3). From start to finish, 25-30 days were required to generate, cure and verify engineered strains, with only six days involving hands-on work (Supplementary Fig. 1).

As part of our verification pipeline, Sanger sequencing was performed on at least one exconjugant for every gene/strain inactivation generated (Table 1, Supplementary Fig 3). In roughly one third of the strains verified (9/25 strains generated), we found the introduction of a stop codon as expected from HR with the repair template. Perhaps due

to lower nucleotide identity in homology arms compared to other strains, one strain had a mutation characteristic of NHEJ (Table 1). We verified the absence of off-target effects in other genomes predicted to be repaired by HR through whole genome sequencing (WAC5374 *ΔstrF*). In the remaining strains (15/25), we were unable to PCR amplify the edited region (Supplementary Fig. 3). Whole genome sequencing of three of these strains revealed large deletions including the entire streptothricin or streptomycin BGC (Table 1). In WAC5374 Δ*strH,* despite repair through HR at the CRISPR-targeted locus, off-target activity led to a ~600 kb deletion elsewhere in the genome.

We observed many phenotypic differences between wildtype/engineered strain pairs including growth rate, sporulation, and pigment production (Fig. 2a and Supplementary Fig. 5) suggesting rewiring of transcriptional and/or precursor supply pathways regulating secondary metabolism in actinomycetes[16,17]. Metabolic profiles were further investigated in five wildtype strains and their respective engineered strains by principle component analysis (PCA) of high resolution LC-MS data to assess gross metabolomic differences (Fig. 2b and Supplementary Fig. 5). We observed effective clustering of replicates of a single strain and clear divisions between the wildtype and each engineered strain, reflecting altered metabolic profiles. Not surprisingly, greater shifts in metabolic profile were associated with large genomic deletions. Unexpectedly, targeting the same gene in the same strain also resulted in different metabolite profiles as was observed for WAC6273 Δ*orf15* strains generated by the pCRISPomyces or pCRISPR-Cas9 system (Fig. 2b). We detected large deletions in genomes of both of these engineered strains which may have led to this distinction (Table 1).

Table 1. CRISPR-engineered mutants generated and their uncovered metabolites.
NS – not significant

	WAC Strain	Gene targeted	CRISPR System	Mutation repair	Uncovered metabolites (Fold production increase vs WT)
Streptomycin	4770	*strI*	pCRISPR	Undefined deletion	
		strF	pCRISPR	HR	
		strH	pCRISPR	HR	
	5467	*strF*	pCRISPR	HR	
		strH	pCRISPR	HR	
	8181	*strH*	pCRISPR	Undefined deletion	
	5374	*strI*	pCRISPR	Undefined deletion	Thiolactomycin (NS)
		strF	pCRISPR	HR	Thiolactomycin (NS)
		strH	pCRISPR	HR and deletion (~600 kb)	5-chloro-3-formylindole (NS)
	8241	*strI*	pCRISPR	Deletion (~500 kb)	Phenanthroviridin aglycone (7x)
Streptothricin	7149	*orf15*	pCRISPR	Undefined deletion	
		orf17	pCRISPR	Undefined deletion	
		orf17	pCRISPomyces	Undefined deletion	
	8452	*orf15*	pCRISPR	Undefined deletion	
		orf17	pCRISPR	NHEJ - 1bp insertion	
	6273	*orf15*	pCRISPR	Deletion (~200kb)	Amicetin (3-20x), Ferrioxamines (9-68x)
		orf15	pCRISPomyces	Undefined deletion	Amicetin (4-22x), Ferrioxamines (11-62x)
		orf17	pCRISPR	Undefined deletion	Amicetin (4-21x), Ferrioxamines (7-50x)
	6776	*orf15*	pCRISPR	HR	
		orf15	pCRISPomyces	Undefined deletion	
		orf17	pCRISPR	Undefined deletion	
		orf17	pCRISPomyces	HR	
	6128	*orf17*	pCRISPomyces	Undefined deletion	
	5950	*orf15*	pCRISPomyces	Deletion (~1.5Mb)	
		orf17	pCRISPomyces	HR	

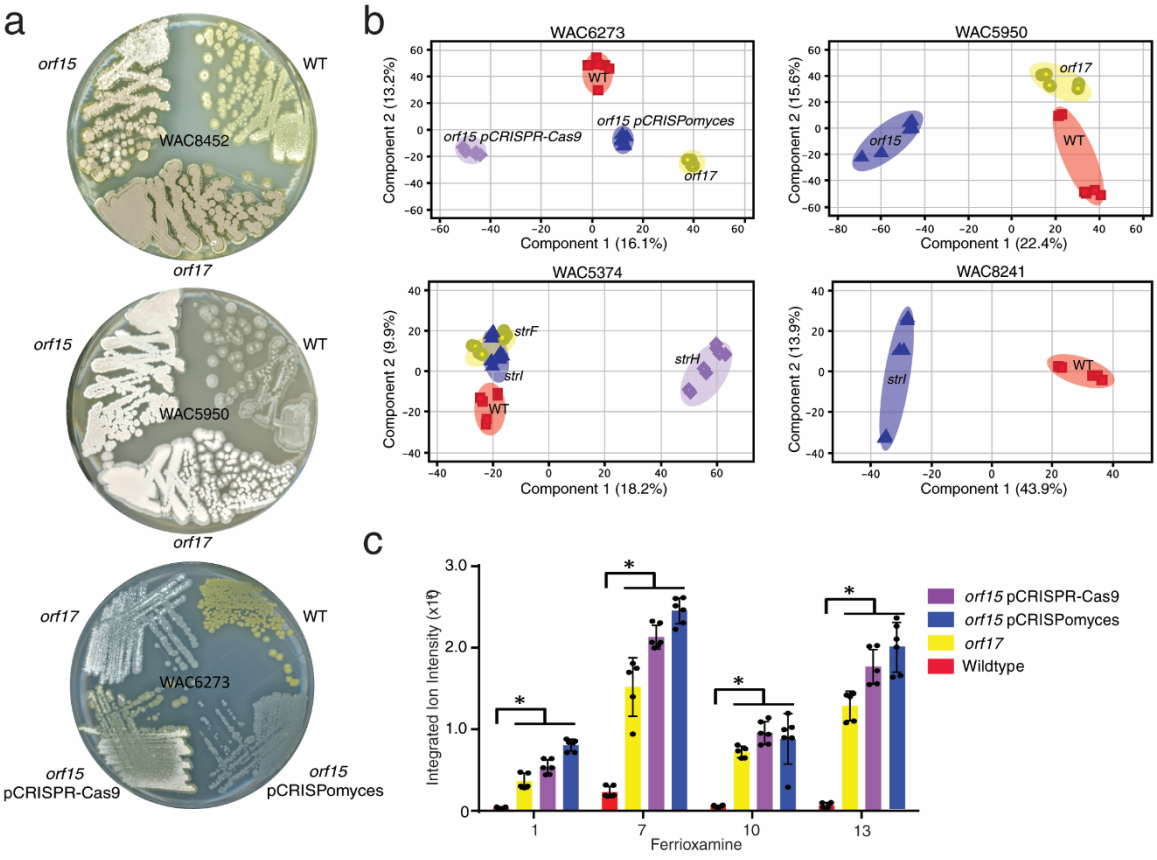


Figure 2. Inactivating antibiotic biosynthesis shifts the metabolic profile of producer bacteria.

Metabolic flux of wildtype and biosynthetically inactivated streptomycin or streptothricin strains were assessed globally by phenotypic analysis (A), high resolution LC-MS metabolomics (B), and on the production of select metabolites (C). Data are shown for representative streptothricin producers (WAC8452, WAC5950, WAC6273), and streptomycin producers (WAC5374, WAC8241). (A) Phenotypic comparisons were made after 7 days growth on Bennett's media. Growth was repeated on three independent occasions showing similar phenotypes. (B) LC-MS data from *n*-butanol extracts were used to construct PCA plots, revealing effective clustering of different engineered strains from their wildtype parent ($n = 6$, three independent fermentations analyzed in technical duplicate). (C) High resolution LC-MS analysis of WAC6273 extracts identifies a family of upregulated metabolites in engineered strains as ferrioxamines, including ferrioxamine 1, 7, 10 and 13 (Supplementary Fig. 6). Mean with error bars showing standard deviation is plotted ($n = 6$, three independent fermentations analyzed in technical duplicate). Significance is tested to $p<0.0001$ by one-way ANOVA with Tukey's post hoc analysis.

Consistent with PCA groupings, we identified metabolites unique or upregulated in engineered strains (Supplementary Fig. 5). To characterize these upregulated molecules, we used the Global Natural Product Social Molecular Networking (GNPS) platform[18]. In all three WAC6273 inactivated strains, we identified ferrioxamines, a suite of siderophores not generally observed in *Streptomyces* under typical culture conditions[19], as a family of highly upregulated metabolites (Fig. 2c, Supplementary Fig. 6)[19]. Altered metabolism in engineered strains therefore enables compounds not detectable in wildtype backgrounds to be readily identified.

To determine the potential chemical diversity available in streptothricin producers, we performed a phylogenetic analysis to examine the distribution of BGCs present in known streptothricin producers. Whole genome sequences of 42 streptothricin producers, pulled from in-house strains and GenBank (Supplementary Table 4), were evaluated for their biosynthetic potential (see Methods) and shown to have the capacity to produce a wide variety of specialized metabolites that rarely overlap in identity (Supplementary Fig. 7). When compared by multi-locus sequence analysis to 145 *Streptomyces* strains selected to represent diversity across the genus[20], streptothricin producers were spread across the phylogenetic tree (Supplementary Fig. Supplementary Fig. 8) These findings further encouraged us to explore the bioactive compounds from our inactivated strains.

From our collection of CRISPR-inactivated streptothricin and streptomycin producers, we selected a set of eight strains that retained antimicrobial activity against our tester strain *Escherichia coli* BW25113 Δ*bamB ΔtolC* (data not shown). We performed fermentations under a variety of conditions and observed antimicrobial activity in all

strains (Supplementary Fig. 9), suggesting that low abundance antibiotics were masked by streptothricin and streptomycin production in wildtype strains since these wildtypes had shown no activity against streptothricin and streptomycin resistant tester strains (Fig. 1, Supplementary Fig. 3). We chose three strains with substantial activity against our tester strain for bioactivity guided purification: one streptothricin inactivated producer, WAC6273, and two streptomycin inactivated producers, WAC5374 and WAC8241.

Activity guided purification from WAC6273 *Δorf17* resulted in the isolation of the rare antibiotic amicetin and associated derivatives, a family of compounds that inhibit translation elongation (Fig. 3a, Supplementary Table 5)[21]. Genome analysis identified the amicetin BGC in WAC6273 (Supplementary Fig. 10). The GNPS workflow revealed many amicetin derivatives in fermentation broths, several of which are novel derivatives with modifications that could be predicted by LC-MS/MS (Supplementary Fig. 11). LC-MS quantification showed that our CRISPR engineered strains produced higher levels of amicetin family compounds compared to the wildtype producer WAC6273 (Fig. 3b, Table 1). RT-PCR on the amicetin BGC indicated similar transcription levels in wildtype and engineered strains, suggesting that redirection of metabolic flux may be the primary driver of improved yield (Fig. 3b).

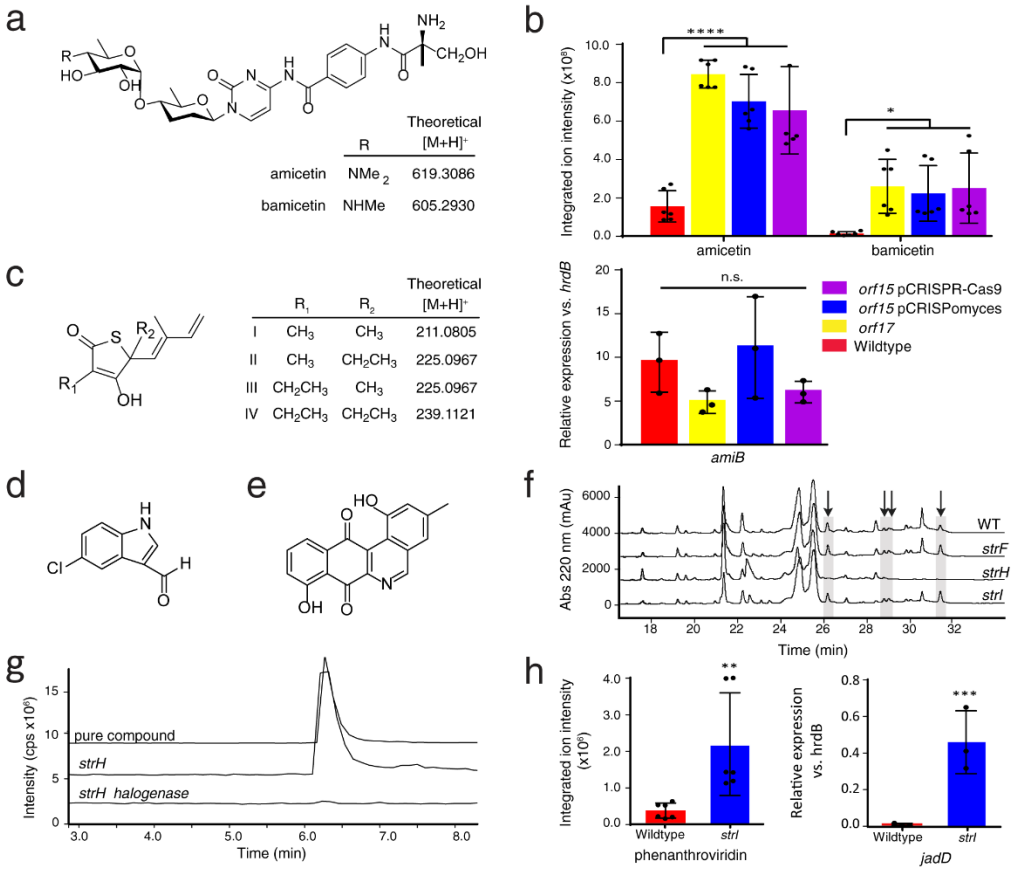


Figure 3. New antibiotic compounds discovered from CRISPR-inactivated strains. (A) Chemical structures of amicetin and bamicetin and (B) relative antibiotic production or expression of a key gene in their BGC, *amiB*, in WAC6273 wildtype and engineered strains. WAC6273 Δ*orf15* targeted mutants generated by the pCRISPomyces and pCRISPR-Cas9 systems were both analyzed. (C-E) Chemical structure of thiolactomycins I-IV (C), 5-chloro-3-formylindole (D), and phenanthroviridin aglycone (E). (F) HPLC analysis of WAC5374 crude extracts identifies thiolactomycins I-IV, indicated by black arrows, produced in all strains except WAC5374 Δ*strH*. (G) Relative amounts of 5-chloro-3-formylindole present in crude extracts as determined by extracted ion chromatograms (negative mode *m/z* 179.23-179.73). The WAC5374 Δ*strH* mutant was compared to derivatives with the tryptophan halogenase deleted. (H) Relative amounts of phenanthroviridin aglycone and expression of a key gene in its BGC, *jadD*, in WAC8241 wildtype and engineered strains. (B, G, H) For LC-MS production quantification, mean with error bars showing standard deviation is plotted ($n = 6$, three independent fermentations analyzed in technical duplicate). For gene expression, mean with error bars showing standard deviation is plotted ($n = 3$, three independent fermentations). Multiple comparison significance is tested to **p<0.0005 or *p<0.05 by one way ANOVA with Tukey's post hoc analysis (B), or pairwise to **p = 0.011 or ***p = 0.0103 by unpaired two-sided t test (F, G). n.s. not significant (p = 0.18). All experiments (B, F-H) were repeated on two independent occasions with similar results.

Activity guided purification from two different WAC5374 engineered strains, Δ*strI* and Δ*strH*, identified two different antibacterial compounds, thiolactomycin and 5-chloro-3-formylindole, respectively (Fig. 3c,d, Table 1). Thiolactomycin is a rare fatty acid synthesis inhibitor[22], while the antibiotic activity of 5-chloro-3-formylindole has not been previously described (Supplementary Table 5). The isolation of two different compounds from two mutants of the same parent strain is consistent with the distinct metabolic PCA profiles of Δ*strH* when compared to wildtype and other mutant strains (Fig. 2b). Thiolactomycin was found in crude fermentation extracts from wildtype, Δ*strF* and Δ*strI* mutants but not WAC5374 Δ*strH* (Fig. 3f), consistent with deletion of the thiolactomycin BGC in WAC5374 Δ*strH*. Genome analysis resulted in the identification of the thiolactomycin BGC (Supplementary Fig. 12)[23]. The predicted BGC for 5-chloro-3-formylindole was identified by searching the WAC5374 genome for a tryptophan halogenase and confirmed by deletion of this halogenase (Fig 3g, Supplementary Fig.13).

Activity guided purification from WAC8241 Δ*strI* resulted in the purification of phenanthroviridin aglycone, a low abundance derivative of the antibiotic jadomycin (Fig. 3e)[24] and identification of its BGC (Supplementary Fig. 14). By PCA, the metabolic profile of the Δ*strI* mutant was also distinct from the wildtype (Fig. 2b), and correspondingly the mutant produced higher levels of phenanthroviridin than wildtype (Fig. 3h). Levels of transcription of the identified BGC were similarly increased, indicating that transcriptional rewiring is at least partially responsible for improved yields in the engineered strain (Fig. 3h). Transcriptional rewiring in engineered strains may be the result of accumulation of biosynthetic intermediates that can interact with

transcription factors in feed-forward mechanisms to activate expression of later-stage biosynthetic enzymes[12] or cross-talk with other biosynthetic pathways[25]. Together with the liberation of precursors, these mechanisms result in shifts in metabolic profile to allow for the identification of compounds in engineered strains not detectable in the wildtype.

In order to assess the ability of our platform to discover rare, novel antibacterials from actinomycetes, we used BLAST to identify the number of BGCs of our discovered compounds in GenBank. Three unique biosynthetic genes were identified for each BGC and genomes that contained all three were taken to contain the cluster. At the time of this study, only five and four amicetin and thiolactomycin producers, respectively, could be identified in GenBank, much smaller than the 42 streptothricin producers found. While this proxy for compound rarity is heavily skewed by sampling bias, it does indicate that uncommon antibiotics can be found in streptothricin and streptomycin inactivated strains.

Our platform is widely applicable to most strains, and despite a few failures to induce mutations, we were successful in choosing conserved protospacers and inactivating streptothricin or streptomycin in 11/14 of our selected strains. While a drawback of all genetic platforms for BGC activation is the requirement to manipulate potentially intractable environmental isolates, accessing even only the easily engineered strains amounts to a substantial proportion of collections given the number of streptothricin/streptomycin producers they contain. Furthermore, the platform could be expanded to access difficult strains using technologies to manipulate non-laboratory or undomesticated strains[26,27].

When strains producing common compounds are triaged after initial screens, all other bioactive compounds potentially produced by these strains are simultaneously discarded. We report that CRISPR-Cas9 targeted inactivation of commonly found BGCs allows mining of a greater proportion of strain collections for new antibiotics. The concept is not limited to antibiotic biosynthesis and could be applied to target other specialized metabolites and bioactivities such as antibiotic adjuvants or anti-cancers. Our simple strategy to inactivate ubiquitous BGCs could allow researchers to tap into rare BGCs that are difficult to access by conventional screening of strain collections. Continuing to develop alternative approaches such as these will propel antibiotic discovery forward in the genomics age.

ACKNOWLEDGEMENTS

Computational resources for genome assembly and analysis were provided by Andrew G. McArthur at McMaster University. This research was funded by a Canadian Institutes of Health Research grant (MT-14981), the Ontario Research Fund, and by a Canada Research Chair (to G.D.W.). E.C. was supported by a CIHR Vanier Canada Graduate Scholarship. G.Y. was supported by a M.G. DeGroote Fellowship Award and a CIHR postdoctoral fellowship. N.W. was supported by a CIHR Canada Graduate Scholarship Doctoral Award. We thank Christy Groves for graphical edits to figures.

AUTHOR CONTRIBUTIONS

E.C., G.Y. and G.D.W. conceived the study and designed the experiments. W.W. performed bioactivity guided purification. N.W. assembled whole genome sequences and performed phylogenetic and BGC content analysis. A.C.P. designed the computer script for sgRNA identification. E.C. and G.Y. performed all other experiments. E.C., G.Y. and G.D.W wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Whole genome sequences of actinomycete strains used in this study, including 19 wildtype streptomycin or streptothricin producers and six CRISPR engineered derivatives (WAC5950 *Δorf15* pCRISPomyces, WAC6273 *Δorf15* pCRISPR-Cas9, WAC8241 *ΔstrI,* WAC5374 *ΔstrF,* WAC5374 *ΔstrH* and WAC5374 *ΔstrI*), are available in Genbank with the Bioproject accession number PRJNA504665, Supplementary Table 1.

CODE AVAILABILITY STATEMENT

A custom python script was used to identify conserved sgRNA target sites in a BGC of interest and is provided in Supplementary Note 1 together with instructions for use.

METHODS

Strains and Culture Conditions

E. coli strains and plasmids are listed in Supplementary Table 6. All *Streptomyces* strains are listed in Table S3. *E. coli* was grown at 37°C, 250 rpm, in LB Broth (Bioshop, Canada). *Streptomyces* strains were grown at 30°C, 250 rpm, in Tryptone-Soya-Broth (BD Biosciences) for starter cultures and genomic DNA preparation or SM-$MgCl_2$ (2% each D-mannitol, soya flour, agar, 10 mM $MgCl_2$) for sporulation and conjugation. Fermentations were performed in Bennett's media (1% potato starch, 0.2% casamino acids, 0.18% yeast extract, 0.02% KCl, 0.02% $MgSO_4 \cdot 7H_2O$, 0.024% $NaNO_3$, $4x10^{-4}$ % $FeSO_4 \cdot 7H_2O$) or SMP media (2.5% soluble starch, 0.146% glutamine, 0.1% K_2HPO_4, 0.1% NaCl, 0.05% $MgSO_4 \cdot 7H_2O$, $4x10^{-6}$ % $ZnCl_2$, $2x10^{-5}$ % $FeCl_3 \cdot 6H_2O$, $1x10^{-6}$ % each $CuCl_2 \cdot 2H_2O$, $MnCl_2 \cdot 4H_2O$, $Na_2B_4O_7 \cdot 10H_2O$, $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$), as indicated. Antibiotics were supplemented as necessary (100 µg/mL ampicillin, 50 µg/mL apramycin, 35 µg/mL chloramphenicol, 25 µg/mL nalidixic acid).

In silico detection of conserved protospacer motifs

A python script was used to identify conserved sgRNA sites across a set of 12 or 29 exemplar streptomycin or streptothricin BGCs, respectively, pulled from Genbank and in-house genome sequences (Supplementary Table 1). First, all protospacer motifs of the form N_{20}-NGG were identified in select gene targets, and only those present in all given BGCs were kept. Off-target sites were checked by alignment of protospacer motifs to whole genomes (streptomycin: WAC4770, streptothricin: WAC5950, *Streptomyces* sp.

TOR3209 accession no. NZ_AGNH00000000.1 and *Streptomyces* sp. PCS3-D2 accession no. JDUZ01000001.1) using Bowtie2[28], then prioritized based on proximity to a gene's 5' end (Supplementary Fig. 1). Protospacer motifs selected to target streptothricin and streptomycin BGCs are listed in Supplementary Table 2. Sequence alignments of *orf15* and *orf17* in Supplementary Figure 1 were generated using Geneious V8.1.

pCRISPR-Cas9-LigD and pCRISPomyces sgRNA and repair template cloning

Primers used for construction of all CRISPR constructs are listed in Supplementary Table 2. For simplicity, while Tong et al.[15] distinguishes between pCRISPR-Cas9 and pCRISPR-Cas9-LigD, in this work the pCRISPR-Cas9 system refers to a plasmid containing both *ligD* (an enzyme required for NHEJ) and a homology template. Similarly, while Cobb et al.[14] distinguishes between two forms of the pCRISPomyces system (pCRISPomyces-1 and 2), only the pCRISPomyces-2 system was used in this work.

pCRISPR-Cas9 was previously designed to be able to carry either *ligD* or a homology repair template, but not both. We modified the pCRISPR-Cas9 plasmid to heterologously express *ligD* and carry the homology repair template, thereby allowing either DNA repair pathway to occur. Since we used the same repair template for all strains, we reasoned that if the HR template lacked sufficient identity in a particular strain's genome, LigD might allow for efficient repair via NHEJ. Therefore, we first modified the plasmid, adding a *StuI* site in front of the *ligD* expression cassette to allow a location for repair template cloning. *ligD* was PCR amplified from the original pCRISPR-

Cas9-LigD construct and cloned back into *StuI* digested pCRISPR-Cas9 by Gibson Assembly (NEB), this time designing the primers to regenerate the *StuI* site upstream of *ligD.*

Repair templates for HR were designed in the same way for both pCRISPR-Cas9 and pCRISPomyces systems. Approximately 1 kb on either side of the sgRNA target site was amplified from a representative BGC (*Streptomyces sp.* WAC5950 for streptothricin, WAC4770 for streptomycin) using primers designed to remove the protospacer motif while introducing an in-frame stop codon and in some cases, a *Hind*III site. Overlap-extension PCR or Gibson assembly was used to combine the two sides of the repair template and clone them into *Xba*I digested pCRISPomyces-2 or *Stu*I digested pCRISPR-Cas9.

Desired protospacer sequences were introduced into primers and cloned into pCRISPomyces using Golden Gate assembly[29]. Alternatively, they were cloned into pCRISPR-Cas9 between *Nco*I and *Sna*BI restriction sites using traditional ligation based methods. Streptothricin repair templates and sgRNAs were subcloned from pCRISPomyces plasmids to pCRISPomyces. All plasmids described were sequence verified and transformed into ET12567/pUZ8002[30].

Generation and verification of CRISPR-mediated BGC inactivation strains

All constructs were moved into *Streptomyces* spp. via conjugations from *E. coli* ET12567/pUZ8002 using standard protocols[31]. Exconjugants were selected using 50 µg/mL apramycin and 25 µg/mL nalidixic acid. For the pCRISPR-Cas9 system, twelve exconjugants were subsequently restreaked on Bennett's or Nutrient Broth (BD

Biosciences) plates with 1 μg/mL thiostrepton to induce *cas9* expression. For both systems, single colonies were finally restreaked onto Bennett's media, or inoculated into 3 mL Bennett's or D media and grown for 7 days. Agar plugs or 30 μL of spent media was tested by a Kirby-Bauer assay on cation adjusted Mueller-Hinton agar (BD Biosciences) against *E. coli* BW25113 Δ*bamB ΔtolC* with or without pGDP3:*aph(6)*Ia and pGDP1:*STAT* [2], for streptomycin and streptothricin detection, respectively. Exconjugants that produced no differential zone of inhibition around wildtype and resistant *E. coli* strains were predicted to be successfully engineered (Fig. 1b and Supplementary Fig. 3). Candidate engineered strains were cured of targeting plasmids, all of which carry a temperature sensitive *ori* derived from pSG5, by growth at 37-40°C. Curing was confirmed by PCR and apramycin susceptibility testing.

Candidate inactivated strains were verified by Sanger sequencing and LC-MS. PCR sequencing primers were designed in conserved regions roughly 500 bp on either side of the target site, allowing amplification from all strains using the same primers (Supplementary Table 2). Samples for LC-MS verification were prepared by growing strains in liquid media under production conditions, as described above, and vortexing spent media with an equal volume of chloroform then centrifuging to remove protein and cell material. Samples were analyzed in positive mode using an Agilent 110 Series HPLC system and Applied Biosystems QTRAP LC-ESI/MS in positive mode, with a Sunniest RP-Aqua 4.6x150 mm 5μm C28 column and LC conditions as follows: 0-2 min 2% B, 2-4 min 2-15% B, 4-5 min 15-20% B, 5-7 min 20-100% B, 7-9 min 100% B, 9-9.5 min

100-2% B, 9.5-11.5 min 2% B (A: water+0.1% formic acid, B: acetonitrile+0.1% formic acid, 0.8 mL/min).

Preparation of crude CRISPR-mediated BGC inactivated strain extracts

Verified streptothricin and streptomycin null strains were inoculated 1:100 from TSB seed cultures to 50 mL cultures in a 250 mL flask and grown for 7 days at 30°C, 250 rpm. Bennett's media was used for metabolic profiling, and up to seven different fermentation conditions were used for bioactivity testing. Whole cultures were extracted with 15 mL *n*-butanol, dried under vacuum, and resuspended in 100 µL DMSO.

Metabolic profiling and PCA analysis of CRISPR-mediated BGC inactivated strains

Extracts from three biological replicates were run in technical duplicate in positive mode using an Agilent 1290 Infinity II Series HPLC system and 6550 iFunnel QTOF LC-ESI/MS, with an Agilent XDB-C8 3.5 µm 2.1x100 mm column. LC conditions were as follows: 0-1 min 5% B, 1-7 min 5-97% B, 7-7.5 min 97% B, 7.5-8 min 97-5% B, 8-10min 5% B (A: water+0.1% formic acid, B: acetonitrile+0.1% formic acid, 0.4 mL/min). For compound abundance comparisons between strains, extracted ion chromatograms for the mass of interest were integrated. Standard curves were used to verify linearity over the dynamic range of interest. Full metabolomic comparisons were made using Agilent MassHunter Profinder software for feature extraction and chromatographic alignment followed by Mass Profiler Professional for PCA.

Molecular networking of CRISPR-mediated BGC inactivated strain extracts

Extracts from three biological replicates were analyzed using the same QTOF LC-MS system as for metabolic profiling. LC conditions were as follows: 0-1.5 min 5% B,

1.5-3 min 5-40% B, 3-4 min 40% B, 4-10min 40-100% B, 10-11min 100% B, 11-11.5 min 100-5% B, 11.5-12 min 5% B (A: water+0.1% formic acid, B: acetonitrile+0.1% formic acid, 0.4 mL/min). AutoMS/MS acquisition settings used were as follows: positive ion mass range 200-2000, MS scan rate 3 spectra/sec, MS/MS scan rate 2 spectra/sec, collision energy 20 eV. The top five most abundant precursor ions were selected at every scan for fragmentation, and after three cycles these precursor masses were excluded from fragmentation for 45 sec.

Molecular networks were generated using the online workflow at GNPS[18]. Data was filtered to remove MS/MS peaks within 17 Da of the precursor m/z, parent masses were clustered with a tolerance of 0.1 Da, and MS/MS fragments clustered with an ion tolerance 0.1 Da. Networks were generated where edges have a cosine score above 0.7 and more than 5 matched MS/MS peaks.

Genome sequencing of wild type and CRISPR inactivated strains

Strains for sequencing were grown in TSB at 30°C, 250 rpm to midlog phase. Cells were pelleted and lysed with standard lysozyme, proteinase K and SDS treatment, followed by phenol/chloroform cleanup and ethanol precipitation or column purification. Illumina MiSeq sequencing (300 bp, paired end reads) was performed by the Farncombe Genomics Facility.

Nanopore MinION sequencing was performed in-house with 1D, R9.4 chemistry. Libraries were made using standard ligation sequencing kit protocols (Nanoporetech, catalog number SQK-LSK109), omitting the FFPE repair enzyme mix and performing clean up with LFB buffer. Genomic libraries were run on the flow cell (FLO-MIN-106)

and base calling was performed on the raw nanopore signals using Albacore 2.3.1 with default parameters to generate reads in fast5 and fastq format. Passing reads were adapter trimmed using porechop version 0.2.3 (https://github.com/rrwick/Porechop), and assembled using minimap2 version 2.11[36] (using the -x ava-ont switch) and miniasm 0.3-r179[37] with default parameters. The assembly graph in GFA format was manually inspected using Bandage version 0.8.0[38] to make sure there were no unusual edges in the assembly. An improved consensus sequence was generated by mapping the trimmed reads to the initial assembly with minimap2 (using the -x map-ont switch) and polishing with Racon (version 1.3.1)[39] using default parameters. A second round of contig polishing was performed using Nanopolish 0.10.1[40] using the reads mapped to the improved racon assembly as well as the passing Albacore fast5 output with the -methylation-aware=dam,dcm option.

The size of large genomic deletions in knockout strains was determined by comparing the full genome assembly size of the wildtype and knockout strains. Where nanopore sequencing was used to build complete genomes of knockout strains, Illumina contigs from corresponding wildtype strains were aligned onto Nanopore assemblies using BLAST to determine the location and contents of deletions.

WAC5374 ΔstrH halogenase gene inactivation

The putative tryptophan halogenase or related NRPS in WAC5374 was disrupted using pCRISPR-Cas9, without LigD system, in a similar way to that described above. The pCRISPR-Cas9 construct was designed to introduce a 654 bp in frame deletion using primers listed in Table S2 and verified by PCR and Sanger sequencing. Disruption of the

chlorinated indole aldehyde production was verified by high resolution LC-MS analysis of *n*-butanolic extracts from two independent fermentations.

Assessing streptothricin producer BGC diversity

In addition to the genome sequences harbouring a streptothricin BGC from our in-house strain collection, we added 145 *Streptomyces* genomes described previously[20] and 17 other actinomycete genomes that were available from public databases. The entire set of streptothricin BGC containing genome sequences were subjected to analysis by antiSMASH v4 using the '—smcogs' '—knownclusterblast' and '—full-hmmer' in addition to default options[32]. Each BGC in each genome was extracted and examined individually by a custom Python script. Coding sequences (CDSs) in each cluster were extracted whole if they contained zero or one 'aSdomain' features. However, if they contained two or more 'aSdomain' features the coding sequence was divided at the borders of each 'aSdomain' and extracted separately, including the sequences, if any, at the beginning and end of each CDS and between labelled domains. The entire set of extracted CDS sequences were then subjected to clustering using USEARCH v8.1.181[33] using the 'cluster_fast' mode with a 60% identity cutoff. After clustering, each CDS and CDS fragment was labelled with the cluster number it was assigned, so each BGC can be described as a sequence of labelled CDS fragments. We compared the overall biosynthetic capacity pairwise between each genome by implementing the Jaccard Index component of the Lin index[34,35] using the unique elements (genes/domains) of the set of combined genome BGC CDS fragment labels.

Phylogenetic analysis

The hidden Markov models (HMMs) corresponding to every family listed under TIGRFAM genome property 0799 'bacterial core gene set, exactly 1 per genome' (https://genome-properties.jcvi.org/cgi-bin/Listing.cgi, accessed Sept. 12th 2018) were collected. HMMER3 was used to analyze every genome using each model's trusted cutoff[41]. The top hits for each model were retained and aligned as a group against the model HMM. If a genome lacked a hit for a model, gaps equal to the length of the missing sequence were added to the alignment for that genome. These aligned model families were subsequently concatenated into an overall alignment which was inspected manually. This alignment was used for phylogenetic analysis using fasttree2 using the WAG substitution model and otherwise default parameters[42].

RT PCR on WAC6273, WAC5374 and WAC8241

Strains for analysis were inoculated 1:100 from a TSB seed culture into 60 mL Bennett's (WAC6273) or SMP media (WAC8241) and pellets were taken at early stationary phase, 16 hr (WAC8241) or 20 hr (WAC6273). Cells were lysed by bead beating mycelium with 4 mm glass beads in 5 mL TRIzol reagent (Invitrogen), and RNA was extracted using the manufacturer's recommendations. RNA from the resulting aqueous phase was extracted a second time using acid phenol/chloroform, then combined with a half volume of anhydrous ethanol and finally purified using PureLink RNA Mini Kit (Invitrogen). Maxima H Minus First Strand cDNA synthesis kit with dsDNase (Thermo Scientific) was used for cDNA synthesis, and PowerUp SYBR Green master mix (Applied Biosystems) was used for RT-PCR quantification on a BioRad CFX96 real

time system. Primers targeting major biosynthetic operons in each BGC of interest were designed (Supplementary Table 2) and 90-100% efficiency was verified prior to quantification. Analysis was performed on three or four independent fermentations and quantified in technical triplicate. Technical triplicates for each biological replicate were averaged, then fold change expression for each replicate was calculated by normalizing to *hrdB* expression using the ΔCt method.

Thiolactomycin purification from WAC5374 ΔstrI

WAC5374 Δ*strI* was inoculated 1:100 from a TSB seed culture into two 3 L flasks each with 600 mL SMP media. After 7 days of growth at 30°C, spent media was adjusted to pH 4 and extracted with 1 L of ethyl acetate. The ethyl acetate layer was dried under vacuum, resuspended in methanol and loaded onto a 5 g C18 prepacked cartridge. Reverse phase flash chromatography was performed as follows: 0-5 min 5% B, 5-25min 5-75% B, 25-28 min 75% B, 28-30min 75-95% B (A: water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid, 30 mL/min). Active fractions containing pure compound were identified as thiolactomycin I-IV by analysis using QTOF LC-MS and LC-MS/MS under the same conditions as for metabolic profiling.

Phenanthroviridin aglycone purification from WAC8241 ΔstrI

Phenanthroviridin aglycone was isolated by activity guided purification from WAC8241 Δ*strI* fermented in SMP media. After seven days of fermentation, spent media was treated with Diaion HP-20 resin (Sigma-Aldrich, Canada). The HP-20 resin was eluted with 20, 40, 80 and 100% methanol (MeOH). The 80 and 100% MeOH fractions were acidified with 0.3% acetic acid and extracted twice with hexane to remove highly

hydrophobic compounds. The MeOH layer was put on a C18 vacuum liquid chromatography (VLC) column and eluted with 20, 40, 60, 80 and 100% MeOH. The 60% and 80% MeOH fractions were subjected to a Sephadex LH20 (Sigma-Aldrich, Canada) column using MeOH:dichloromethane (2:1) as a running solvent. Active fractions were run on a silica gel (Sigma-Aldrich, Canada) VLC column, eluting with hexane, hexane:ethyl acetate (1:1), ethyl acetate, ethyl acetate:MeOH (9:1), ethyl acetate:MeOH (8:2), ethyl acetate:MeOH (1:1), MeOH and MeOH:water (7:3). The active fractions (ethyl acetate and ethyl acetate:MeOH (9:1)) were further purified by HPLC on a Waters Atlantis T3 prep column using a linear gradient from 95% to 5% solvent A (solvent A water, solvent B acetonitrile, both acidified with 0.1% trifluoroacetic acid) to yield the active compound phenanthroviridin aglycone.

Purification of amicetin and derivatives from WAC6273 Δorf17

WAC6273 Δ*orf17* was fermented in Bennett's media for five days. Amicetin and its derivatives were isolated by activity guided purification. Spent media was treated with Diaion HP-20 resin. The HP-20 resin was eluted with 5, 20, 50 and 100% MeOH. The active fractions were subject to a C18 CombiFlash column using a linear gradient from 95% to 5% solvent A (solvent A water, solvent B acetonitrile, both acidified with 0.1% formic acid). Active fractions were purified by HPLC on a Waters Atlantis T3 prep column using a linear gradient from 95% to 5% solvent A (solvent A water, solvent B acetonitrile, both acidified with 0.1% trifluoroacetic acid). The active fractions contained amicetin and its derivatives.

Purification of 5-chloro-3-formylindole from WAC5374 ΔstrH

WAC5374 *ΔstrH* was fermented in SMP media for eight days. The culture was treated with Diaion HP-20 resin and the resin was eluted with 10, 20, 40, 60, 80 and 100% MeOH. The 60 and 80% MeOH fractions were run on a Sephadex LH20 column using 350 ml of MeOH, 200 ml dichloromethane:MeOH (1:1) and 100 mL MeOH as running solvents. The active fractions were put on a C18 vacuum column and eluted with 20, 40, 50, 60, 70, 80 and 100% MeOH. The 50 and 60% MeOH fractions were combined and subjected to a C18 combiFLASH column using a linear gradient from 95% to 5% solvent A (solvent A water, solvent B acetonitrile, both acidified with 0.1% formic acid). The active fractions were fractionated by HPLC on a Waters Atlantis T3 prep column using a linear gradient from 95% to 5% solvent A (solvent A water, solvent B acetonitrile, both acidified with 0.1% trifluoroacetic acid) to give 5-chloro-3-formylindole.

Statistical analysis

Statistical analysis of production as quantified by LC-MS and gene expression as quantified by RT-PCR was performed using GraphPad Prism V6. Pairwise comparisons were made by unpaired two-sided t test (n = 6 for LC-MS data and n = 3 for RT-PCR, as indicated) and p values are reported in figure legends. Multiple comparisons were made by one-way ANOVA with Tukey's post hoc analysis (n = 6). Parameters are as follows for LC-MS production quantification: WAC8241 (t=3.107, df = 10), WAC6273 ferrioxamine 1 (F = 116.7, df = 20), ferrioxamine 7 (F = 143, df = 20), ferrioxamine 10 (F = 34.34, df = 20), ferrioxamine 13 (F = 103.3, df = 20), WAC6273 amicetin (F = 15.57,

df = 20), WAC6273 bamicetin (F = 3.67, df = 20). Parameters are as follows for gene expression quantification: WAC8241 (t = 4.562, df = 4). All results are representative of two independent experiments.

Reporting summary

Further information on research design is available in the Life Sciences Reporting Summary linked to this article.

REFERENCES

1. Barka, E. A. *et al.* Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev.* **80,** 1–43 (2016).

2. Cox, G. *et al.* A common platform for antibiotic dereplication and adjuvant discovery. *Cell Chem. Biol.* **24,** 98–109 (2017).

3. Wright, G. D. Something old, something new: revisiting natural products in antibiotic drug discovery. *Can. J. Microbiol.* **60,** 147–54 (2014).

4. Wright, G. D. Solving the antibiotic crisis. *ACS Infect. Dis.* **1,** 80–84 (2015).

5. Lewis, K. Antibiotics: Recover the lost art of drug discovery. *Nature* **485,** 439–40 (2012).

6. Baltz, R. H. Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J. Ind. Microbiol. Biotechnol.* **33,** 507–513 (2006).

7. Allard, P.-M. *et al.* Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Anal. Chem.* **88,** 3317–3323 (2016).

8. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13,** 509–523 (2015).

9. Daniel-Ivad, M. *et al.* An engineered allele of *afsQ1* facilitates the discovery and investigation of cryptic natural products. *ACS Chem. Biol.* **12,** 628–634 (2017).

10. Mao, D., Okada, B. K., Wu, Y., Xu, F. & Seyedsayamdost, M. R. Recent advances in activating silent biosynthetic gene clusters in bacteria. *Curr. Opin. Microbiol.* **45,** 156–163 (2018).

11. 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 (2019).

12. Zhang, Y. *et al.* JadR*-mediated feed-forward regulation of cofactor supply in jadomycin biosynthesis. *Mol. Microbiol.* **90,** 884–897 (2013).

13. Maruyama, C. *et al.* A stand-alone adenylation domain forms amide bonds in streptothricin biosynthesis. *Nat. Chem. Biol.* **8,** 791–797 (2012).

14. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4,**

723–728 (2015).

15. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* **4,** 1020–1029 (2015).

16. Craney, A., Ozimok, C., Pimentel-Elardo, S. M., Capretta, A. & Nodwell, J. R. Chemical Perturbation of Secondary Metabolism Demonstrates Important Links to Primary Metabolism. *Chem. Biol.* **19,** 1020–1027 (2012).

17. Thykaer, J. *et al.* Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster. *Metab. Eng.* **12,** 455–461 (2010).

18. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34,** 828–837 (2016).

19. Sidebottom, A. M., Johnson, A. R., Karty, J. A., Trader, D. J. & Carlson, E. E. Integrated metabolomics approach facilitates discovery of an unpredicted natural product suite from *Streptomyces coelicolor* M145. *ACS Chem. Biol.* **8,** 2009–2016 (2013).

20. McDonald, B. R. & Currie, C. R. Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *MBio* **8,** e00644-17 (2017).

21. Zhang, G. *et al.* Characterization of the amicetin biosynthesis gene cluster from *Streptomyces vinaceusdrappus* NRRL 2363 implicates two alternative strategies for amide bond formation. *Appl. Environ. Microbiol.* **78,** 2393–401 (2012).

22. Slayden, R. A. *et al.* Antimycobacterial action of thiolactomycin: an inhibitor of fatty acid and mycolic acid synthesis. *Antimicrob. Agents Chemother.* **40,** 2813–2819 (1996).

23. Tang, X. *et al.* Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10,** 2841–2849 (2015).

24. Frendrich, G. *et al.* Phenanthridine derivatives, process for the preparation thereof, and compositions containing them. (1990).

25. de la Fuente, A., Lorenzana, L. M., Martin, J. F. & Liras, P. Mutants of *Streptomyces clavuligerus* with disruptions in different genes for clavulanic acid biosynthesis produce large amounts of holomycin: possible cross-regulation of two unrelated secondary metabolic pathways. *J. Bacteriol.* **184,** 6559–6565 (2002).

26. Li, L. *et al.* CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in *Streptomyces*. *Appl. Environ. Microbiol.* **84,** e00827-18 (2018).

27. Brophy, J. A. N. *et al.* Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nat. Microbiol.* **3,** 1043–1053 (2018).

28. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–9 (2012).

29. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: A one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4,** e5553 (2009).

30. Paget, M. S., Chamberlin, L., Atrih, A., Foster, S. J. & Buttner, M. J. Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181,** 204–11 (1999).

31. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (John Innes Foundation, 2000).

32. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45,** W36–W41 (2017).

33. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).

34. Lin, K., Zhu, L. & Zhang, D.-Y. An initial strategy for comparing proteins at the domain architecture level. *Bioinformatics* **22,** 2081–2086 (2006).

35. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158,** 412–421 (2014).

36. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34,** 3094–3100 (2018).

37. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32,** 2103–10 (2016).

38. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31,** 3350–2 (2015).

39. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27,** 737–746 (2017).

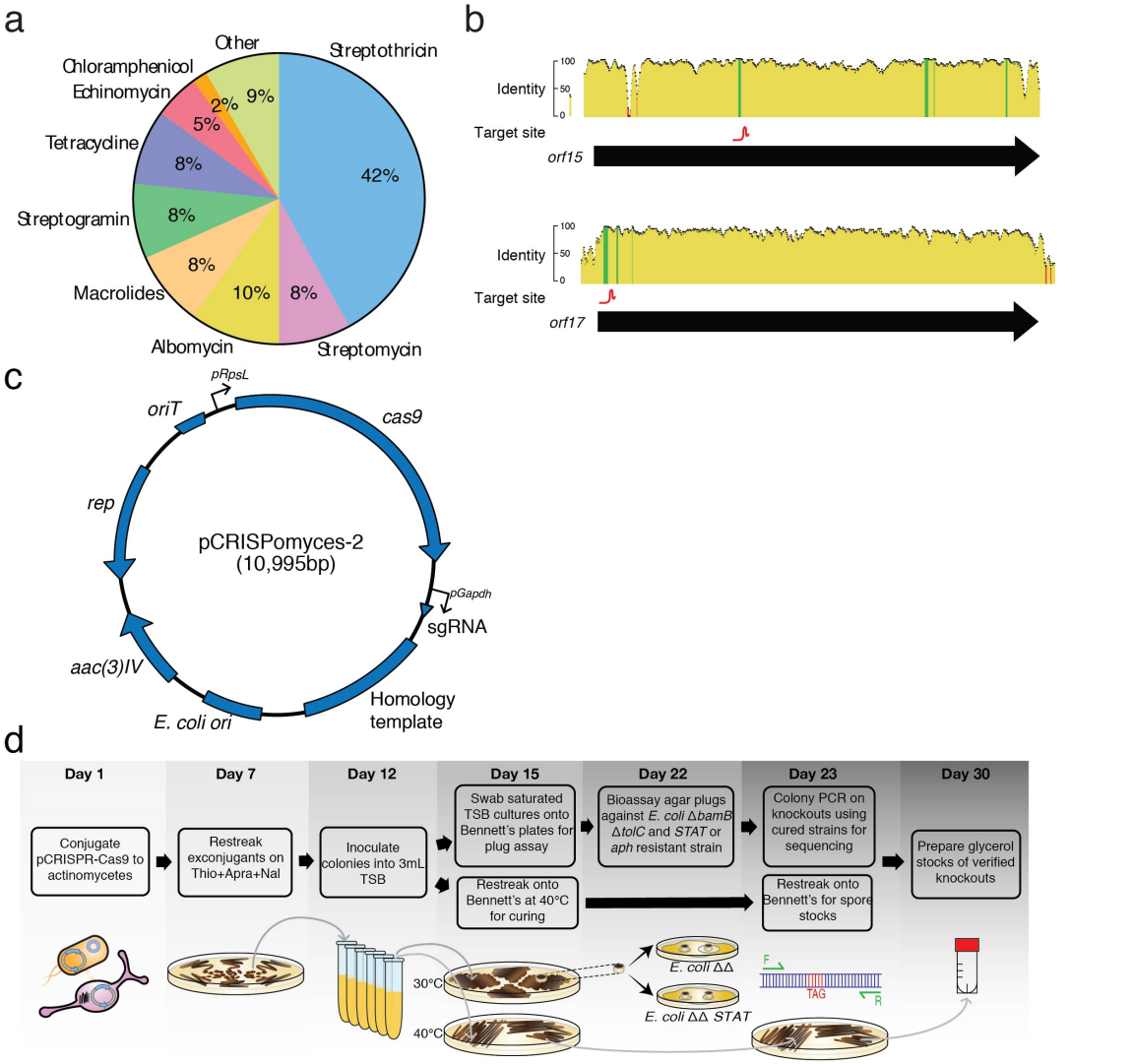
40. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14,** 407–410 (2017).

41. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7,** e1002195

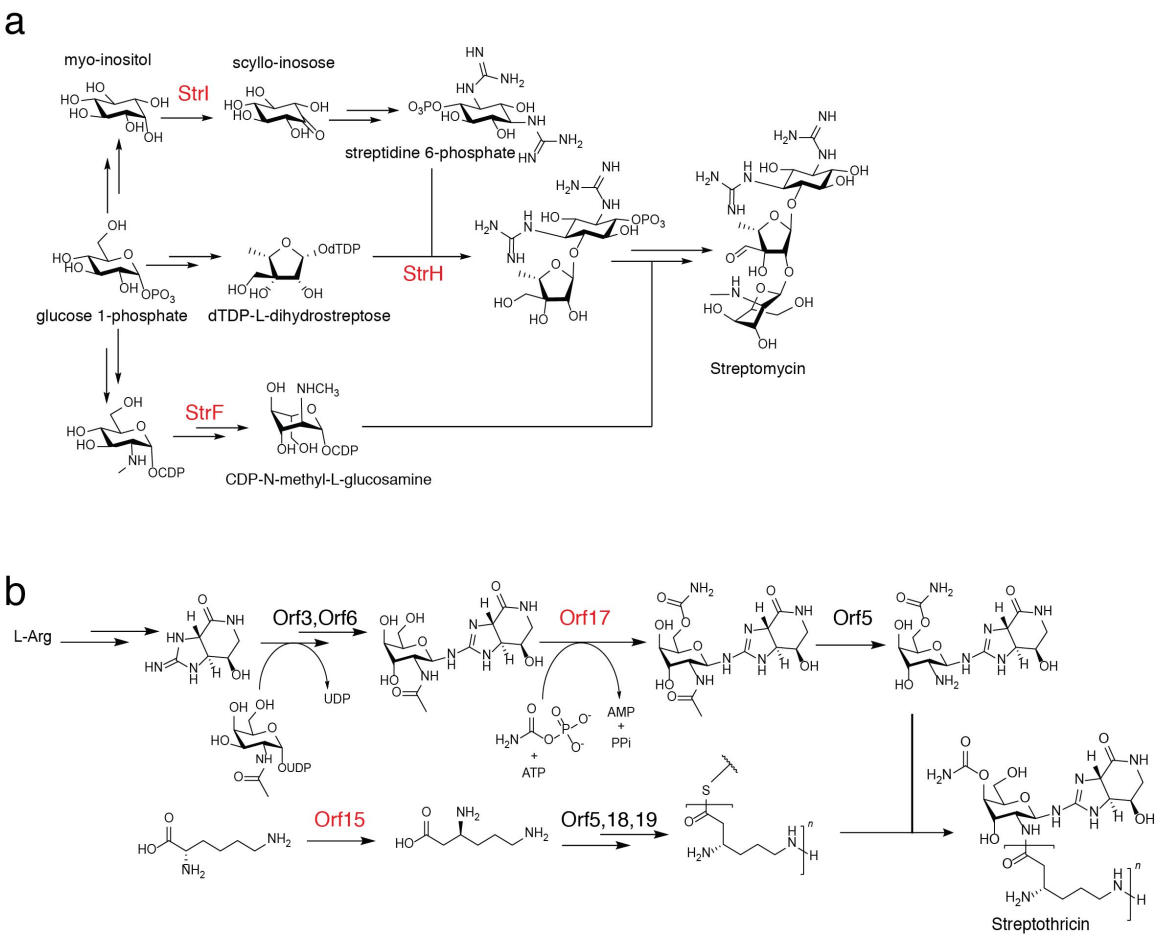
(2011).

42. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).

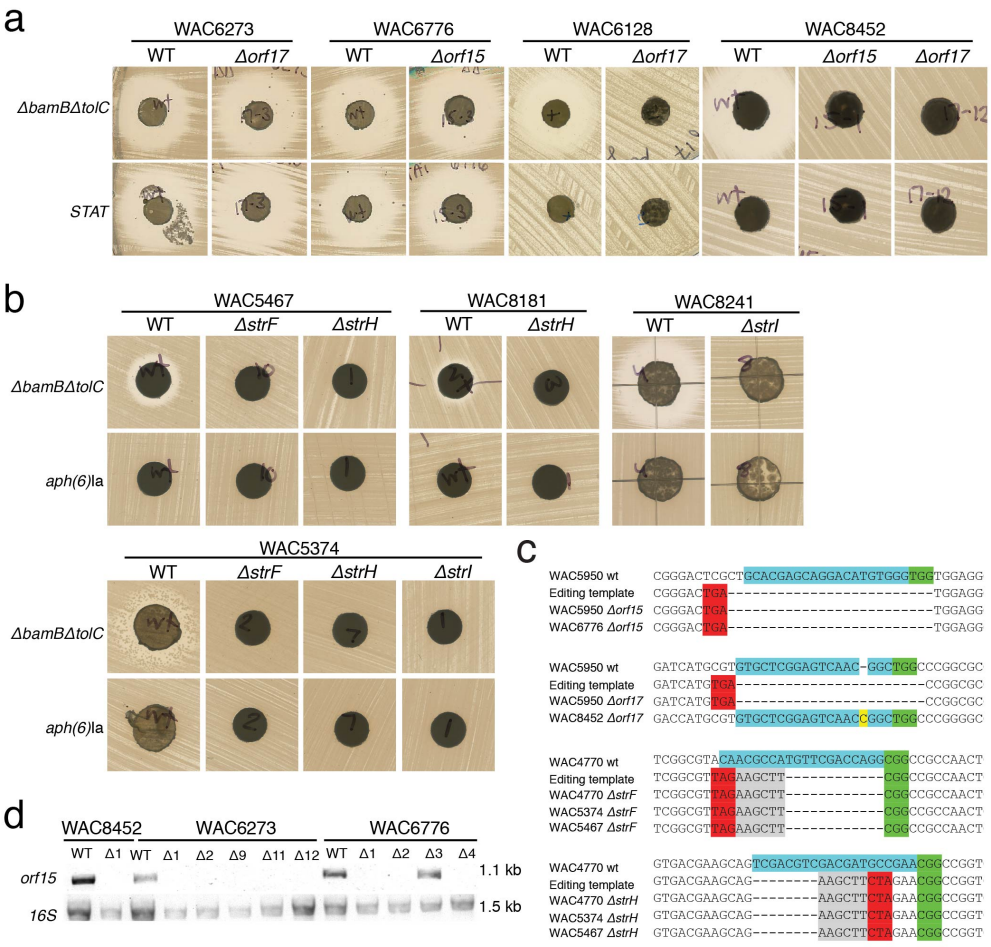
SUPPLEMENTARY INFORMATION



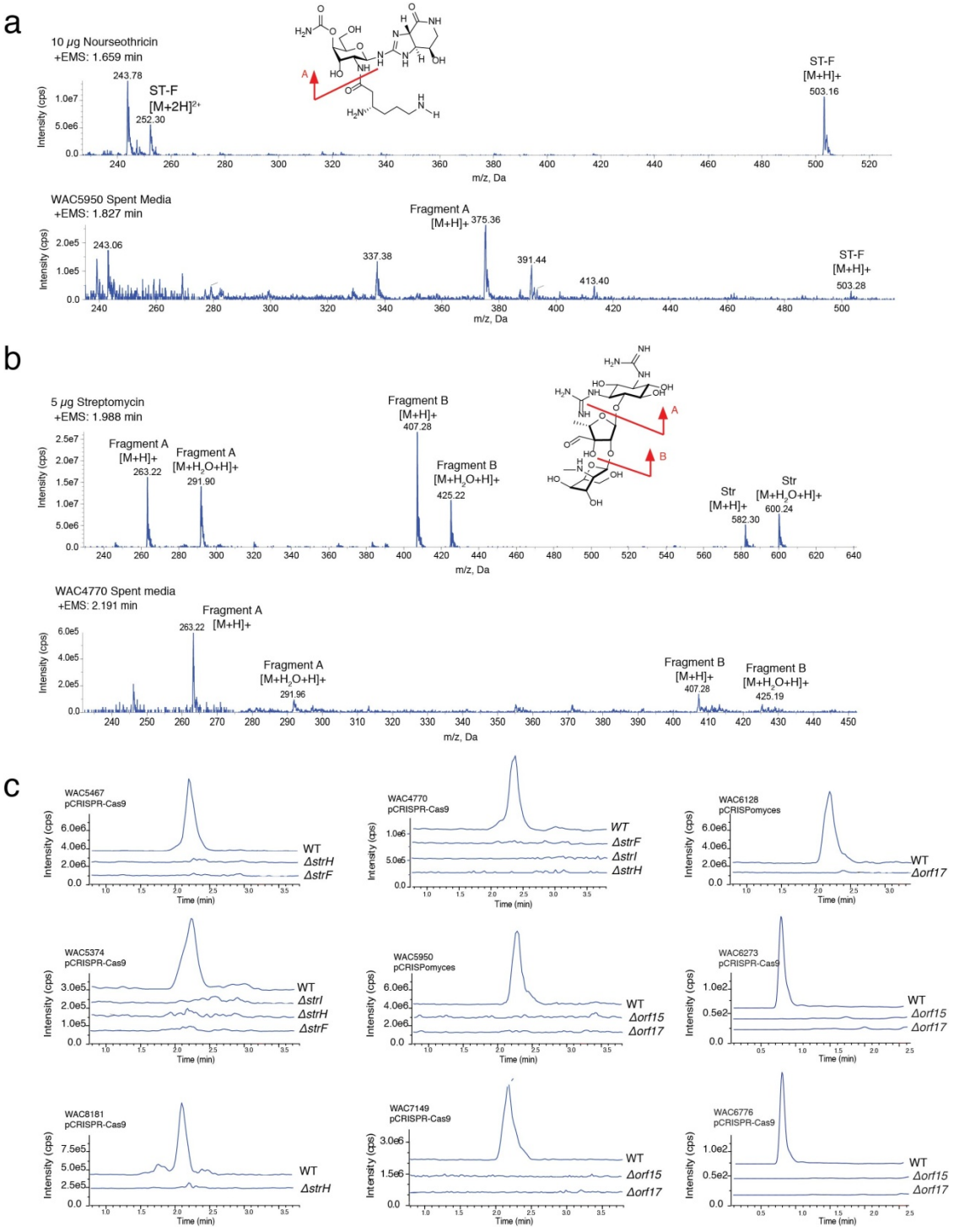
Supplementary Figure 1. **Rationale and workflow for CRISPR-Cas9 inactivation of common antibiotics.** (A) Pie chart showing the frequency of discovery of different antibiotics from 60 Gram negative active actinomycete extracts. Frequencies are calculated from Cox *et al.*[2] (B) Nucleotide identity (%) is shown across targeted biosynthetic genes in the streptothricin gene cluster. Identity was calculated by alignment of 29 streptothricin BGCs. The locations of target sites are shown by red sgRNAs and regions of 100% sequence conservation are indicated with green vertical bars. (C) Plasmid map of the pCRISPomyces-2 system[14]. (D) Key steps for a streamlined and generalizable strategy to efficiently inactivate production of a commonly found antibiotic using the pCRISPR-Cas9 system. Once targeting plasmids are constructed, the process can be completed in 30 days with just six days of hands-on work.



Supplementary Figure 2. Overview of streptomycin (A) and streptothricin (B) biosynthetic pathways. Enzymes targeted for inactivation are shown in red. StrF is an isomerase involved in CDP-N-methyl-L-glucosamine synthesis. StrH is a glycosyltransferase responsible for condensing streptidine-6-phosphate and dihydrostreptose. StrI is a dehydrogenase involved in streptidine 6-phosphate production. Orf15 is a transaminase responsible for the conversion of α to β-Lys and Orf17 is a carbamoyltransferase acting on the D-gulosamine sugar. The number of ß-lysines (n) present in streptothricin can vary from one to seven and correspond to streptothricin F through A and X, respectively[13].

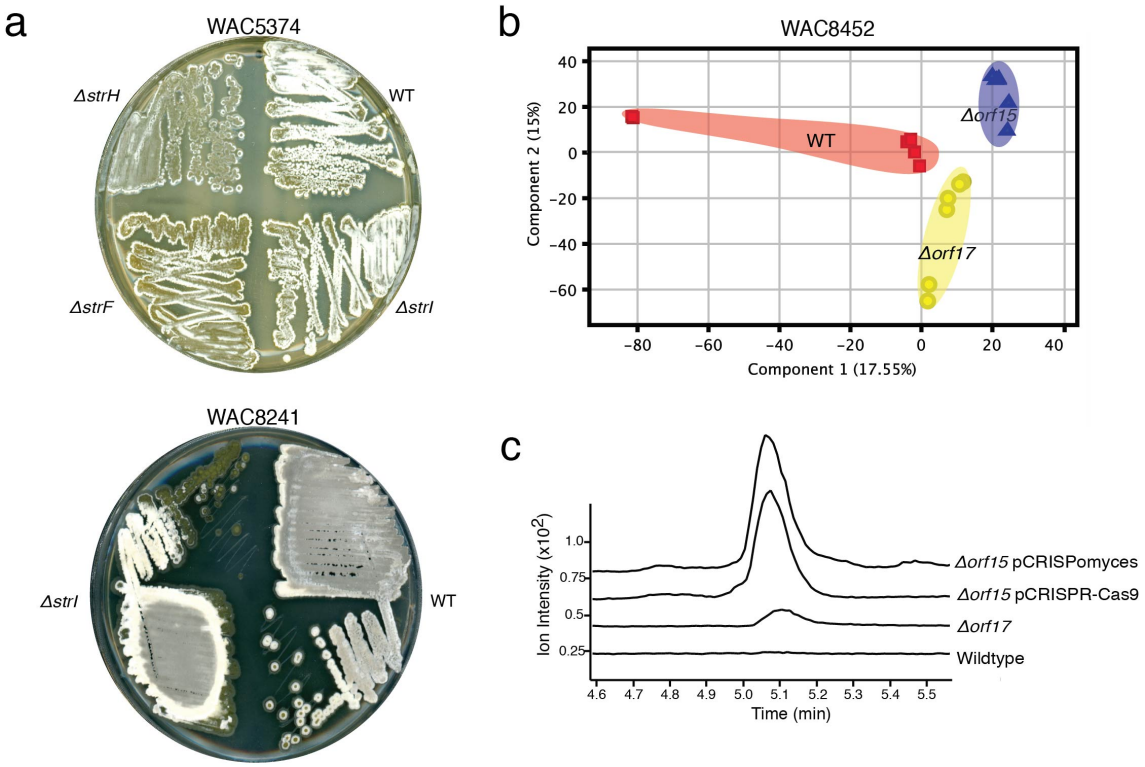


Supplementary Figure 3. Bioactivity and sequence verification of streptothricin and streptomycin inactivated strains. Streptothricin (A) and streptomycin (B) CRISPR-mediated gene inactivations were verified by agar plug bioassays with susceptible and resistant *E. coli.* Representative data is shown for one clone of each species/gene pair that was successfully knocked out. Two independent fermentations and bioassays for each strain gave similar results. (C) Sanger sequencing of CRISPR-engineered strains. Sequence alignments highlight the sgRNA target site (blue) and PAM (green). HR results in the introduction of a stop codon (red) and in some cases a *Hin*dIII site (grey). NHEJ resulting in indels (yellow) was also detected. (D) Attempts to PCR amplify the targeted gene failed to yield an amplicon in some cases and was interpreted as an undefined deletion in the genome. Representative results are shown for several clones (Δ1, Δ2, *etc.*) of three streptothricin producers targeted for *orf15.* 16S PCR and Sanger sequencing was used to verify template DNA integrity and the strain's identity. Deletions shown in this figure were generated using the pCRISPR-Cas9 system except for WAC6128 Δ*orf17* generated using pCRISPomyces.

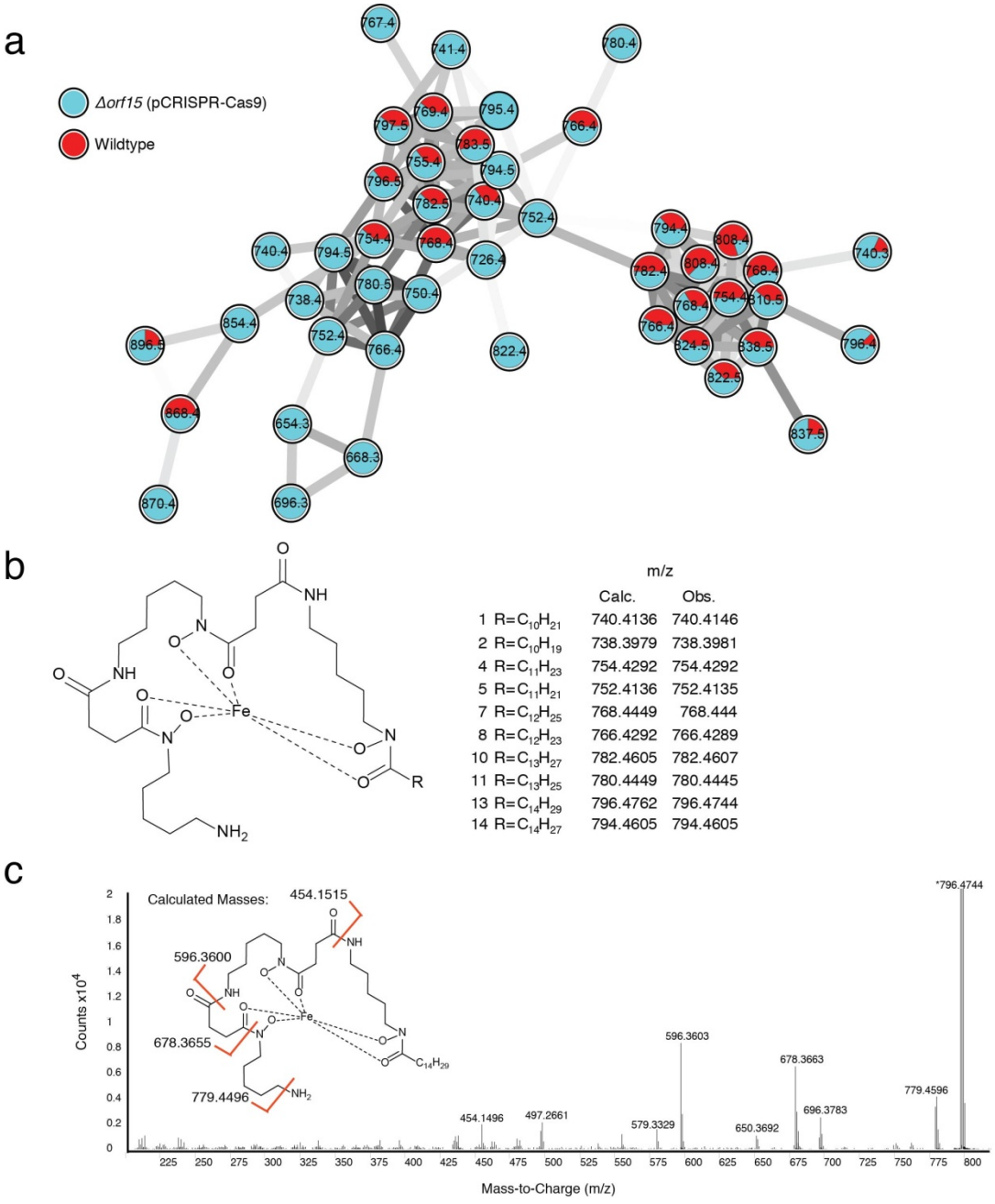


Supplementary Figure 4. LC/MS verification of streptothricin and streptomycin inactivated strains. (A-B) LC-MS profiles of antibiotic standards and spent media confirm antibiotic production in wildtype strains. Nourseothricin is a mixture of streptothricin analogs (F and D). Compound fragments can also be detected, and fragmentation schemes are shown. (C) Extracted ion chromatograms (EIC) verified the

absence of streptomycin or streptothricin for at least one clone of each species/gene pair. LC-MS was performed on a Q-trap instrument in positive mode and extracted for streptothricin F (*m/z*: 502.723-505.924) or an abundant streptomycin fragment (*m/z*: 407.029-407.529). WAC6776 and WAC6273 were analyzed using a high-resolution Q-TOF LC/MS and extracted for *m/z*: 503.25. Two independent fermentations and analysis for each strain gave similar results.

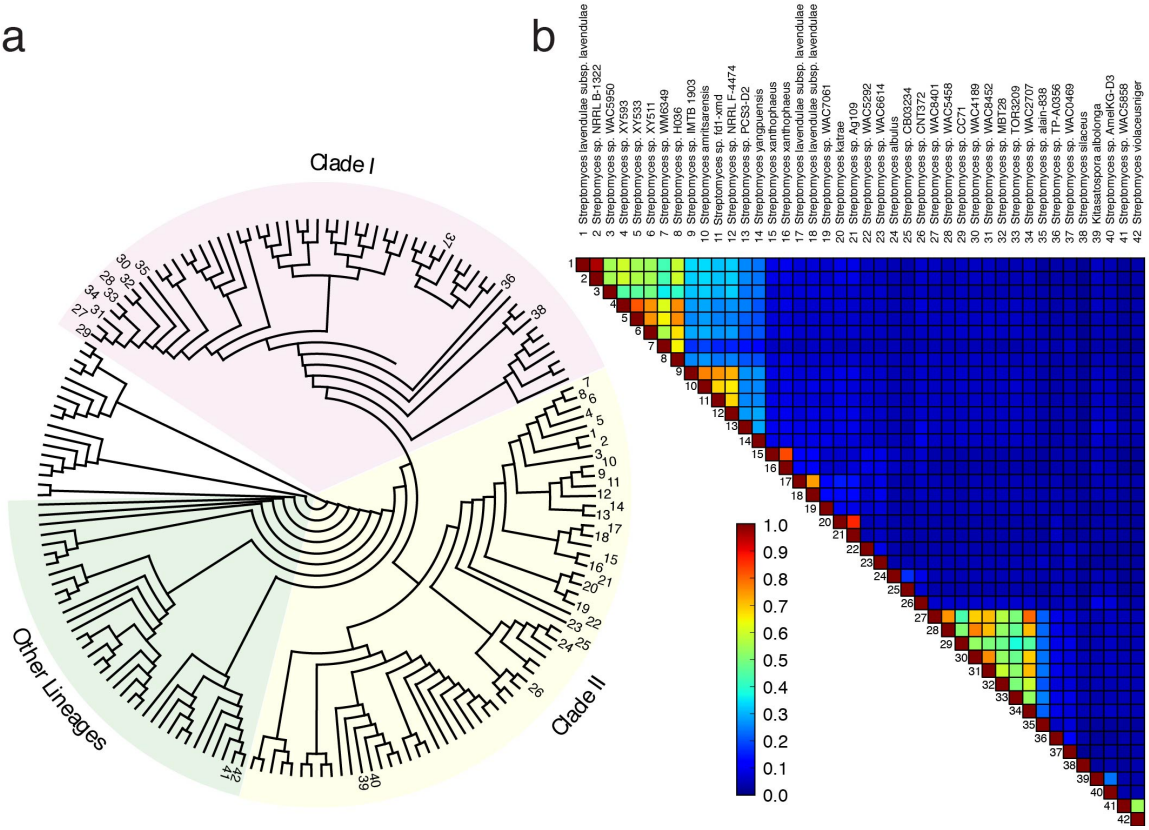


Supplementary Figure 5. CRISPR-inactivated strains have an altered metabolic profile. Global changes in metabolic flux of wildtype and knockout streptomycin or streptothricin producers were assessed globally by phenotypic analysis on solid media (A) and principle component analysis on high resolution LC/MS profiles of *n*-butanolic extracts (B). Strains were analyzed as described in Figure 2 of the main text. The streptomycin producers WAC5374 and WAC8241, and the streptothricin producer WAC8452 are shown here, while remaining strains are shown in Figure 2. Replicates represent three independent fermentation analyzed in technical duplicate ($n = 6$). (C) High resolution LC/MS analysis on the streptothricin producer WAC6273 reveal different metabolite profiles in CRISPR-inactivated strains compared to the wildtype strain. Extracted ion chromatograms for an unknown metabolite with *m/z* 650.333 are shown. Two independent fermentations and analysis gave similar results.

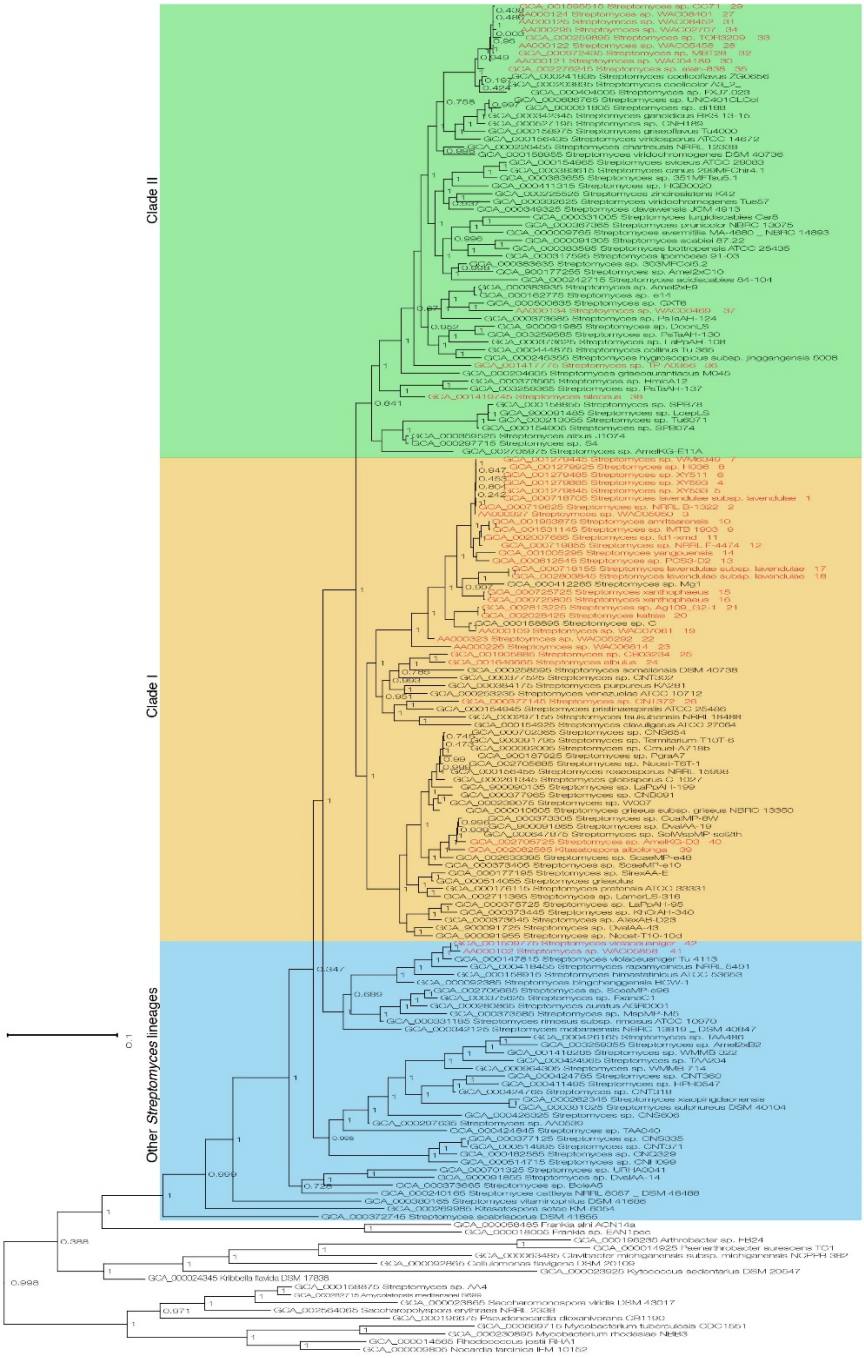


Supplementary Figure 6. Ferrioxamine family of metabolites is upregulated in WAC6273 engineered strains. (A) Metabolic networking of WAC6273 *n*-butanolic extracts clustered a family of masses that were identified as ferrioxamines using the GNPS LC-MS/MS based dereplication platform[18]. Grey edges are shaded from light to dark scaled to the degree of relatedness (cosine score) between nodes. Each node is represented as a pie chart showing the relative abundance of each mass in wildtype WAC6273 (red) or streptothricin knockout *Δorf15* pCRISPR-Cas9 (blue). Almost every member of the network is upregulated in the inactivated strain compared to wildtype, with numerous species being only detectable in the former. Pie charts are drawn based on one

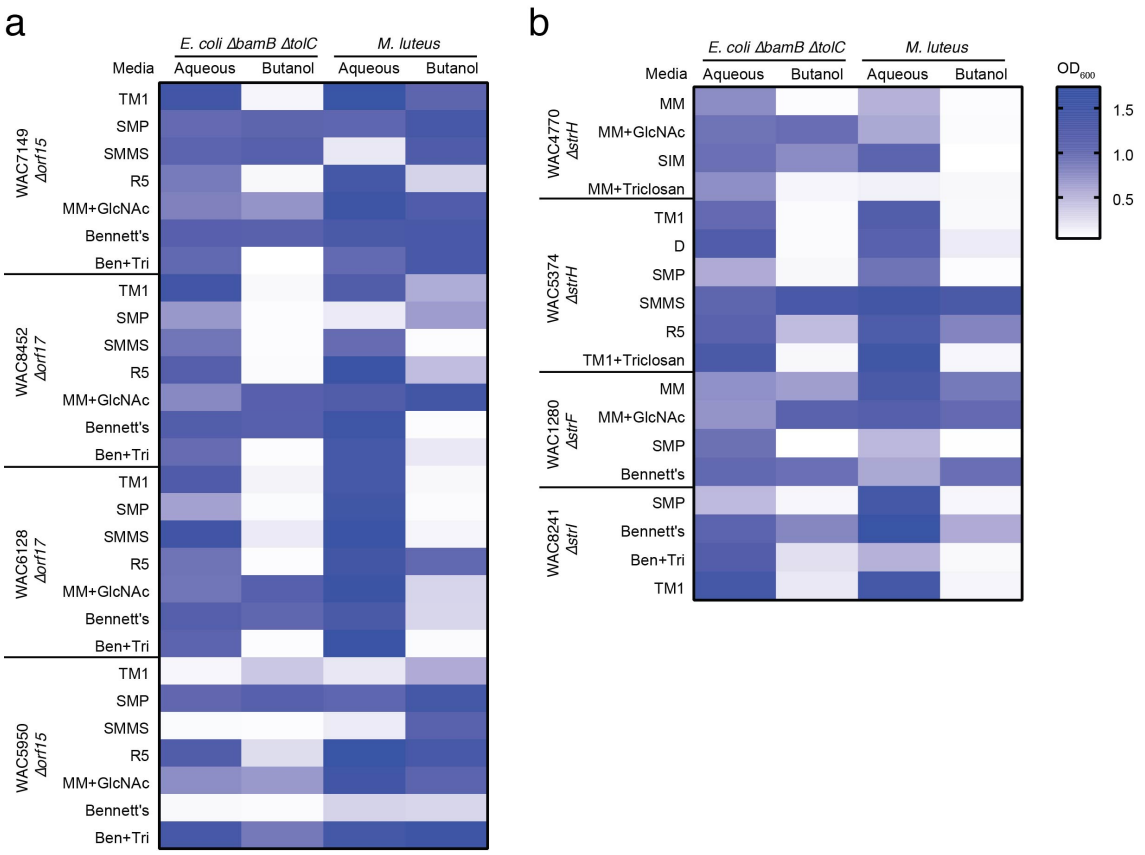
biological replicate but are representative of data collected from three biological replicates. (B) Structure of ferrioxamine family members. Numbering system for different analogs is based on Sidebottom *et al*[19]. Calculated and observed mass of $[M+H]^+$ species are reported for high resolution LC/MS data. (C) High resolution LC-MS/MS was used to confirm network members as ferrioxamines. Calculated masses based on previously reported fragmentation patterns[19] match observed masses in WAC6273 extracts. Fragmentation of a representative member, ferrioxamine 13, is shown. Fragmentation results are representative of two independent analyses.



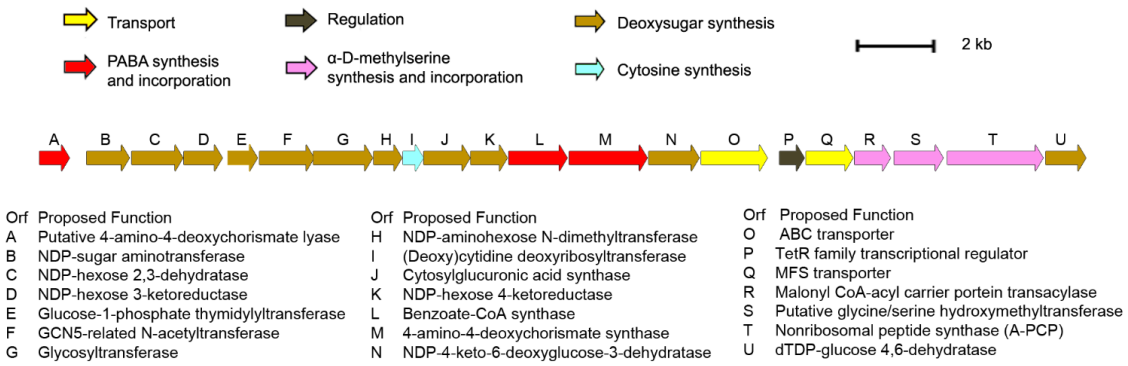
Supplementary Figure 7. Phylogenetic comparison of streptothricin producers and their biosynthetic potential. (A) A phylogenetic tree was constructed using multilocus sequence analysis with 108 core conserved bacterial genes from 173 actinomycetes, including 42 streptothricin producers. This unrooted phylogeny is displayed as a circular cladogram with a midpoint root. Clade II, clade I and other *Streptomyces* lineages are highlighted and described by McDonald *et al*.[20] The uncoloured strains are other non-*Streptomyces* actinomycetes. Streptothricin producers are bolded and numbered in both cladogram and heatmap. An enlarged version of this tree is shown as Supplementary Fig. 8. (B) A heatmap of the similarity between biosynthetic genes/domains of 42 streptothricin producers. Streptothricin producers with similar BGCs are scored closer to 1.0 while strains that share few biosynthetic genes other than those for streptothricin biosynthesis are scored close to 0.



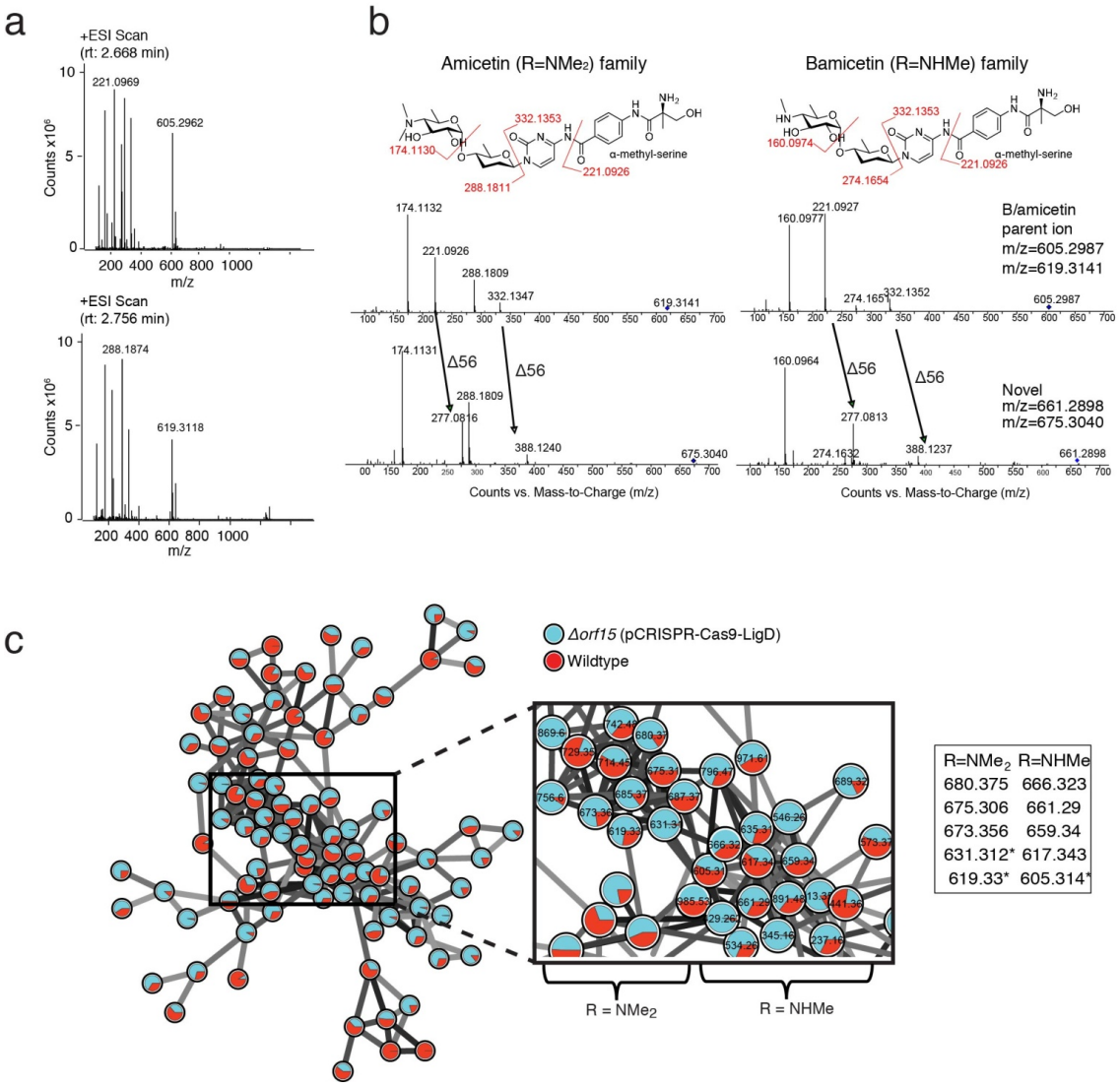
Supplementary Figure 8. Labelled and expanded cladogram of streptothricin producers and other actinomycetes shown in Supplementary Table 4. Interior labels are SH-like support values. Clade II, clade I and other *Streptomyces* lineages are as described by McDonald *et al*.[20] The uncoloured strains are other non-*Streptomyces* actinomycetes. Streptothricin producers are typed in red and numbered as in Supplementary Fig. 7.



Supplementary Figure 9. Antimicrobial activty of spent media of inactivated strains. (A) Streptothricin and (B) streptomycin inactivated producer strains were fermented in various media and *n*-butanolic extracts were tested against *Escherichia coli* BW25113 Δ*tolC* Δ*bamB* and *Micrococcus luteus*. Two independent bioassays gave similar results.

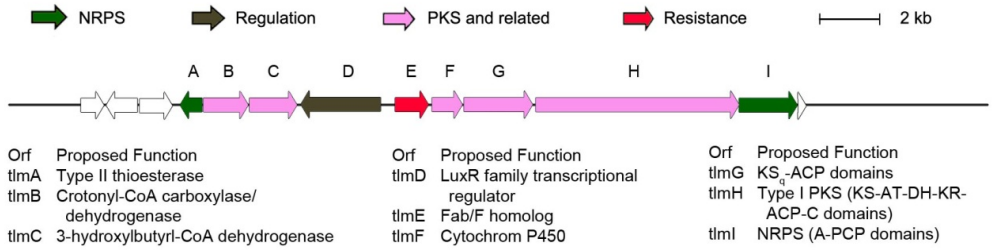


Supplementary Figure 10. **Schematic of amicetin biosynthetic gene cluster in WAC6273.** This BGC has 99.9% nucleotide identity and conserved gene synteny to the amicetin BGC reported from *Streptomyces vinaceusdrappus* NRRL 2363[21].



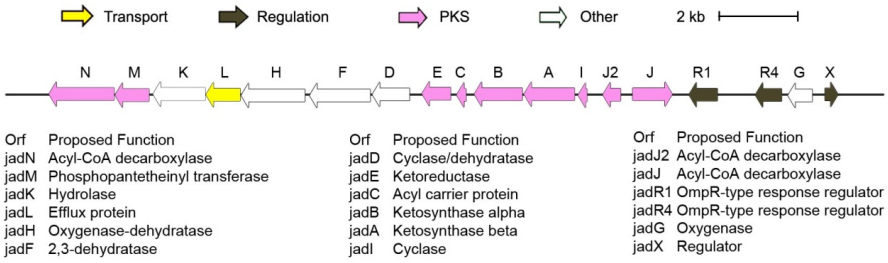
Supplementary Figure 11. **High resolution LC-MS and LC-MS/MS analysis of WAC6273 *Δorf17* fermentation broths.** (A) High resolution LC-MS of WAC6273 *Δorf17* showing the bamicetin (top) and amicetin (bottom) *m/z* values. (B) MS/MS of b/amicetin family of compounds. Both families of compounds show fragments with an additional *m/z* of 56, suggesting that the fermentation broth contains b/amicetin type molecules with replacements of the α-methylserine at the ortho position of *p*-aminobenzoic acid for other amino acids in these novel derivatives, and pointing towards promiscuity in the biosynthetic enzymes responsible for α-methylserine incorporation[21]. LC-MS/MS results are representative of two independent analyses. (C) Molecular network of ions in a WAC6273 *Δorf17* fermentation broth. Edges and nodes are colored as in Supplementary Fig. 6, with each node representing the relative abundance of each mass in wildtype WAC6273 (red) or streptothricin knockout *Δorf15* pCRISPR-Cas9 (blue). The network nodes show parent ions that are related as determined by similar fragmentation patterns present in the MS/MS spectra. The inset lists pairs of parent ions

that have similar fragmentation patterns as b/amicetin. Asterisks indicate masses of the known compounds streptcytosine A, amicetin and bamicetin, suggesting that all other masses represent unknown amicetin derivatives. Pie charts are drawn based on one biological replicate but are representative of data collected from three biological replicates.



Supplementary Figure 12. Schematic of the thiolactomycin biosynthetic gene cluster in WAC5374. This BGC is similar to the *Salinispora pacifica* CNS863 BGC with conserved gene synteny and identity for most genes.

Supplementary Figure 13. Schematic of the 5-chloro-3-formylindole biosynthetic gene cluster in WAC5374. The NRPS encodes a tripeptide with at least one putative tryptophan when analyzed by AntiSMASH[32]. While deletion of the NRPS from WAC5374 Δ*strH* had no effect on the production of 5-chloro-3-formylindole, deletion of the tryptophan halogenase abolished production and confirmed its role in biosynthesis.



Supplementary Figure 14. Schematic of phenanthroviridin aglycone biosynthetic gene cluster in WAC8241. This type II PKS cluster contains the four core genes involved in jadomycin biosynthesis but lacks the jadomycin sugar biosynthesis genes

Supplementary Table 1. Streptothricin and streptomycin BGC and genome sequences used to identify conserved sgRNAs.

	***Streptomyces* strain**	**GenBank accession number**
Streptothricin	WAC5950	NZ_RQJB00000000.1
	TP-A0356	KC935381
	rochei	KT362049
	C	NZ_GG657750.1
	CNT372	NZ_KB897777.1
	TOR3209	NZ_AGNH01000262.1
	lavendulae NRRL B-2774	NZ_JOEW01000007.1
	NRRL B-1322	NZ_JOHF01000002.1
	lavendulae NRRL WC-3542	NZ_JOBU01000003.1
	NRRL F-4474	NZ_JOIB01000008.1
	NRRL F-4835	KL591937.1
	xanthophaeus NRRL B-5414	NZ_JOFT01000018.1
	xanthophaeus NRRL B-3004	NZ_JNZH01000016.1
	PCS3-D2	NZ_JDUZ01000001.1
	MBT28	NZ_LARV01000148.1
	yangpuensis fd2-tb	NZ_LBMK01000002.1
	XY533	NZ_LGDL01000178.1
	WM6349	NZ_LGCZ01000174.1
	XY511	NZ_LGDM01000254.1
	H036	NZ_LGDS01000186.1
	XY593	NZ_LGDJ01000036.1
	TP-A0356	NZ_BBZJ01000015.1
	IMTB 1903	NZ_LQYB01000003.1
	rochei NBRC 12908	AB684619
	SS52	CP039123.1
	fd1-xmd	CP019798.1
	rochei	KT362049.1
	lavendulae BCRC 12163	KF498701.1
	lavendulae NBRC 12789	AB684620.1
Streptomycin	WAC4770	RQIR00000000.1
	WAC1526	VHQN00000000
	bikiniensis NRRL ISP-5582	JOAU01000026.1
	FxanaC1	AQWO01000031.1
	bikiniensis NRRL B-1049	JNWL01000024.1
	bikiniensis NRRL ISP-5580	JNXJ01000025.1
	albulus B-3066	LWBU01000300.1
	griseus DSM 40236	AJ862840
	platensis CR50	GU384160
	griseus NRRL WC-3480	JOBR01000011.1
	MnatMP-M77	FLTQ01000005.1
	ACT-1 XylebKG-1	ADFC02000002.1

Supplementary Table 2. Primers and protospacer sequences. Bold text shows *Hind*III site, red shows stop codons, underlined are sgRNA target sites.

Name/Gene	Sequence	Purpose
Primer Sequences		
ScaligD-F	TCGTCGAAGGCACTAGA**AGGCCT**GCGGTCGATCTTGACGGCTG	
ScaligD-R	GGTCGATCCCCGCATATAGGTGCCGCCGGGCGTTTTTTAT	
sgRNA-F-StrF	CATGCCATGGCAACGCCATGTTCGACCAGGGTTTTAGAGCTAGAAATAGC	
sgRNA-F-StrI	CATGCCATGGACGACATCGACGCCGTGCACGTTTTAGAGCTAGAAATAGC	sgRNA cloning into pCRISPR-Cas9
sgRNA-F-StrH	CATGCCATGGTCGACGTCGACGATGCCGAAGTTTTAGAGCTAGAAATAGC	
sgRNA-F-orf15	CATGCCATGGGCACGAGCAGGACATGTGGGGTTTTAGAGCTAGAAATAGC	
sgRNA-F-orf17	CATGCCATGGGTGTGCTCGGAGTCAACGGCGTTTTAGAGCTAGAAATAGC	
sgRNA-R	ACGCCTACGTAAAAAAAGCACCGACTCGGTGCC	
GGsgRNA-R-orf17	AAACGCCGTTGACTCCGAGCACAC	
GGsgRNA-R-orf15	AAACCCCACATGTCCTGCTCGTGC	sgRNA cloning into pCRISPomyces via golden gate
GGsgRNA-F-orf17	ACGCGTGTGCTCGGAGTCAACGGC	
GGsgRNA-F-orf15	ACGCGCACGAGCAGGACATGTGGG	
Left_StrF_F	TCGTCGAAGGCACTAGAAGGCCTGCTGACCAAACTCGGCGTC	
Left_StrF_R	GGCCGAAGCTTCTAACGCCGAGAACAGTGAACGC	
Right_StrF_F	GGCGTTAGAAGCTTCGGCCGCCAACTCCGACA	
Right_StrF_R	CAGCCGTCAAGATCGACCGC**AGGCCT**AGTTGCGCAGGTGCACGC	
Left_StrI_F	TCGTCGAAGGCACTAGAAGGCCTCCACCTCCACAACGGCAAC	
Left_StrI_R	CAGCCAAGCTTCTAGATGAGGCAGTCGGTGAAGATGTC	Homology template cloning into pCRISPR-Cas9 for streptothricin or streptomycin knockout
Right_StrI_F	TCATCTAGAAGCTTGGCTGACCGGACAGCCC	
Right_StrI_R	CAGCCGTCAAGATCGACCGC**AGGCCT**CGTCTGCTCGACGACGGAAC	
Left_StrH_F	TCGTCGAAGGCACTAGAAGGCCTGGTGATCGACGAACTGGTCG	
Left_StrH_R	AGCAGAAGCTTCTAGAACGGCCGGTCCGTGTT	
Right_StrH_F	CGTTCTAGAAGCTTCTGCTTCGTCACCGACACC	
Right_StrH_R	CAGCCGTCAAGATCGACCGC**AGGCCT**GACGTCGTGGACCACCAGTG	
Left_orf17_F	CCGGCGCCAGCCACGACGG	
Left_orf17_R	AAA**TCTAGA**GCCGGATGCGGTCGCCTTCC	
Right_orf17_R	CCGTCGTGGCTGGCGCCGGTCACATGATCAGTCCCCAAAG	
Right_orf17_F	AAA**TCTAGA**TTGTACGGGACGGCGTTC	Homology template cloning into pCRISPomyces-1
Left_orf15_R	AAA**TCTAGA**CCACTCCTCCAGCACCGG	
Left_orf15_F	TGGAGGGTCTCACCCACCGC	
Right_orf15_F	AAA**TCTAGA**GGCCTGCTTGCGCAGAGC	
Right_orf15_R	GCGGTGGGTGAGACCCTCCATCAGTCCCGGCTGGCCATCGG	

Name/Gene	Sequence	Purpose
pCRISPR-Right-orf15-R	CAGCCGTCAAGATCGACCGCAGGCCTACCACTCCTCCAGCACCGG	Homology template subcloning from pCRISPomyces to pCRISPR-Cas9-LigD
pCRISPR-Left-orf15-F	AGATGGAGTTCTGAGGTCATTACTGGGCCTGCTTGCGCAGAGC	
pCRISPR-Right-orf17-R	CAGCCGTCAAGATCGACCGCAGGCCTCCGGATGCGGTCGCCTTC	
pCRISPR-Left-orf17-F	AGATGGAGTTCTGAGGTCATTACTGTTGTACGGGACGGCGTTC	
orf17-seq-F1	CGCAACGAGGTCGAGCAG	Sequence verification of knockouts
orf17-seq-R1	CGTCGCTGTCGAACCAGTC	
orf17-seq-F2	ACCGGCAGCATGTARCGSC	
orf17-seq-R2	CCGTCCACGACGAGGCAG	
orf15-seq-F	TGCTGTCAGGCGATACGG	
orf15-seq-R	ACGGAGTCGATGTCCAGC	
strI-seq-F	GTGCCGGTGTTCTGCGAGAA	
strI-seq-R	CGAAGCGGTCGGTGAAGTCC	
strF-seq-F	CGGGCAGCCCGGTGTTCAC	
strF-seq-R	GTAGTCGGTGACCCGGCCG	
strH-seq-F	CCCCGTGGTTCTGCTTCTTCTC	
strH-seq-R	CGTCGAAGACGACGCGGAT	
sgRNA-F-halo	CATGCCATGGCCGAATAAACATCACGCTCGGTTTTAGAGCTAGAAATAGC	pCRISPR-Cas9 WAC5374 halogenase knockout
Left_halo_F	TCGTCGAAGGCACTAG**AAGGCCT**GTGAGCTTCTCGGTCAGGTG	
Left_halo_R	TGGTA**AAGCTT**CTATATTCGGTCACCGAACGC	
Right_halo_F	GAATATAG**AAGCTT**TACCAGGACACCCTTCCCAA	
Right_halo_R	GGTCGATCCCCGCATAT**AGGCCT**TGATCTCCCCGATGACCTCC	
Halo-scrn-F	TCCGCCACTTCTTCGAGTTC	
Halo-scrn-R	ACGGAGTACGTCGTCAGGTA	
Halo-seq-F	CTGACATCGCGTCAAGCATG	
Halo-seq-R	TACGGATGTGATTGGCCTCG	
Protospacer sequences		
orf17	GTGTGCTCGGAGTCAACGGC	
orf15	GCACGAGCAGGACATGTGGG	
strI	ACGACATCGACGCCGTGCAC	
strH	TCGACGTCGACGATGCCGAA	
strF	CAACGCCATGTTCGACCAGG	
WAC5374 halogenase	CCGAATAAACATCACGCTCG	

Supplementary Table 3. Summary of streptothricin and streptomycin CRISPR-mediated BGC inactivations. Deletion efficiency is compared to homology arm nucleotide identity and predicted off-target sites. Off-target sites are considered significant if there are less than five mismatches to the actual protospacer sequence and is adjacent to NGG. NS-not significant. * cut off at the edge of a contig, so exact number not available. X conjugation not successful. Ŧ genome sequence not available.

	WAC	Gene	sgRNA identity	Identity to HR template: Left	Identity to HR template: Right	# mismatch to closes off-target site	CRISPR System	Editing efficiency	Sequencing results
Streptomycin	4770	StrI	20/20	100%	100%	2	pCRISPR	82%	Undefined deletion
		StrF	20/20	100%	100%	3	pCRISPR	100%	HR
		StrH	20/20	100%	100%	3	pCRISPR	100%	HR
	5467	StrI	20/20	97%	90%	1	pCRISPR	0%	
		StrF	20/20	96%	98%	NS	pCRISPR	25%	HR
		StrH	20/20	97%	96%	3	pCRISPR	100%	HR
	8181	StrI	20/20	97%	91%	1	pCRISPR	0%	
		StrF	20/20	95%	98%	3	pCRISPR	0%	
		StrH	20/20	97%	97%	3	pCRISPR	21%	Undefined deletion
	5374	StrI	20/20	91%	*64%	2	pCRISPR	12%	Undefined deletion
		StrF	20/20	97%	99%	NS	pCRISPR	60%	HR
		StrH	20/20	97%	84%	3	pCRISPR	42%	HR and deletion (~600 kb)
	8241	StrI	20/20	97%	91%	1	pCRISPR	7%	Deletion (~500kb)
		StrF	20/20	95%	98%	3	pCRISPR	0%	
		StrH	20/20	97%	97%	2	pCRISPR	0%	
Streptothricin	7149	orf15	20/20	80%	91%	3	pCRISPR	8%	Undefined deletion
							pCRISPomyces	0%	
		orf17	20/20	78%	89%	5	pCRISPR	8%	Undefined deletion
							pCRISPomyces	17%	Undefined deletion
	8452	orf15	20/20	86%	72%	3	pCRISPR	25%	Undefined deletion
							pCRISPomyces	0%	
		orf17	20/20	88%	67%	NS	pCRISPR	67%	NHEJ - 1bp insertion
							pCRISPomyces	0%	
	6273	orf15	20/20	75%	86%	3	pCRISPR	92%	Deletion (~200kb)
							pCRISPomyces	25%	Undefined deletion

WAC	Gene	sgRNA identity	Identity to HR template: Left	Identity to HR template: Right	# mismatch to closes off-target site	CRISPR System	Editing efficiency	Sequencing results
	orf17	20/20	54%	88%	NS	pCRISPR	55%	Undefined deletion
						pCRISPomyces	0%	
6776 Ŧ	orf15	20/20				pCRISPR	58%	HR
						pCRISPomyces	8%	Undefined deletion
	orf17	20/20				pCRISPR	75%	Undefined deletion
						pCRISPomyces	8%	HR
5458	orf15	20/20	86%	72%	3	pCRISPR	x	
						pCRISPomyces	x	
	orf17	20/20	88%	67%	NS	pCRISPR	x	
						pCRISPomyces	0%	
6128	orf15	20/20	75%	86%	3	pCRISPR	x	
						pCRISPomyces	x	
	orf17	20/20	52%	88%	NS	pCRISPR	x	
						pCRISPomyces	50%	Undefined deletion
4189	orf15	20/20	86%	72%	3	pCRISPR	x	
						pCRISPomyces	x	
	orf17	20/20	88%	68%	NS	pCRISPR	0%	
						pCRISPomyces	x	
8401	orf15	20/20	86%	72%	4	pCRISPR	x	
						pCRISPomyces	x	
	orf17	20/20	88%	67%	NS	pCRISPR	0%	
						pCRISPomyces	x	
5950	orf15	20/20	100%	100%	3	pCRISPomyces	17%	One by HR, One large deletion (~1.5Mb)
	orf17	20/20	100%	100%	NS	pCRISPomyces	17%	HR

Supplementary Table 4. Additional data for Supplementary Fig. 7-8. The number of BGCs annotated by antiSMASH[32], the number of genes/domains in the BGCs of the given organism and the number of unique genes/domains in the BGCs of the given organism are listed. If a genome is separated into many contigs, the BGCs are likely present on more than one contig, each partial piece being counted as a BGC by antiSMASH. Thus, the number of unique genes/domains are likely a better representation of BGC diversity compared to the number of BGCs. Unique genes/domains were used for constructing the heatmap in Supplementary Fig. 7. Labels correspond to those in Supplementary Fig. 7-8.

Label	Deposited Taxon	# of BGCs	# of Genes/ Domains	# Unique Genes/ Domains
1	*Streptomyces lavendulae* subsp. lavendulae strain NRRL WC-3542	33	1732	1358
2	*Streptomyces* sp. NRRL B-1322	33	1412	1349
3	*Streptomyces* sp. WAC 05950	46	1180	855
4	*Streptomyces* sp. XY593	35	931	880
5	*Streptomyces* sp. XY533	34	829	789
6	*Streptomyces* sp. XY511	30	899	815
7	*Streptomyces* sp. WM6349	27	778	660
8	*Streptomyces* sp. H036	37	1027	961
9	*Streptomyces* sp. IMTB 1903	28	1138	954
10	*Streptomyces* amritsarensis strain MTCC 11845	33	899	798
11	*Streptomyces* sp. fd1-xmd	28	1282	1054
12	*Streptomyces* sp. NRRL F-4474	30	1183	924
13	*Streptomyces* sp. PCS3-D2 isolate P1	27	1390	954
14	*Streptomyces* yangpuensis strain fd2-tb	28	1363	946
15	*Streptomyces* xanthophaeus strain NRRL B-5414	40	1399	1134
16	*Streptomyces* xanthophaeus strain NRRL B-3004	35	1628	1055
17	*Streptomyces* lavendulae subsp. lavendulae strain NRRL B-2774	44	1752	1072
18	*Streptomyces* lavendulae subsp. lavendulae strain CCM 3239	28	1775	1264
19	*Streptomyces* sp. WAC 07061	53	1114	640
20	*Streptomyces* katrae strain S3 genome	23	2270	1389
21	*Streptomyces* sp. Ag109_G2-1	25	2028	1481
22	*Streptomyces* sp. WAC 05292	40	955	562
23	*Streptomyces* sp. WAC 06614	46	1008	602
24	*Streptomyces* albulus strain B-3066	37	1710	1024
25	*Streptomyces* sp. CB03234	35	1850	1117
26	*Streptomyces* sp. CNT372	29	1401	846
27	*Streptomyces* sp. WAC 08401	38	1477	1141
28	*Streptomyces* sp. WAC 05458	36	1499	1098
29	*Streptomyces* sp. CC71	34	781	671
30	*Streptomyces* sp. WAC 04189	33	1584	1062
31	*Streptomyces* sp. WAC 08452	33	1226	955
32	*Streptomyces* sp. MBT28	47	1143	875
33	*Streptomyces* sp TOR3209	55	824	668
34	*Streptomyces* sp. WAC 02707	39	1313	1054
35	*Streptomyces* sp. alain-838	41	633	455
36	*Streptomyces* sp. TP-A0356	21	1635	929

Label	Deposited Taxon	# of BGCs	# of Genes/ Domains	# Unique Genes/ Domains
37	*Streptomyces* sp. WAC 00469	20	597	372
38	*Streptomyces silaceus* strain NRRL B-24166	58	2562	1354
39	*Kitasatospora albolonga* strain YIM 101047	37	2671	1598
40	*Streptomyces* sp. AmelKG-D3	39	2158	1319
41	*Streptomyces* sp. WAC 05858	79	3447	1761
42	*Streptomyces violaceusniger* strain NRRL F-8817	61	2349	1555

Supplementary Table 5. **MIC (µg/ml) of bioactive compounds measured against various strains including common pathogens.**

Compound	Amicetin	Bamicetin	5-chloro-3-formylindole
E.coli BW25113	256	512	
E.coli BW25113 Δ*tolC* Δ*bamB*	2	4	64
B. subtilis 168	4	4	
K. pneumoniae ATCC 33495	256	>512	
P. aeruginosa PAO1	>512	>512	
A. baumannii ATCC 17978	512	256	
S. aureus ATCC 29213	4	8	
M. smegmatis mc^2155	5	2.5	

Supplementary Table 6. Strains and plasmids used in this study.

Strains	Description	Source
Escherichia coli		
Top10	General cloning and plasmid maintenance	Invitrogen
ET12567	*dam-13*::Tn*9 dcm-6 hsdMRS* Cmr	FEMS Microbiol. Lett. 155, 223–229 (1997)
BW25113 Δ*bamB ΔtolC*	hyperpermeable strain for bioassay	Nature 510, 503–6 (2014).
BW25113 Δ*bamB ΔtolC* pGDP1:S*TAT*	streptothricin resistant strain for bioassay	2
BW25113 Δ*bamB ΔtolC* pGDP3:*aph(6)*Ia	streptomycin resistant strain for bioassay	2
Other		
Micrococcus luteus ATCC4698	strain for bioassay	ATCC
WAC####	environmental isolates	Wright Lab, Bioproject PRJNA504665
Plasmids		
pUZ8002	conjugation helper plasmid, RK2 derivative with *oriT* mutation	30
pCRISPR-Cas9	*oriT*pSG5*ori aac(3)*IV PtipA::*Cas9* and *scaligD* with extra *stuI* site for homology arm cloning	9 and this work
pCRISPomyces-2	*oriT*pSG5*ori aac(3)*IV PrpsL::*Cas9*	14
pCRISPR-Cas9-orf15	pCRISPR-Cas9 derivative targeting orf15 in streptothricin gene cluster	This work
pCRISPR-Cas9-orf17	pCRISPR-Cas9 derivative targeting orf17 in streptothricin gene cluster	This work
pCRISPR-Cas9-strI	pCRISPR-Cas9 derivative targeting strI in streptomycin gene cluster	This work
pCRISPR-Cas9-strF	pCRISPR-Cas9 derivative targeting strF in streptomycin gene cluster	This work
pCRISPR-Cas9-strH	pCRISPR-Cas9 derivative targeting strH in streptomycin gene cluster	This work
pCRISPomyces-orf15	pCRISPomyces-2 derivative targeting orf15 in streptothricin gene cluster	This work
pCRISPomyces-orf17	pCRISPomyces-2 derivative targeting orf17 in streptothricin gene cluster	This work

Chapter III – Evolution-guided discovery of antibiotics with a novel mode of action

PREFACE

The work presented in this chapter was previously published in:

Culp, E.J., Waglechner, N., Wang, W. *et al.* Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* **578,** 582–587 (2020). https://doi.org/10.1038/s41586-020-1990-9

E.J.C., N.W. and G.D.W. conceived the study and designed experiments. N.W. performed phylogenetic analysis, structural predictions and resistant mutant genome analysis. W.W. performed compound purification and structural elucidation. E.J.C., A.F.C. and B.K.C. designed animal studies and A.F.C. performed animal experiments. E.J.C. and D.S. performed PG and enzyme purification. E.J.C. and K.K. performed BODIPY derivative synthesis. Y.H. and Y.V.B. designed FDAA studies, M.V. provided FDAAs and Y.H. performed experiments. E.J.C. performed all other experiments. E.J.C. and G.D.W. prepared the manuscript.

ABSTRACT

Addressing the ongoing antibiotic crisis requires the discovery of compounds with novel mechanisms of action capable of treating drug resistant infections[1]. The source of many antibiotics are specialized metabolites produced by bacteria, particularly of the Actinomycete family[2]. While actinomycete extracts have traditionally been screened using activity-based platforms, this approach has fallen out of favor due to the frequent rediscovery of known compounds. Genome sequencing of actinomycetes reveals an untapped reservoir of biosynthetic gene clusters (BGCs), but requires prioritization to predict which gene clusters may yield promising novel chemical matter[2]. Here we make use of the phylogeny of biosynthetic genes along with the lack of known resistance determinants to predict divergent members of the glycopeptide family of antibiotics (GPAs) that are likely to possess new biological activities. Using these predictions, we uncovered two members of a new functional class of GPAs, the known GPA complestatin and a newly discovered compound corbomycin, that have a novel mode of action. We show that by binding to peptidoglycan, complestatin and corbomycin block the action of autolysins, essential peptidoglycan hydrolases required for remodeling of the cell wall during growth. Corbomycin and complestatin have low levels of resistance development and are effective in reducing bacterial burden in a mouse model of skin MRSA infection.

MAIN

Throughout evolution, BGCs are sculpted at least in part by selective pressure on the resultant metabolite's biological activity. Therefore, BGCs that have evolutionarily diverged biosynthetic genes might also produce novel biological activity. Such

phylogeny-guided discovery has previously been applied to identify divergent members of natural product families by generating phylogenetic trees from short sequence tags of select biosynthetic genes[3–5]. Alternatively, we have previously shown that by combining the selection of antibiotic resistance with phylogenetic analysis of concatenated biosynthetic gene segments, novel members of an antibiotic class retaining the same mechanism of action can be identified[6]. These approaches are fruitful in prioritizing BGCs producing undiscovered compounds with biological activity but do not inform on mechanism of action. Antibiotic BGCs encode not only biosynthetic machinery, but also resistance genes to protect from self-intoxication. Searching for the presence of resistance genes in BGCs has previously directed genome mining for antibiotics with known or predicted mechanisms[7,8]. Interested instead in new modes of action, we hypothesized that by identifying phylogenetically distinct BGCs lacking known self-resistance genes, we could find antibiotics with different and possibly novel mechanisms.

Resistance and phylogeny guide discovery

We applied our hypothesis to the GPA family, chosen for its BGC diversity, high degree of tailoring, and distinctive self-resistance genes through target modification (e.g. *vanHAX*). We collected 71 GPA and GPA family BGCs from in-house genome sequences and public repositories and constructed phylogenetic trees of every shared gene and gene segment found in these BGCs[9]. By integrating our knowledge about species phylogeny with the lineage of individual genes/domains, we identified trees informative about the conservation or divergence of components due to function (e.g. non-ribosomal peptide synthetase, NRPS) rather than the relationship between strains on a species level (e.g.

precursor supply)[9]. The relationship we observed in a tree of particular NRPS condensation domains exemplifies divergence due to changing function, and was present in several phylogenies analyzed (Fig. 1a and Extended data fig. 1a,b). Mapping the presence of self-resistance genes onto these trees, we found that BGCs containing common resistance determinants for 'true' GPAs (e.g. *vanHAX* and *vanY* for antibiotics binding D-Ala-D-Ala of lipid II) fell within a single clade (Fig. 1a). However, in BGCs with divergent evolutionary histories and a distinct common ancestor from 'true' GPAs, we noted two clades that lacked any known GPA resistance determinant indicating that they may also possess divergent biological activity.

One of these divergent clades contained a known compound, complestatin (Fig. 1b), as well as other members predicted to be structural variants. We purified complestatin from *Streptomyces* sp. WAC01325 harbouring this BGC (Extended data fig. 1c)[10]. The second divergent clade contained BGCs with no characterized members and so we used our phylogenetic trees to make structural predictions and guide purification of the resulting metabolite from *Streptomyces* sp. WAC01529 (Fig. 1b, Extended data fig. 1d,e and Supplementary figures 1-6). We named it corbomycin (after *Nid de Corbeau*, or Crow's Nest Pass in Alberta, Canada where *Streptomyces* sp. WAC01529 was collected) and the corresponding BGC *crb* (Extended data fig. 1d).

Corbomycin and complestatin belong to the type V family of GPAs[11] including other known members chloropeptin I, neuroprotectin A/B and kistamicins[12,13]. Antibiotic activity of these compounds has been noted but not investigated in depth[14,15]. The lack of

known resistant determinants in these BGCs therefore warranted further investigation of their mechanism of action.

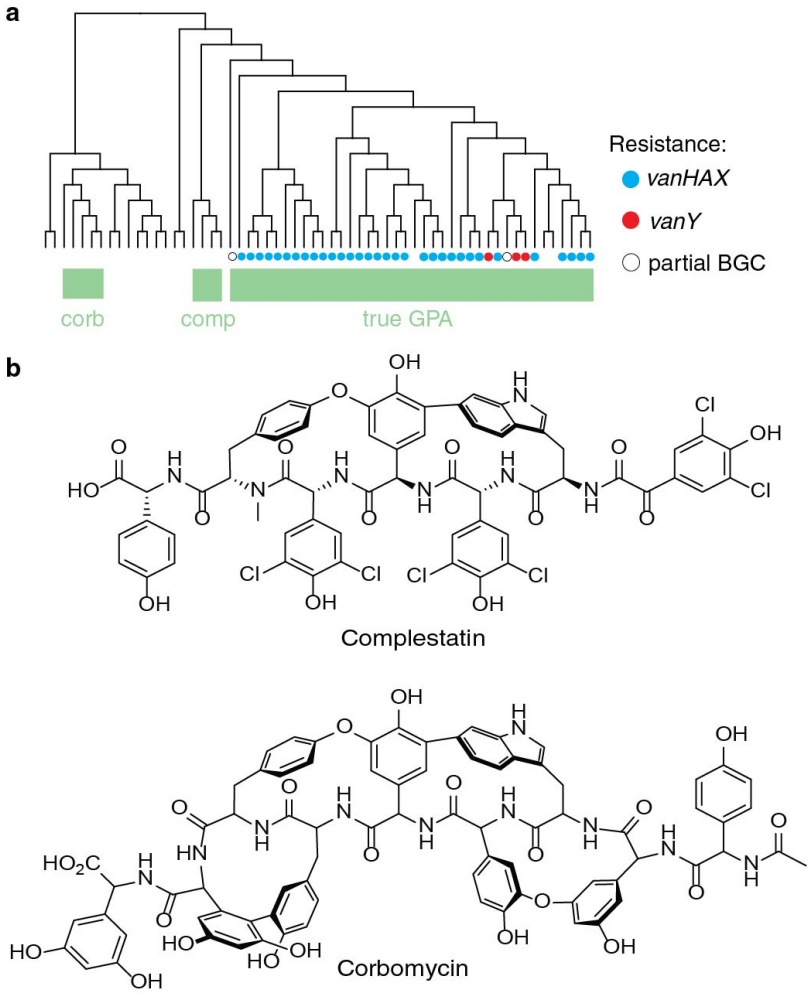


Figure 1. Phylogeny-guided discovery of complestatin and corbomycin. (A) A maximum-likelihood phylogenetic tree of condensation domains at the C-2 module according to labelling scheme shown in Extended data fig. 1a, and the full tree as Extended data fig. 1b. The presence of true GPA self-resistance genes including *vanHAX* and *vanY* present within each corresponding BGC is marked by colored circles. Tree is shown as a rectangular cladogram with midpoint root. Labels indicate well supported clades (> 0.89 SH-like support values) of condensation domains for the true GPAs, complestatin (comp)-type BGCs and corbomycin (corb)-type BGCs. Branches not marked by one of these BGC family labels are uncharacterized. Scaled and terminal node labeled trees are available at http://github.com/waglecn/GPA_evolution. (B) Structures of complestatin and corbomycin.

Identification of a novel mode of action

Corbomycin and complestatin have primarily anti-Gram positive activity with minimum inhibitory concentrations (MICs) ranging 0.5-4 μg/mL, a potency comparable to vancomycin (Extended data table 1). They are active against a range of laboratory strains and clinically relevant pathogens including methicillin-resistant and daptomycin resistant *Staphylococcus aureus* (MIC = 0.5-2 μg/mL). Remarkably, they are also active against Vancomycin Resistant *Enterococcus* (VRE) and Vancomycin Intermediate *S. aureus* (VISA), indicating that they have a different mode of action from typical GPAs. Indeed, isothermal titration calorimetry failed to show binding to PG stem pentapeptide under conditions where vancomycin binding is easily measured (K_d = 14.3 μM).

Complestatin's activity against VRE has been previously observed[16], but work done *in vitro* and in *S. aureus* suggested the mechanism was inhibition of fatty acid synthesis by targeting the enoyl-ACP reductase FabI[15]. We questioned whether this was the physiologically relevant target given that complestatin's size (MW = 1328) and physiochemical properties make it unlikely to penetrate the cell membrane to reach intracellular FabI. Target overexpression, knockout and exogenous fatty acid supplementation (Extended data fig. 2) did not support fatty acid synthesis as the primary target of complestatin or corbomycin in *S. aureus* or *B. subtilis*.

Given GPA family molecule's cell wall activity, we next tested whether corbomycin and complestatin affect peptidoglycan (PG) metabolism. The promoter of *ywaC* in *B. subtilis* is activated almost exclusively by cell envelope acting antibiotics[17], and was robustly activated by corbomycin and complestatin (Fig. 2a). P_{ywaC} is also

activated by compounds targeting the cell membrane[17], but permeabilization was not observed using the voltage-sensitive fluorescent dye $DiSC_3$ (Fig. 2b). Teichoic acid biosynthesis inhibition was tested using a *tagO* mutant, an enzyme involved in early-stage teichoic acid biosynthesis whose knockout abolishes teichoic acid inhibition lethality[18]. Corbomycin and complestatin were equally active against Δ*tagO* and wildtype *B. subtilis* (MIC = 1 μg/mL, Extended data table 1), ruling out an effect on teichoic acid synthesis. We thus focused on PG metabolism as the potential target for corbomycin and complestatin.

To probe which stage of PG metabolism is blocked by complestatin and corbomycin, we measured the buildup of the final cytoplasmic intermediate of PG synthesis, uridine 5'-diphosphate-*N*-acetylmuramic acid-pentapeptide (UDP-MurNAc-PP). Treatment of *S. aureus* with antibiotics acting on lipid-linked steps of PG biosynthesis such as transglycosylation results in UPD-MurNAc-PP accumulation (Fig. 2c). Corbomycin and complestatin had no effect (Fig. 2c), indicating PG metabolism downstream of transglycosylation is affected.

We next tested whether transpeptidation was inhibited by making use of fluorescent D-amino acids (FDAAs) such as HCC-amino-D-alanine (HADA) that are incorporated into actively growing PG by transpeptidases[19]. At concentrations 2xMIC, PG synthesis inhibition by the β-lactam ampicillin was detected by decreased HADA incorporation while corbomycin and complestatin caused no significant change over 1 hr (Fig. 2d, Extended data fig. 3a-c). Furthermore, while all antibiotics that block PG synthesis result in cell lysis through an imbalance between PG synthesis and degradation,

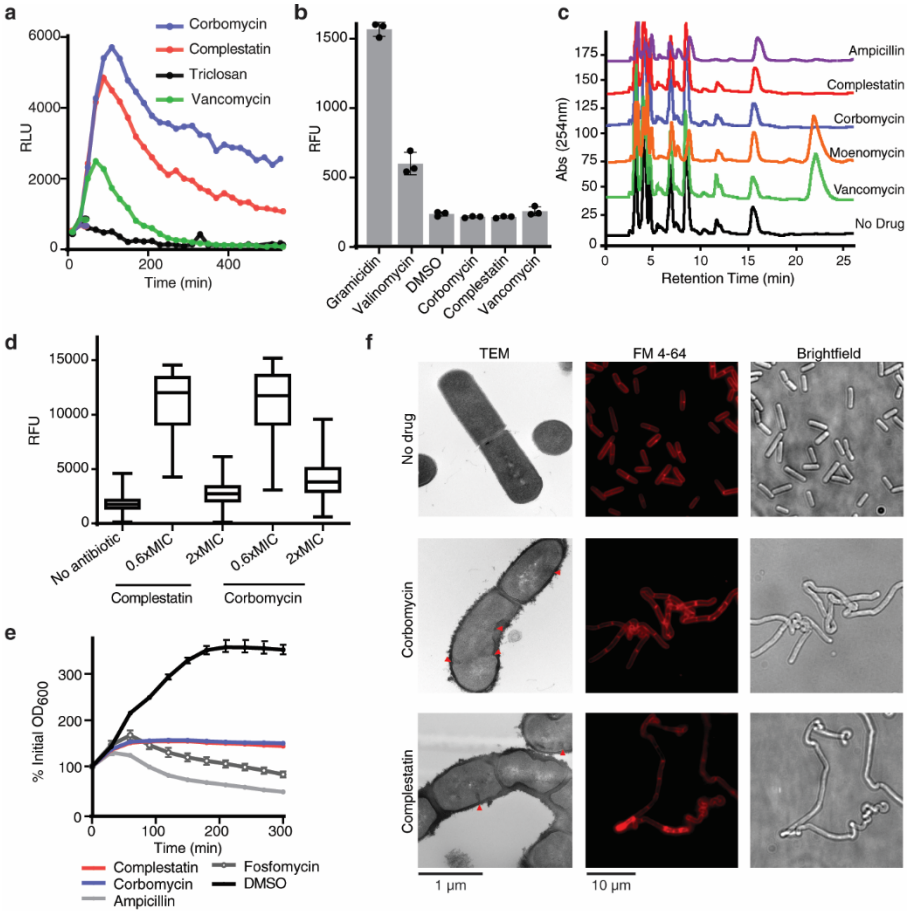


Figure 2. Corbomycin and complestatin affect PG metabolism.
(A) Complestatin and corbomycin strongly activate P_{ywaC} controlling the expression of *lux* genes in *B. subtilis,* as shown in relative luminescence units (RLU). (B) Membrane disruption was measured using release of the voltage sensitive fluorescent dye $DiSC_3$ from *B. subtilis* membranes, quantified as maximal relative fluorescent units (RFU). Gramicidin and valinomycin are membrane active antibiotics and positive controls for this assay. (C) HPLC chromatograms show UDP-MurNAc-PP accumulation in *S. aureus* after treatment with 10x MIC antibiotic, as seen by the peak at 23 minutes in the vancomycin and moenomycin samples. (D) Incorporation of HADA was used to measure PG synthesis and transpeptidase activity in antibiotic treated *B. subtilis*. For samples shown left to right, individual cells quantified $n = 202$, 81, 169, 138, 187. Whiskers show min-max of dataset, box shows upper and lower quartiles, line shows median. (E) Lysis of *B. subtilis* treated with various antibiotics above their MIC shows that corbomycin and complestatin are bacteriostatic, in contrast to PG biosynthesis inhibitors ampicillin and fosfomycin. (F) Microscopy of *B. subtilis* grown in 0.6xMIC corbomycin or complestatin shows a characteristic twisted phenotype. Red triangles on TEM images mark sites of aberrant division septa formation, thickened cell wall and granular formations. For panels B and E, mean of three biological replicates is shown with error bars showing standard deviation. All experiments were repeated at least twice with similar results.

corbomycin and complestatin are bacteriostatic (Fig. 2e, Extended data fig. 3d,e). Complestatin and corbomycin were not synergistic with various cell wall and membrane active antibiotics, but were additive with each other (Extended data fig. 4a). Corbomycin and complestatin therefore target PG metabolism through a similar mechanism that is distinct from any previously reported antibiotic.

MOA through autolysin inhibition

Given complestatin and corbomycin's unprecedented activity on PG metabolism, we examined the phenotype of *B. subtilis* grown in sub-MIC levels of antibiotic. At 0.6xMIC, cells formed septa effectively but failed to divide, instead forming twisted and knotted chains of cells (Fig. 2f). Transmission electron micrographs (TEM) showed aberrant septal structure, absence of flagella, a "shaggy" and thickened cell wall, and distortion of the cytoplasm. This distinctive phenotype was inconsistent with several control antibiotics (Extended data fig. 4b), but matched that of *B. subtilis* strains defective in autolysins[20,21]. This diverse group of PG degrading enzymes is essential for normal PG metabolism by allowing insertion of new material into the existing cell wall and cleavage of PG at the division septa.

We attempted selection for spontaneous resistant mutants on 4x and 8xMIC agar, but were unsuccessful for both *B. subtilis* 168 (frequency of resistance $<3x10^{-9}$) and *S. aureus* ATCC 29213 (frequency of resistance $<10^{-10}$). *B. subtilis* was instead serially passaged in sub-MIC levels of antibiotic for 25 days. Resistance was slow and difficult to develop, reaching a maximum MIC increase of only 4-fold for both corbomycin and complestatin (final MIC = 4 μg/mL, Fig. 3a). We observed similar low levels of

resistance development in serially passaged *S. aureus* (Extended data fig. 5a). Resistant mutants raised separately on corbomycin (strain COR14 and COR25) or complestatin (strain COM20 and COM25) displayed cross-resistance with each other but did not share cross-resistance to other cell wall or membrane active antibiotics (Table 1, Extended data fig. 5b,c), consistent with their distinct activity from these antibiotics.

Whole genome sequencing identified mutations in protein-coding regions of the *B. subtilis* resistant mutants (Fig. 3b). We identified mutations in various proteins that modify the cell wall directly (e.g. the autolysin CwlO[20]), linked to the regulation of autolysins (e.g. the phosphodiesterase YmdB[22], CwlO regulator FtsX[20,23], actin homolog Mbl[20]), or cell division in general (e.g. the metalloprotease FtsH[24], pyruvate kinase Pyk[25])(see supplemental discussion). Effects of single gene deletions were characterized (see supplemental discussion and Extended data fig. 5d), but only conferred up to 2-fold reduced susceptibility to corbomycin and complestatin (Extended data fig. 5e). This is within the accepted error for MICs and indicates a possible polygenic mechanism by which they confer resistance and a non-protein or multisubunit target for corbomycin and complestatin. Based on these resistant mutations and phenotypic evidence, we hypothesized that corbomycin and complestatin block the function of autolysins.

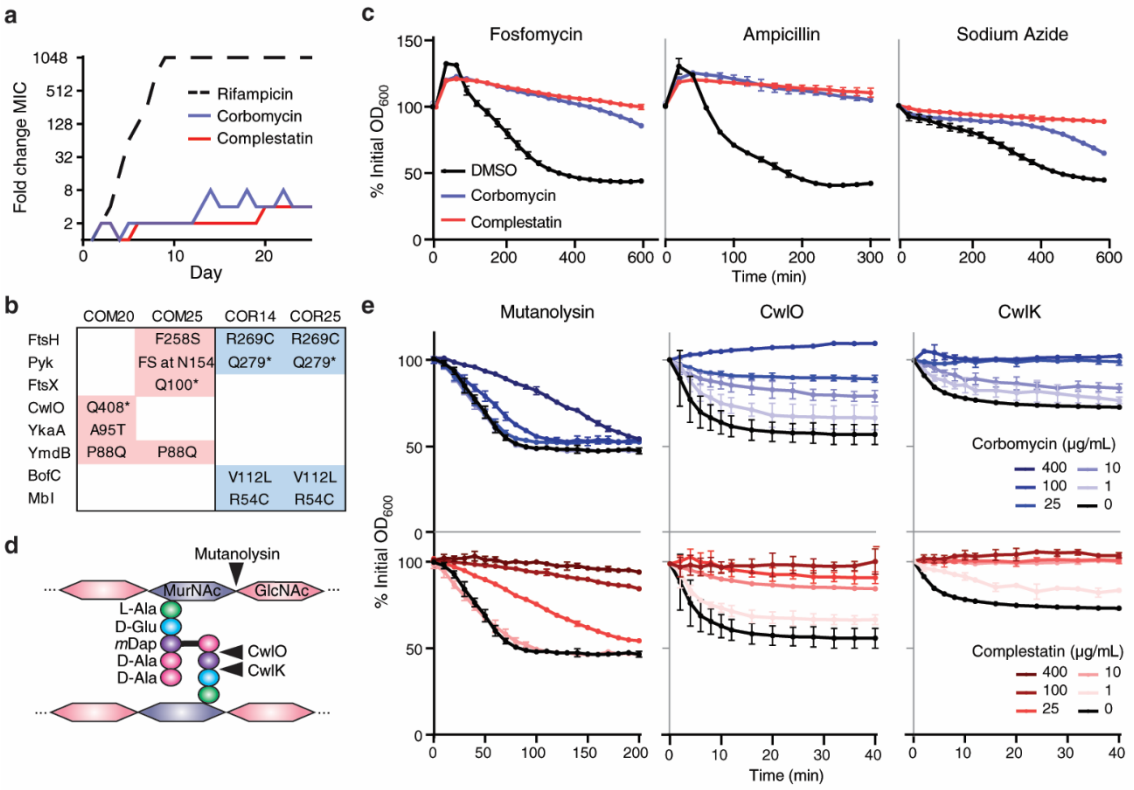


Figure 3. Corbomycin and complestatin inhibit autolysins.
(A) MICs of serially passaged *B. subtilis* in sub-lethal concentrations antibiotic over 25 days was tracked and plotted. As a comparator, rapid development of high-level resistance to rifampicin was tracked. Serial passaging was completed on two independent occasions with similar results. (B) Mutations identified in complestatin raised isolates on day 20 (COM20) or 25 (COM25), or corbomycin raised mutants on day 14 (COR14) or 25 (COR25). Amino acid mutations are shown, with FS representing a frame shift. The CwlO Q408* mutation truncates the catalytic histidine (H431), resulting in an inactive enzyme. (C) Corbomycin and complestatin protects *B. subtilis* from lysis. Cells were treated with agents to induce lysis in addition to corbomycin, complestatin or solvent control. (D) Bond cleavage specificities of various autolysins. (E) *In vitro* digestion of *B. subtilis* PG by various hydrolases was monitored after preincubation with complestatin or corbomycin. Mean values and standard deviation from biological triplicates are plotted for panels C and E.

To test whether corbomycin and complestatin inhibit autolysins in whole cells, we examined the effects of the compounds on *B. subtilis* lysis. By blocking PG synthesis with ampicillin or fosfomycin, net PG degradation by autolysins leads to lysis, but could be antagonized by corbomycin or complestatin (Fig. 3c). Some bacteriostatic drugs, especially those targeting the ribosome such as chloramphenicol, are known to antagonize bactericidal antibiotics by preventing the synthesis of lytic enzymes (Extended data fig. 6a)[26]. Therefore we induced lysis using sodium azide to dissipate the proton motive force (Fig. 3c). Autolysin activity is inhibited by low pH next to the cell membrane but becomes unregulated as this pH rises upon azide treatment[27,28]. Corbomycin and complestatin protected cells from sodium azide-induced lysis whereas chloramphenicol did not (Fig. 3c and Extended data fig. 7a). Similar results were observed for *S. aureus* lysis induced by PG synthesis inhibition (Extended data fig. 6b). These results support that corbomycin and complestatin inhibit autolysins in whole cells.

Next, we simplified our whole cell system to an *in vitro* assay containing PG and a variety of PG hydrolases with different bond specificities. Intact PG is insoluble, and so digestion can be monitored by tracking the decrease in optical density as PG is solubilized. We first tested two muramidases, lysozyme from hen egg-white and mutanolysin from *Streptomyces globisporus*. These enzymes hydrolyze β-1,4-glycosidic bonds between *N*-acetylmuramic acid (MurNAc) and *N*-acetyl-D-glucosamine (GlcNAc) of PG's glycan backbone (Fig. 3d). Complestatin strongly inhibited PG digestion by both muramidases while corbomycin blocked digestion only at higher concentrations, 400 μg/mL per 1 mg/mL PG (Fig. 3e and Extended data fig. 6c). Next, we purified and tested

two *B. subtilis* endopeptidases, CwlK (YcdD) and CwlO (YcvE), that cleave PG's peptide stem between either L-Ala and D-Glu or D-Glu and *m*DAP, respectively[29,30] (Fig. 3d). We chose these enzymes as representative physiologically relevant autolysins active during vegetative *B. subtilis* growth but targeting different bonds with unrelated catalytic domains. The action of both enzymes on PG was strongly inhibited to similar degrees by corbomycin and complestatin at concentrations as low as 10 μg/mL per 1 mg/mL PG (Fig. 3e). Corbomycin and complestatin therefore broadly block PG hydrolase activity, irrespective of enzyme family.

Measuring PG incorporation using FDAA after treatment of *B. subtilis* with 0.6xMIC corbomycin resulted in significant FDAA signal increase in contrast to 2xMIC treated cells (Fig. 2d and Extended data fig. 3b). A similar increase in nascent PG labeling was observed with fluorescent Vancomycin-BODIPY FL (Extended data fig. 6d) and has been documented for multi-autolysin *B. subtilis* mutants[31]. This phenotype is consistent with slowing of PG hydrolysis without affecting PG synthesis resulting in accumulation of labeled PG. At 2xMIC when growth is halted completely (Extended data fig. 3b), PG synthesis may slow in balance with degradation and result in no significant change in HADA incorporation (Fig. 3c and Extended data fig. 3b). Combined with results above, complestatin and corbomycin therefore inhibit autolysins both *in vitro* and in whole cells.

Corbomycin and complestatin bind PG

Corbomycin and complestatin's ability to inhibit a broad range of structurally unrelated autolysins points toward binding to the common substrate, PG. The inability to digest PG preincubated with antibiotic then washed before treatment with CwlO further

supported inhibition resulting from an interaction between antibiotic and PG, not enzyme (Extended data fig. 7a). We tested binding by incubating the antibiotics with insoluble PG at a range of weight/weight ratios, collecting insoluble material including any bound antibiotic, and quantifying the remaining soluble antibiotic by HPLC. Validating our assay, vancomycin was shown to bind PG while negative controls daptomycin and rifampin did not (Fig. 4a and Extended data fig. 7b). Corbomycin and complestatin were bound and removed from solution by both *B. subtilis* and *S. aureus* PG in a dose-dependent manner (Fig. 4a).

To further visualize PG binding, we synthesized fluorescently labeled corbomycin and complestatin by derivatizing the carboxy-terminus with BODIPY-FL-EDA (corb-BODIPY, comp-BODIPY; Extended data fig. 7c). These derivatives were shown to retain on-target activity (see methods, Extended data fig. 7d). We stained *B. subtilis* cells with 2xMIC antibiotic and observed staining of the outside of the cell including the division septa, consistent with their binding to PG (Fig. 4b,c). Generating protoplasts by removing the cell wall with lysozyme abolished staining by corb- or comp-BODIPY (Fig. 4b,c). Collectively, these results show a specific interaction between PG and corbomycin and complestatin at a motif that is widely found in PG from different species including *Staphylococcus* and *Bacillus*.

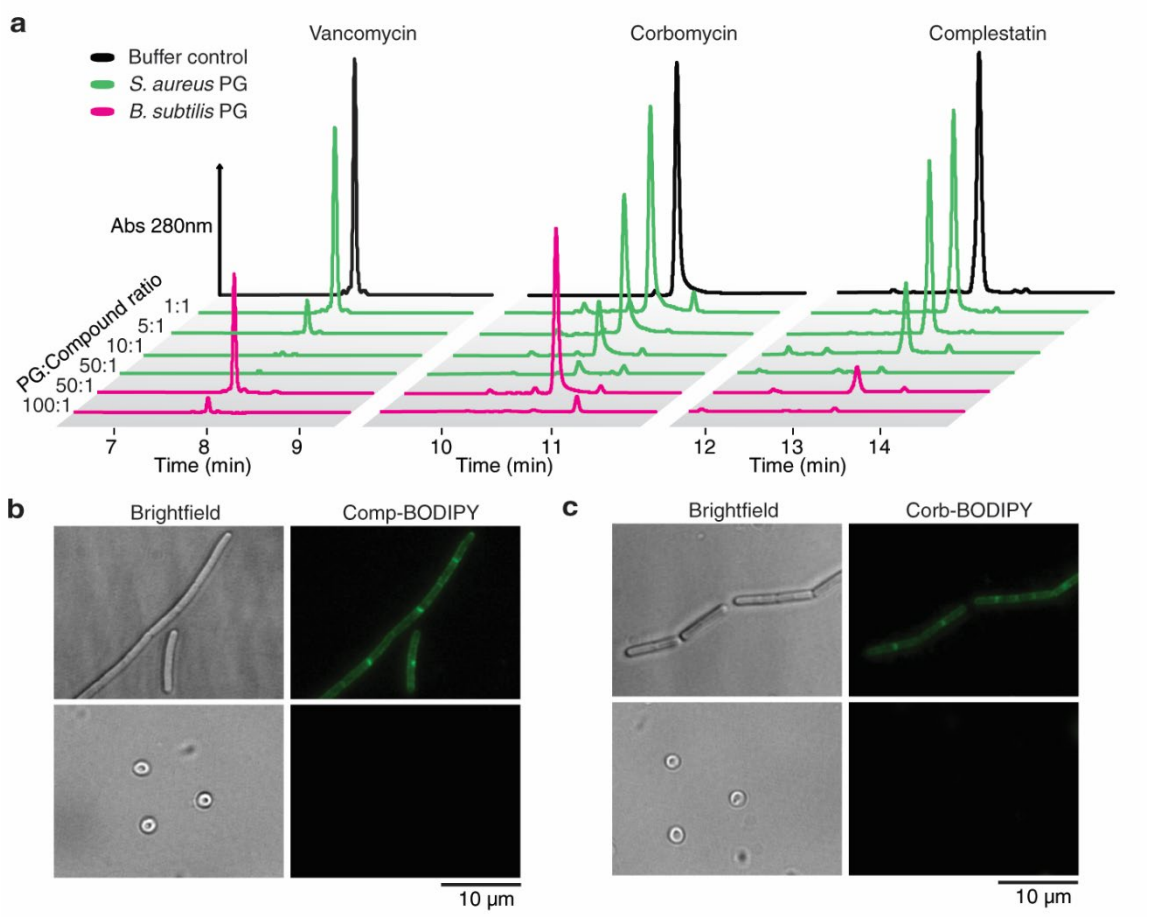


Figure 4. Corbomycin and complestatin bind to PG.
(A) HPLC chromatograms of antibiotic left unbound after incubation with either *B. subtilis* or *S. aureus* PG. Antibiotic and PG were combined in various w/w ratios, and the shortening of a peak in comparison to the buffer control that lacked PG represents compound binding and removal from solution. (B-C) Imaging of *B. subtilis* rods (top) or protoplasts (bottom) stained with fluorescent complestatin (comp-BODIPY; B) or corbomycin (corb-BODIPY; C). Identical staining conditions, exposure time and image adjustments were used for rod and protoplast images. All experiments (A-C) were performed on two separate occasions with similar results.

Our results support an unprecedented mechanism of action whereby corbomycin and complestatin bind to PG and block its access to autolysins. While autolysins are a highly redundant family of enzymes and no single enzyme is usually essential, by binding the PG itself corbomycin and complestatin block most if not all autolysin activity that is required for cell wall expansion, thereby inhibiting growth.

Murine *in vivo* efficacy

We investigated the *in vivo* efficacy of corbomycin and complestatin at treating an infection. Notably, both compounds are non-toxic to eukaryotic cells including yeasts and HEK cells (Extended data table 1). Due to poor water solubility of the compounds, we chose to formulate the antibiotics as a topical lotion for preliminary studies. We established a methicillin-resistant *Staphylococcus aureus* (MRSA; Rosenbach ATCC 33591) superficial skin infection in neutropenic mice and applied complestatin or corbomycin in a petroleum jelly-based lotion. Fusidic acid, a common topically applied antibiotic, was used as a comparison. Both complestatin and corbomycin significantly reduced bacterial load by ~100 fold at 33 h post-infection with similar efficacy to fusidic acid (Extended data fig. 8a) and in line with previous studies[32–34]. Mouse weight loss during infection was significantly reduced and wound scabbing and necrosis were noticeably improved compared to vehicle controls (Extended data fig. 8b,c).

DISCUSSION

While actinomycetes harbor a wealth of potential chemical diversity, prioritization of BGCs for follow up remains a bottleneck. By combining phylogenetic divergence with the absence of dedicated self-resistance determinants to prioritize BGCs, we describe the discovery of a new functional class of GPAs with a novel mechanism of action. This approach is similar to those employing phylogenetic analysis of sequence tags PCR amplified from metagenomic samples[3,4,35], but incorporates information from multiple genes in an intact BGC rather than single sequence tags. Tracking the phylogenetic history of complete GPA BGCs allows identification of divergent BGC since not all genes within a cluster follow the same evolutionary lineage. This strategy also provided information about grouping and location of domains in NRPS as well as important structural predictions for purifying corbomycin. Importantly, in contrast to previous methods we also focused on BGCs expressly lacking a certain resistance mechanism. Though biological activity of identified compounds cannot be predicted *a priori,* this prioritizes compounds that are more likely to have new modes of action. The selection of BGCs lacking known resistance genes is generalizable to any antibiotic class with a specific resistant mechanism (e.g. target modification).

Antibiotics blocking nearly every step in PG synthesis have been described, but complestatin and corbomycin are the first to inhibit PG remodeling. Different members of the family may have slightly different PG binding sites, which could explain subtle differences in the activity between corbomycin and complestatin. By acting as chemical

probes for PG binding and autolysin inhibition, corbomycin and complestatin along with their fluorescent derivatives will be useful tools in studying PG hydrolases.

To avoid cross-resistance with existing antibiotics, development of compounds with novel targets is a coveted but elusive goal. While others have explored the idea of inhibiting autolysins in the context of virulence[36,37] and β-lactam potentiation[38], the lack of an essential single protein prevented autolysins from emerging as a primary antibiotic target. The finding that corbomycin and complestatin inhibit most or all autolysins by binding PG, therefore, represents the discovery of antibiotics with a long-sought-after mechanism of action. Corbomycin and complestatin are active against multidrug-resistant clinical isolates, display low resistance development, and are effective *in vivo* in a murine skin infection model, making them an exciting avenue for future development.

DATA AVAILABILITY

Phylogenetic trees including those of 84 GPA genes and gene segments from 71 BGCs that were analyzed in this study are available at http://github.com/waglecn/GPA_evolution. *Streptomyces* sp. WAC01529 and *Streptomyces* sp. WAC01325 genome sequences are available in GenBank with accession numbers NZ_CP029617.1 and NZ_QHKK00000000.1, respectively. Source data for animal experiments are included online.

ACKNOWLEDGEMENTS

We thank Shawn French for help imaging Vancomycin-BODIPY stained cells. We thank Vishwas Rao and Venkateswarlu Yarlagadda for *N. gonorrhoeae* MIC measurement. This research was funded by a Canadian Institutes of Health Research grant (FRN-148463), the Ontario Research Fund, and by a Canada Research Chair to G.D.W. It was also supported by National Institute of Health grants R35GM122556 to Y.V.B. and 5R01GM113172 to M.V. and Y.V.B., and a Canada 150 Research Chair in Bacterial Cell Biology to Y.V.B. E.J.C. was supported by a CIHR Vanier Canada Graduate Scholarship. N.W. was supported by a CIHR Canada Graduate Scholarship Doctoral Award.

AUTHOR CONTRIBUTIONS

E.J.C., N.W. and G.D.W. conceived the study and designed experiments. N.W. performed phylogenetic analysis, structural predictions and resistant mutant genome analysis. W.W. performed compound purification and structural elucidation. E.J.C., A.F.C. and B.K.C. designed animal studies and A.F.C. performed animal experiments. E.J.C. and D.S. performed PG and enzyme purification. E.J.C. and K.K. performed BODIPY derivative synthesis. Y.H. and Y.V.B. designed FDAA studies, M.V. provided FDAAs and Y.H. performed experiments. E.J.C. performed all other experiments. E.J.C. and G.D.W. prepared the manuscript.

COMPETING INTERESTS

E.J.C, N.W, W.W and G.D.W. are inventors on a provisional patent application that covers the use of complestatin and corbomycin as antimicrobial therapies.

METHODS

Strains and culture conditions

All strains, sourced from this study and previous studies[17,39–41], are listed in Extended data table 2. Strains were grown under standard culturing conditions at 37°C in either Mueller Hinton Broth (MHB; BD Biosciences) or Lennox Broth (LB; Bioshop). Antibiotics were supplemented as necessary for *B. subtilis* strains (7.5 µg/mL kanamycin for BKK strains, 1 µg/mL erythromycin and 12 µg/mL lincomycin for BKE strains, 10 µg/mL chloramphenicol for pSWEET). Antibiotics were purchased from Sigma except for moenomycin (Cayman chemicals), ampicillin and kanamycin (Bioshop).

Cluster identification

We sourced GPA BGC clusters consisting of previously published sequences (31 BGCs), and used GPA fingerprint sequences[6] and BLASTp followed by antiSMASH[42] to identify new BGCs from Genbank (18 BGCs) and our in-house collection (22 BGCs). Any sequence with a hit (e-value < E-05) was subsequently run through antiSMASH (version 4.2, options "--smcogs --knownclusterblast --full-hmmer"), and looking through the results for a GPA or GPA-like cluster which we defined as an NRPS cluster with 4-hydroxyphenylglycine and/or 3,5-dihydroxyphenylglycine biosynthesis, and optionally tailoring enzymes (halogenase, glycosyltransferase, acylation, methyltransferase, etc). This generated a list of 71 whole or partial clusters from whole or partial genome sequences, and included known published and putative unknown BGCs. A more complete analysis of these clusters is found in Waglechner et al. 2019[9].

Tree construction

The NRPS detection rules in antiSMASH identifies well-known functional domains in the large multimodular synthase genes in order to facilitate structural predictions of the products of the NRPS enzymes. These domains include the peptidyl-carrier protein or thiolation domains, adenylation domains that activate precursor amino acids before they can be incorporated into the peptide scaffold, epimerization domains that catalyze amino acid conversion from D to L epimers, in GPA BGCs the so-called X-domains that facilitate oxidative crosslinking[43], and condensation domains that catalyze peptide bond formation. The sequences identified as C-domains were aligned using MUSCLE (default parameters), and manually inspected for the presence of gaps and mis-aligned regions. This alignment was used to compute a maximum-likelihood tree using fasttree2, using the WAG substitution model and CAT approximation with the default number of rate categories. Each block of condensation domains labelled in the phylogeny for the numbered positions in GPA scaffolds are robust, with SH-like support values > 0.949 except for one, 0.891 for C+2 corbomycin. Further bioinformatic methods are available in Waglechner et al. 2019[9].

Due to variation in the number of NRPS-encoded condensation domains in each BGC, we applied a labeling system for condensation domains based on location with respect to the shared centrally encoded 4-hydroxyphenylglycine, position 0, to allow us to visualize trends within this tree (Extended data fig. 1a). Using this labeling system, the relationship between condensation domains of different modules and between strains

within an individual module can be observed (Extended data fig. 1b). The location C-2 in this domain structure labeling scheme is shown as an example in Figure 1a.

Antibiotic susceptibility testing

Antibiotic susceptibility was tested using standard procedures in MHB for most strains. *Enterococcus* strains were grown in Brain Heart Infusion media (BHI; BD Biosciences). *Streptomyces* MICs were determined in Tryptic Soy Broth (TSB; BD Biosciences) with 0.5% yeast extract. *N. gonorrhoeae* inoculum for MIC determination was prepared from strains streaked on enriched chocolate agar and resuspended in GW liquid media[44], and microtitre plates were incubated shaking at 500 rpm, 37°C, 5% CO_2 and 90% humidity for 12 hr. MIC determination of daptomycin and bacitracin were supplemented with 1.25 mM $CaCl_2$ or 40 μg/mL $ZnCl_2$, respectively. 2D checkerboards were set up according to standard protocol.

Fermentation and purification of corbomycin

Structural predictions for corbomycin were informed using phylogenetic trees of NRPS adenylation domains, including 3,5-dihydroxyphenylglycine specificity, something notoriously difficult to predict for this amino acid[45]. The NPRS was predicted to encode a nonapeptide, in contrast to most GPA antibiotic's heptapeptide backbone. These structural predictions guided initial purification of corbomycin independent of antibiotic activity.

WAC1529 mycelium from 50 mL TSB seed culture was inoculated into each of eighteen 2.8 L flasks containing 600 mL Bennett's media (1% potato starch, 0.2% casamino acids, 0.18% yeast extract, 0.02% KCl, 0.02% $MgSO_4 \cdot 7H_2O$, 0.024% $NaNO_3$,

$4x10^{-4}$ % $FeSO_4 \cdot 7H_2O$). After 4 days, fermentations were fed with 0.2 mM each cysteine, histidine, glutamine and tyrosine. Amino acid supplements were prepared as 100x stocks in water (Cys, His, Glu) or 10 mM HCl (Tyr) and neutralized with two equivalents of $NaHCO_3$ after addition to fermentations.

Spent media was extracted with 8% (W/V) HP-20 (Diaion) resin. Cell pellets were extracted twice with 500 mL methanol (MeOH) and was concentrated under vacuum with 100 g HP-20 (Diaion) resin. These resins were combined and eluted with H_2O (2 L), 20% MeOH (2 L), 40% MeOH (2 L), and 100% MeOH (4 L). Analysis of fractions by HPLC and liquid chromatography-mass spectrometry (LC-MS) identified a peak with molecular weight (~1300-1600 Da) and UV-profile maxima (220 nm, 280 nm) consistent with the predicted structure. This fraction (100% MeOH) was extracted with ethyl acetate, MeOH/H_2O (1:4) and DMSO. The DMSO subfraction was found to contain the predicted glycopeptide and was applied to reverse-phase CombiFlash ISCO (RediSep Rf C18, Teledyne) and eluted with a linear gradient system (5-100% water/acetonitrile, 0.1% formic acid) to give 136 fractions. Fractions containing the predicted glycopeptide were combined and subject to Sephadex LH-20 column (400 mL), eluting with MeOH/Acetonitrile/H_2O (1:2:1), to yield 36 subfractions. Identified subfractions were combined and further purified with Agilent Eclipse XDB-C8 column (5 μm, 9.4 × 250 mm), to yield 23.6 mg of corbomycin.

Fermentation and purification of complestatin

WAC1325 was fermented using conditions identical to WAC1529 except for amino acid feeding. After 3 days growth at 30°C, 250 rpm, fermentations were fed with

0.2 mM each 4-hydroxyphenylglycine, tryptophan and tyrosine. Amino acids were prepared as a 100x stock solution in 10 mM HCl and neutralized with two equivalents of $NaHCO_3$ after addition to fermentations. Growth was allowed to continue for a total of 8 days and complestatin was purified from spent media and cell pellet in a similar manner to corbomycin. High resolution-electrospray ionisation-mass spectrometry (HR-ESI-MS) and ^{1}H- nuclear magnetic resonance (NMR) confirmed the compound as complestatin.

Structural characterization of corbomycin

To elucidate its structure, corbomycin was subjected to 1D and 2D NMR and HR-ESI-MS. See Supplementary figures 1-6 and Supplementary table 1 for all structural data. 1D and 2D NMR experiments were performed using a Bruker AVIII 700 MHz instrument equipped with a cryoprobe in deuterated DMSO. Chemical shifts are reported in parts per million relative to tetramethyl silane using the residual solvent signals at 2.50 ppm in proton NMR and 39.5ppm in carbon NMR as an internal signals. HR-ESI-MS was acquired using an Agilent 1290 UPLC separation module and qTOF G6550A mass detector in negative ion mode. Interpretation of MS and NMR data for structural determination are provided in the supplemental discussion.

Fab and CwlO overexpression and knockout in B. subtilis 168

Full length *fabI, fabL, cwlO* and truncated $cwlO_{1\text{-}407}$ were cloned from *B. subtilis* 168 gDNA into pSWEET[46] for overexpression under a xylose inducible promoter[46]. Both pSWEET-*bgaB* and PCR amplicons were digested with PacI and BamHI restriction enzymes before ligation and electroporation into *E. coli* Top10 and selection with 100 µg/mL ampicillin. For transformation into *B. subtilis* 168, 1 µg of plasmid DNA was

digested with PstI and transformed as previous described using selection with 10 µg/mL chloramphenicol[41]. As *B. subtilis ΔcwlO* displays loss of genetic competence[47], *B. subtilis* Δ*cwlO* pSWEET combination strains were generated by subsequently transforming 2 µg BKE34800 gDNA into respective pSWEET containing strains and selected with appropriate antibiotics. For antibiotic susceptibility testing, overexpression of genes was induced using 3% w/v xylose in LB media.

P_{ywaC} luminescence testing

Activation of P_{ywaC} was tested in *B. subtilis* EB1385 as previously described with minor modifications[17]. EB1385 was grown overnight in MHB with 1 µg/mL erythromycin, then diluted to an OD_{600} of 0.15 in fresh MHB. 100 µL of inoculum was dispensed in a white 96 well plate with a clear bottom. Antibiotics were added in triplicate in a two-fold serial dilution from 16 µg/mL to 0.125 µg/mL. The plate was covered with an optically clear film, and incubated at 37°C with shaking, with OD_{600} and luminescence (0.1 sec integration time) monitored using an EnVision Multilabel plate reader (PerkinElmer) at 10 minute intervals. Luminescence is reported for the lowest concentration that fully inhibited growth at this inoculum (2 µg/mL vancomycin, 4 µg/mL corbomycin, 8 µg/mL complestatin, 1 µg/mL triclosan).

$DiSC_3$ dye fluorescence assay

Membrane permeability was tested using the voltage sensitive dye $DiSC_3$. Midlog phase *B. subtilis* 168 grown in LB (OD_{600} ~0.6) were washed twice in 5 mM HEPES (pH 7.3), and finally resuspended in 5 mM HEPES with 20 mM glucose (pH 7.3), adjusting the OD_{600} to 0.1. 200 µL of the cell suspension was dispensed in a black 96 well plate and

$DiSC_3$ dye in DMSO was added to a final concentration of 1 µM. Cells were incubated at room temperature for 30 minutes with shaking to allow for dye uptake, and fluorescence quenching was monitored using a Synergy H1 microplate reader (excitation/emission 600/660 nm). Once fluorescence stabilized, test compounds were added to a final concentration of 10-12 µM (18 µg/mL gramicidin, 11 µg/mL valinomycin, 16 µg/mL vancomycin, corbomycin and complestatin) and fluorescence was monitored. Maximal fluorescence, observed 250 sec after compound addition for all antibiotics, is reported. The experiment was performed in triplicate on two independent occasions.

UDP-MurNAc-PP accumulation in S. aureus

UDP-MurNAc-PP accumulation was measured in *S. aureus* as previously described[48]. Briefly, midlog phase *S. aureus* ATCC 29213 pretreated with 130 µg/mL chloramphenicol were split into 10 mL aliquots and exposed to 10x MIC test antibiotic (10 µg/mL corbomycin, 20 µg/mL complestatin, 1.25 µg/mL ampicillin, 10 µg/mL vancomycin, 1.25 moenomycin, 160 µg/mL bacitracin zinc, 20 µg/mL fosfomycin, 160 µg/mL carbenicillin, 80 µg/mL mecillinam, 640 µg/mL aztreonam). After 1 hr incubation at 37°C, 250 rpm, cells were pelleted and extracted with boiling water. Soluble extracts were lyophilized, then resuspended in 100 µL water and 20 µL was analyzed on an Inertsil ODS-4 4x150 mm column run with isocratic elution (0.5 mL/min, 50 mM sodium phosphate buffer, pH 5.2) at 37°C. The results reported are representative of two independent experiments.

Bacteriostatic cfu determination

Midlog phase *B. subtilis* 168 (OD_{600} = 0.25) in LB was dispensed into 1.5 mL aliquots and treated with 4xMIC (4 μg/mL) or 8xMIC (8 μg/mL) complestatin, corbomycin, chloramphenicol, or left untreated. Cells were incubated at 37°C for 6 hr, then cells were washed twice with LB to remove antibiotic that could inhibit growth, OD_{600} was adjusted to 0.4, and cfu were enumerated in triplicate. For growth curves, these cells were inoculated 1:200 into fresh LB and dispensed into a microtitre plate. The plate was covered with a clear, breathable film and OD_{600} was monitored on a Tecan sunrise microplate reader at 37°C with shaking.

Single step selection of spontaneous resistant mutants

Single step selection of resistant mutants in *B. subtilis* 168 and *S. aureus* ATCC29213 on solid media was performed according to standard protocol. Mueller Hinton media with 1% agarose and 4x or 8x solid MIC complestatin or corbomycin (32-64 μg/mL) were plated with $10^9 – 10^{10}$ cfu from an overnight culture. Cfu were enumerated directly from this overnight culture. Plates were incubated for 48 hr at 37°C and no resistant colonies were detected. Experiment was performed on two independent occasions.

Raising resistance mutants through serially passaging

Complestatin and corbomycin resistant mutants were raised beginning with the laboratory strain *B. subtilis* 168 or *S. aureus* ATCC 29213. To begin, a single colony was inoculated into 1 mL MHB in a sterile test tube with 0.25xMIC, 0.5xMIC, 1xMIC and 2xMIC where MIC = 1 μg/mL for both complestatin and corbomycin, and 0.0625 μg/mL

for rifampicin. After 24 hr growth with shaking, the lowest concentration with no growth was taken as the new MIC, and cells were subcultured into fresh tubes 1 in 100 from the highest concentration that supported growth. This process was continued for 25 days, and glycerol stocks were taken whenever there was a shift in MIC. At the end of 25 days, glycerol stocks were streaked on non-selective media (Mueller Hinton agar) and single colonies were isolated for two generations. The MIC of purified strains was measured by microbroth dilution. Serial passaging was performed in biological duplicate using two independent lines.

For one line of the serially passaged cells, whole genome sequencing was performed on the strain on the final day (COM25 and COR25), or at the earliest time point that the highest MIC was reached. For corbomycin, this strain arose at day 14 (*B. subtilis* COR14), and for complestatin, day 20 (*B. subtilis* COM20). Whole genome sequencing on resistant mutants, as well as our laboratory *B. subtilis* 168, was performed with Illumina MiSeq (300 bp, paired end reads) by the Farncombe Genomics Facility. To identify mutations unique to our evolved mutants versus wildtype, each of the three sequenced strains were compared to the published *B. subtilis* 168 reference genome (accession number AL009126.3) using breseq (version 0.33.1)[49] to generate a list of differences. Changes in protein coding regions that were unique to resistant mutants and not present in our laboratory *B. subtilis* 168 strain were identified for follow up. Sequencing two individual colonies isolated from day 25 gave identical genotypes.

Cell lysis assay

Early exponential phase *B. subtilis* 168 or *S. aureus* ATCC29213 (OD_{600} ~0.25) grown in LB media was dispensed 100 µL per well into a round bottom 96 well plate. Antibiotics or $MgSO_4$ were supplemented to the appropriate concentration, and where appropriate, lytic agents were added (50 µg/mL fosfomycin, 100 µg/mL ampicillin, 75 mM sodium azide). Excess Mg^{2+} inhibits azide induced lysis by a mechanism not well understood but thought to be partially through modulating autolysin activity[50,51]. The plate was covered with a clear, breathable film and OD_{600} was monitored on a Tecan sunrise microplate reader for 10 hr at 37°C with shaking.

Autolysin overexpression and purification from E. coli

CwlK and the catalytic CwlO domain with signal peptides removed were cloned into pET28a (EMD Biosciences) using primers CwlK_Full-F (GTCACCCATGGGCCATGAATGGCATTCTCAAAA) and CwlK_Full-R (GTCACCTCGAGGTTAGGAATCATCTCCAAGTG), or CwlO_Cat-F (GTCACCCATGGGCACTGTTATCAGCAACTCTGG) and CwlO-R (GTCACCTCGAGTTGAACAACACGTCTTACAAC), respectively, and *NcoI* and *XhoI* restriction sites for introduction of C-term 6xHis. Cloning was performed in *E. coli* Top10 followed by transformation into *E. coli* BL21(DE3) pLysS for expression. For expression, 1 L of LB media supplemented with 50 µg/mL kanamycin and 35 µg/mL chloramphenicol was inoculated 1:50 from an overnight culture and grown at 37°C, 250 rpm until OD_{600} reached 0.6. Cells were induced with 2 mM IPTG at 37°C for 3 hr (CwlO), or 18 hr at 18°C with 1 mM IPTG (CwlK). Cell pellets were resuspended in 20

mL lysis buffer (25 mM HEPES (pH 8), 300 mM NaCl, 10 mM imidazole) and frozen at -80°C. For purification, pellets were thawed, treated with Pierce protease inhibitor tablets (Thermo Scientific), 10 mg/mL lysozyme and 5 µg/mL DNase, and lysed on ice by sonication. Clarified lysates were loaded onto equilibrated 2 mL Ni-NTA agarose (Qiagen), washed (25 mM HEPES (pH 8), 300 mM NaCl, 25 mM imidazole) and eluted (25 mM HEPES (pH 8), 300 mM NaCl, 250 mM imidazole). Elutions were dialyzed overnight in buffer A for downstream anion/cation exchange. CwlO was purified by cation exchange with a HiTrap SP HP 5 mL column (GE Healthcare) on an AKTA Explorer system (Buffer A: 25 mM HEPES (pH 8); Buffer B: 25 mM HEPES (pH 8), 1 M NaCl). CwlK was similarly purified by anion exchange on a HiTrap Q HP 5 mL column (GE Healthcare) (Buffer A: 25 mM Tris-HCl (pH 8); Buffer B: 25 mM HEPES (pH 8), 1 M NaCl). Buffer exchange and concentration of fractions containing pure protein was performed using Amicon Ultra centrifugal filters with 3K cutoff. Protein purity was >95% as assessed by SDS-PAGE analysis.

Specific activity of the purified enzymes was confirmed by performing PG digestion, as described below, with inactivated enzyme. CwlK, a Zn^{2+} dependent metalloprotease[30], could be inhibited by addition of 0.5 mM EDTA, while CwlO, a NplC/P60 cysteine peptidase[29], could be inhibited by preincubation with 4 mM maleimide.

Peptidoglycan isolation and digestion

To prepare PG as a substrate for binding or digestion experiments, *B. subtilis* 168 or *S. aureus* ATCC 29213 was grown to OD_{600} 0.6-0.7 in LB media and PG was purified

as previously described with minor modification[52]. Briefly, cell pellets were boiled in 4% SDS, then washed, sonicated to break up sacculi, treated with α-amylase and DNase, and finally digested with pronase overnight. The PG was again boiled in 2% SDS, washed, and teichoic acids were hydrolyzed with 1N HCl. The pure PG was washed with water until the pH reached 6 and lyophilized.

For enzymatic digestion of PG, conditions for each enzyme were optimized as follows: mutanolysin (20 µg/mL, 1 mg/mL PG, 20 mM sodium acetate, pH 6.5), lysozyme (100 µg/mL, 1 mg/mL PG, 20 mM sodium acetate, pH 6.5), CwlO (20 µg/mL, 0.5 mg/mL PG, 20 mM MES-NaOH, pH 5.5), CwlK (16 µg/mL, 1 mg/mL PG, 10 µM $ZnSO_4$, 20 mM sodium acetate, pH 6.5). To test corbomycin and complestatin's effect on digestion, 100 µL of PG in buffer, as indicated, was preincubated with either antibiotic or a corresponding volume of DMSO with shaking for 30 minutes, then dispensed into a round bottom 96 well plate and enzyme added. The plate was covered with a clear, breathable film to prevent evaporation, and OD_{600} was tracked on a Tecan sunrise microplate reader shaking at 37°C. Results show average and standard deviation of triplicate wells.

FDAA imaging

Bacillus subtilis 3610 (wildtype) was grown in LB medium at 37 °C to exponential phase (OD_{600} 0.1-0.2), then diluted with fresh LB medium 1:10 and allowed to grow to OD_{600} = 0.2. Complestatin and corbomycin were added to exponentially growing cell cultures (0.3 mL) to a final concentration of 0.6x or 2xMIC and incubated for 5 minutes at 37°C. Alternatively, ampicillin was added at 0.5x, 1x, 2x and 5xMIC and

incubated for 1 minute at 37°C. FDAAs (100 mM DMSO solution) were added to the cultures to a final concentration of 0.5 mM and incubated at 37°C with shaking for 1 hour (complestatin and corbomycin). A shorter incubation time, 5 minutes, was used for ampicillin treated cells to prevent lysis. Cells were fixed by adding pure ethanol to the cultures to a final concentration of 70% (v/v), incubating on ice for 1 hour, and finally washing twice with PBS. Failure to label cells with the L-isomer HCC-amino-L-alanine (HALA) confirmed that incorporation was conducted by transpeptidases (Extended data fig. 3c).

For cell imaging, cells were applied to a coverslip (24x50 mm; #1.5) and covered with an 8x8-mm wide, 2-mm thick PBS-agarose pad (SeaKem LE Agarose). The coverslip-pad combination was placed onto a customized slide holder on the microscope with the pad facing upwards. Phase-contrast and fluorescence images were acquired using a Nikon Ti-E inverted microscope equipped with a 1.4NA Plan Apo 60X oil objective and Andor iXon EMCCD camera. NIS-Element AR (Version 5.02) was used for image acquisition. Identical conditions (exposure, light source power and EM gain) were applied to all the samples for valid comparisons. Filters for HADA imaging: Ex 395/25 nm (DAPI) and Em 435/26 nm.

Image processing was performed in FIJI (Version 1.51). Images were scaled without interpolation, cropped and rotated. Linear adjustment was performed to optimize contrast and brightness of the images. Quantitative measurement of FDAA labeling intensity was achieved using a FIJI plugin, MicrobeJ, where cells were identified in the

phase contrast channel with width limit from 0.3 to 2 μm and length above 1 μm. FDAA labeling intensity was then quantified and averaged so that $n > 100$ for each condition.

Peptidoglycan binding assay

To assess PG binding, 1.6 mg/mL antibiotic dissolved in DMSO was diluted to 0.1 mg/mL in 20 mM phosphate buffer (pH 7.1) with 0 mg/mL, 0.1 mg/mL, 0.5 mg/mL, 1 mg/mL, 5 mg/mL or 10 mg/mL PG for 0, 1:1, 5:1, 10:1, 50:1 and 100:1 w/w ratios, respectively. *B. subtilis* 168 and *S. aureus* ATCC 29213 PG was tested similarly. Reaction volumes were adjusted so that a minimum of 24 μg PG was used in the lowest ratio condition to allow for a visible pellet of PG to form. Mixtures were incubated for 1-2 hr at 37°C, then centrifuged for 10 min to thoroughly pellet insoluble material. 50 μL of supernatant containing unbound antibiotic (maximally 5 μg) was injected and detected by HPLC using WAT094269 Symmetry Shield RP 8 3.5 μM 4.6 x 150 mm column and the following gradient at 0.8 mL/min, 40°C: 0-2 min 0% B, 2-12 min 0-95% B, 12-18 min 95% B, 18-20 min 95-0% B (Buffer A: H_2O + 0.1% formic acid, Buffer B: acetonitrile + 0.1% formic acid)

Antibiotic BODIPY semisynthesis

Vancomycin-BODIPY FL was synthesized as described previously[53]. Complestatin and corbomycin were derivatized by combining equimolar amounts of antibiotic, BODIPY FL ethylenediamine (Invitrogen), N,N-diisopropylethylamine and coupling reagents. For complestatin, 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) plus ethyl cyanohydroxyiminoacetate (oxyma) was used, and for corbomycin N,N'-diisopropylcarbodiimide (DIC) plus oxyma

was used. All components were dissolved in dimethylformamide (DMF) except for corbomycin and complestatin in DMSO. The reaction was carried out at room temperature for 24-48 hr then purified by semi-preparative HPLC using an XSelect CSH Prep C18 5 µm 10x100 mm column with a gradient of H_2O + 0.1% formic acid and acetonitrile + 0.1% formic acid. The derivatives were verified by HR-ESI-MS using an Agilent 1290 Infinity II Series HPLC system and 6550 iFunnel QTOF LC-ESI/MS as follows: Corbomycin-BODIPY $[M+2H]^{+2}$ calc = 912.3092, obs = 912.3094; Complestatin-BODIPY $[M+H]^{+}$ calc = 1642.2856, obs = 1642.2832.

Comp- and Corb-BODIPY had MICs against *B. subtilis* 168 of 16 µg/mL, and showed a similar 4 fold MIC increase against their respective resistant mutants (*B. subtilis* COM20 and COR14 = 64 µg/mL). Growth at sub-MIC levels produced twisted chains of cells (Extended data fig. 7d) similar to underivatized antibiotics. Therefore these derivatives have the same mechanism as the parent compounds.

Fluorescence microscopy

To label active peptidoglycan synthesis in antibiotic-treated cells, midlog phase *B. subtilis* 168 (OD_{600} ~0.5) grown in LB media were first pretreated with 10xMIC for 30 min at 37°C (10 µg/mL corbomycin, 20 µg/mL complestatin, 0.039 µg/mL ampicillin, 640 µg/mL aztreonam). Van-BODIPY was then added along with unlabelled vancomycin each at 1 µg/mL, as previously reported[53]. Cells were incubated at room temperature for 15 minutes with shaking, then washed three times with 20 mM HEPES buffer (pH 8) to remove unbound compound.

For staining with comp- and corb-BODIPY, *B. subtilis* 168 was similarly grown to midlog phase in LB, then harvested by centrifugation and resuspended in 20 mM HEPES buffer (pH 8) for rod staining, or MSM buffer (20 mM $MgCl_2$, 0.5 M sucrose, 20 mM maleic acid, pH 7) for protoplast generation. For protoplasts, cells in osmotically protective MSM buffer were treated with 4 mg/mL lysozyme for 1 h at 37°C, then centrifuged and resuspended in fresh MSM buffer prior to staining. Rods or protoplasts were stained with 2xMIC comp- or corbomycin-BODIPY (32 μg/mL) for 15 minutes with shaking, then washed four times in their respective buffers to remove unbound compound.

Prepared cells were mounted on polylysine treated glass and imaged using a Nikon Eclipse Ti inverted microscope with a FITC fluorescence filter for BODIPY (excitation/emission 470/520nm). Equivalent exposure time and image adjustments were used for rod and protoplast stained cells. FM 4-64FX (Invitrogen) was used where indicated with the corresponding fluorescence filter.

Transmission Electron Microscopy

To prepare cells for imaging, *B. subtilis* 168 were grown in LB media with 0.6 μg/mL corbomycin or 1 μg/mL complestatin to OD_{600} 0.5-0.7, then harvested by centrifugation, resuspended in 2% glutaraldehyde (2% v/v) in 0.1M phosphate buffer (pH 7.4) and allowed to fix overnight at 4°C. The samples were prepared for TEM as previously described[54] and viewed in a JEOL JEM 1200 EX TEMSCAN transmission electron microscope (JEOL, Peabody, MA, USA) operating at an accelerating voltage of

80kV. Images were acquired with an AMT 4-megapixel digital camera (Advanced Microscopy Techniques, Woburn, MA).

Murine skin infection model

All mouse experiments were performed in the Central Animal Facility at McMaster University under animal use protocol #17-03-10 as approved by the Animal Research Ethics Board. We have complied with all relevant ethical regulations. Six- to ten-week-old female BALB/c mice (Charles River, Cat.#028) were used for all experiments. The wounded area was prepared as previously described[34]. Briefly, anesthesia was induced with 5 % isoflurane and maintained at 2.5 %. An area of ~ 2 cm^2 on the dorsal region of the neck was stripped using 25 autoclave tape strips (3M) until the skin was visibly shiny but not bleeding. For bacterial preparation, *S. aureus* ATCC 33591 was grown overnight in TSB media, and sub-cultured to OD_{600} ~0.4. Bacteria were adjusted to $5x10^6$ cfu/ml and a 20 µl volume (10^5 cfu) was applied to the wounded area. Mice in groups of two, three or four were randomly relocated from housing cages into singly housed treatment and control cages. Test antibiotics were formulated in petroleum jelly (Vaseline) at 1% (w/w) and animals were treated 8 times over a 33 h period at the following times post infection: 1, 4, 8, 12, 20, 24, 28, 32 h. Each treatment was 30 mg Vaseline formulation by weight, except for the treatment at 12 h which was 50 mg. Fusidic acid dose was chosen to give similar molar concentrations to complestatin and corbomycin and to mimic their bacteriostatic effect, as fusidic acid is mainly bacteriostatic but becomes bacteriocidal at high concentrations[55]. Control mice received 10 % DMSO in vaseline. Mice were euthanized at 33 h post infection. At endpoint, a ~ 2

cm^2 area of skin was excised and homogenized in 1 ml PBS for 15 min at 30 rps (Retsch MM400). Homogenates were serially diluted and plated on Tryptic Soy Agar containing oxacillin (2 μg/ml) for cfu determination. Six animals (technical replicates) were used in each treatment group, divided among two experiments performed on independent occasions (biological replicates). Nine vehicle control mice were used with at least two control mice in each batch of biological replicates performed.

Statistical analysis

Statistical analysis of the murine skin infection model was performed using GraphPad V6. Significance was tested by one-way ANOVA on ranks with Kruskal-Wallis test and Dunn's *post hoc* analysis for multiple comparisons ($n = 6$ for complestatin, corbomycin and fusidic acid, $n = 9$ for vehicle). Mean rank are as follows: vehicle 27.78, 1% complestatin 10.33, 1% corbomycin 11.67, 0.25% fusidic acid 6.5.

REFERENCES

1. Laxminarayan, R. *et al.* Antibiotic resistance-the need for global solutions. *Lancet. Infect. Dis.* **13**, 1057–98 (2013).

2. Wright, G. D. Something old, something new: revisiting natural products in antibiotic drug discovery. *Can. J. Microbiol.* **60**, 147–54 (2014).

3. Kang, H. S. & Brady, S. F. Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. *ACS Chem. Biol.* **9**, 1267–1272 (2014).

4. Peek, J. *et al.* Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat. Commun.* **9**, 4147 (2018).

5. Hover, B. M. *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol.* **3**, 415–422 (2018).

6. Thaker, M. N. *et al.* Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat. Biotechnol.* **31**, 922–927 (2013).

7. Yan, Y. *et al.* Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action. *Nature* **559**, 415–418 (2018).

8. Tang, X. *et al.* Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10**, 2841–2849 (2015).

9. Waglechner, N., McArthur, A. G. & Wright, G. D. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat. Microbiol.* **4**, 1862–1871 (2019).

10. Chiu, H.-T. *et al.* Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc. Natl. Acad. Sci.* **98**, 8548–8553 (2001).

11. Nicolaou, K. C., Boddy, C. N. C., Bräse, S. & Winssinger, N. Chemistry, biology, and medicine of the glycopeptide antibiotics. *Angew. Chemie Int. Ed.* **38**, 2096–2152 (1999).

12. Breazzano, S. P. & Boger, D. L. Synthesis and stereochemical determination of complestatin A and B (neuroprotectin A and B). *J. Am. Chem. Soc.* **133**, 18495–18502 (2011).

13. Nazari, B. *et al. Nonomuraea sp.* ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. *Medchemcomm*

8, 780–788 (2017).

14. Naruse, N. *et al.* New antiviral antibiotics, kistamicins A and B. I. Taxonomy, production, isolation, physico-chemical properties and biological activities. *J. Antibiot. (Tokyo).* **46**, 1804–11 (1993).

15. Kwon, Y., Kim, H. & Kim, W. Complestatin exerts antibacterial activity by the inhibition of fatty acid synthesis. *Biol. Pharm. Bull.* **38**, 715–721 (2015).

16. Park, O. K., Choi, H. Y., Kim, G. W. & Kim, W. G. Generation of new complestatin analogues by heterologous expression of the complestatin biosynthetic gene cluster from *Streptomyces chartreusis* AN1542. *ChemBioChem* **17**, 1725–1731 (2016).

17. Czarny, T. L., Perri, A. L., French, S. & Brown, E. D. Discovery of novel cell wall-active compounds using Pywac, a sensitive reporter of cell wall stress, in the model gram-positive bacterium *Bacillus subtilis*. *Antimicrob. Agents Chemother.* **58**, 3261–3269 (2014).

18. D'Elia, M. A., Millar, K. E., Beveridge, T. J. & Brown, E. D. Wall teichoic acid polymers are dispensable for cell viability in *Bacillus subtilis*. *J. Bacteriol.* **188**, 8313–8316 (2006).

19. Kuru, E., Tekkam, S., Hall, E., Brun, Y. V & Van Nieuwenhze, M. S. Synthesis of fluorescent D-amino acids and their use for probing peptidoglycan synthesis and bacterial growth in situ. *Nat. Protoc.* **10**, 33–52 (2015).

20. Domínguez-Cuevas, P., Porcelli, I., Daniel, R. A. & Errington, J. Differentiated roles for MreB-actin isologues and autolytic enzymes in *Bacillus subtilis* morphogenesis. *Mol. Microbiol.* **89**, 1084–1098 (2013).

21. Blackman, S. A., Smith, T. J. & Foster, S. J. The role of autolysins during vegetative growth of *Bacillus subtilis* 168. *Microbiology* **144**, 73–82 (1998).

22. Diethmaier, C. *et al.* A novel factor controlling bistability in *Bacillus subtilis*: the YmdB protein affects flagellin expression and biofilm formation. *J. Bacteriol.* **193**, 5997–6007 (2011).

23. Meisner, J. *et al.* FtsEX is required for CwlO peptidoglycan hydrolase activity during cell wall elongation in *Bacillus subtilis*. *Mol. Microbiol.* **89**, 1069–83 (2013).

24. Yepes, A. *et al.* The biofilm formation defect of a *Bacillus subtilis* flotillin-defective mutant involves the protease FtsH. *Mol. Microbiol.* **86**, 457–471 (2012).

25. Monahan, L. G., Hajduk, I. V., Blaber, S. P., Charles, I. G. & Harry, E. J.

Coordinating bacterial cell division with nutrient availability: a role for glycolysis. *MBio* **5**, e00935-14 (2014).

26. Kudrin, P. *et al.* Subinhibitory concentrations of bacteriostatic antibiotics induce relA-dependent and relA -independent tolerance to β-lactams. *Antimicrob. Agents Chemother.* **61**, 1–17 (2017).

27. Jolliffe, L. K., Doyle, R. J. & Streips, U. N. The energized membrane and cellular autolysis in *Bacillus subtilis*. *Cell* **25**, 753–763 (1981).

28. Calamita, H. G., Ehringer, W. D., Koch, A. L. & Doyle, R. J. Evidence that the cell wall of *Bacillus subtilis* is protonated during respiration. *Proc. Natl. Acad. Sci.* **98**, 15260–15263 (2001).

29. Yamaguchi, H., Furuhata, K., Fukushima, T., Yamamoto, H. & Sekiguchi, J. Characterization of a new Bacillus subtilis peptidoglycan hydrolase gene, *yvcE* (named *cwlO*), and the enzymatic properties of its encoded protein. *J. Biosci. Bioeng.* **98**, 174–181 (2004).

30. Fukushima, T., Yao, Y., Kitajima, T., Yamamoto, H. & Sekiguchi, J. Characterization of new l,d-endopeptidase gene product CwlK (previous YcdD) that hydrolyzes peptidoglycan in *Bacillus subtilis*. *Mol. Genet. Genomics* **278**, 371–383 (2007).

31. Davies, A. Peptidoglycan architecture and dynamics in Bacillus subtilis. (University of Sheffield, 2014).

32. Pletzer, D., Mansour, S. C., Wuerth, K., Rahanjam, N. & Hancock, R. E. W. New Mouse Model for Chronic Infections by Gram-Negative Bacteria Enabling the Study of Anti-Infective Efficacy and Host-Microbe Interactions. *MBio* **8**, (2017).

33. Chiang, N. *et al.* Infection regulates pro-resolving mediators that lower antibiotic requirements. *Nature* **484**, 524–8 (2012).

34. Kugelberg, E. *et al.* Establishment of a superficial skin infection model in mice by using *Staphylococcus aureus* and *Streptococcus pyogenes*. *Antimicrob. Agents Chemother.* **49**, 3435–3441 (2005).

35. Guo, J., Ran, H., Zeng, J., Liu, D. & Xin, Z. Tafuketide, a phylogeny-guided discovery of a new polyketide from *Talaromyces funiculosus* Salicorn 58. *Appl. Microbiol. Biotechnol.* **100**, 5323–5338 (2016).

36. Atilano, M. L. *et al.* Bacterial autolysins trim cell surface peptidoglycan to prevent detection by the Drosophila innate immune system. *Elife* **3**, e02277 (2014).

37. Humann, J. & Lenz, L. L. Bacterial peptidoglycan degrading enzymes and their

impact on host muropeptide detection. *J. Innate Immun.* **1**, 88–97 (2009).

38. Skalweit, M. J. & Li, M. Bulgecin A as a β-lactam enhancer for carbapenem-resistant *Pseudomonas aeruginosa* and carbapenem-resistant *Acinetobacter baumannii* clinical isolates containing various resistance mechanisms. *Drug Des. Devel. Ther.* **10**, 3013–3020 (2016).

39. King, A. M. *et al.* Aspergillomarasmine A overcomes metallo-β-lactamase antibiotic resistance. *Nature* **510**, 503–6 (2014).

40. Unemo, M. *et al.* The novel 2016 WHO Neisseria gonorrhoeae reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. *J. Antimicrob. Chemother.* **71**, 3096–3108 (2016).

41. Koo, B.-M. *et al.* Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst* **4**, 291–305 (2017).

42. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).

43. Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E. & Cryle, M. J. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **521**, 105–109 (2015).

44. Wade, J. J. & Graver, M. A. A fully defined, clear and protein-free liquid medium permitting dense growth of *Neisseria gonorrhoeae* from very low inocula. *FEMS Microbiol. Lett.* **273**, 35–37 (2007).

45. Gonsior, M. *et al.* Biosynthesis of the peptide antibiotic feglymycin by a linear nonribosomal peptide synthetase mechanism. *ChemBioChem* **16**, 2610–2614 (2015).

46. Bhavsar, A. P., Zhao, X. & Brown, E. D. Development and characterization of a xylose-dependent system for expression of cloned genes in *Bacillus subtilis*: conditional complementation of a teichoic acid mutant. *Appl. Environ. Microbiol.* **67**, 403–10 (2001).

47. Liu, T.-Y., Chu, S.-H. & Shaw, G.-C. Deletion of the cell wall peptidoglycan hydrolase gene *cwlO* or *lytE* severely impairs transformation efficiency in *Bacillus subtilis*. *J. Gen. Appl. Microbiol.* **64**, 139–144 (2018).

48. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).

49. Deatherage, D. E. & Barrick, J. E. Identification of Mutations in Laboratory-

Evolved Microbes from Next-Generation Sequencing Data Using breseq. in *Methods in molecular biology (Clifton, N.J.)* **1151**, 165–188 (2014).

50. Dajkovic, A. *et al.* Hydrolysis of peptidoglycan is modulated by amidation of meso-diaminopimelic acid and Mg2+ in *Bacillus subtilis*. *Mol. Microbiol.* **104**, 972–988 (2017).

51. Formstone, A. & Errington, J. A magnesium-dependent *mreB* null mutant: implications for the role of *mreB* in *Bacillus subtilis*. *Mol. Microbiol.* **55**, 1646–1657 (2005).

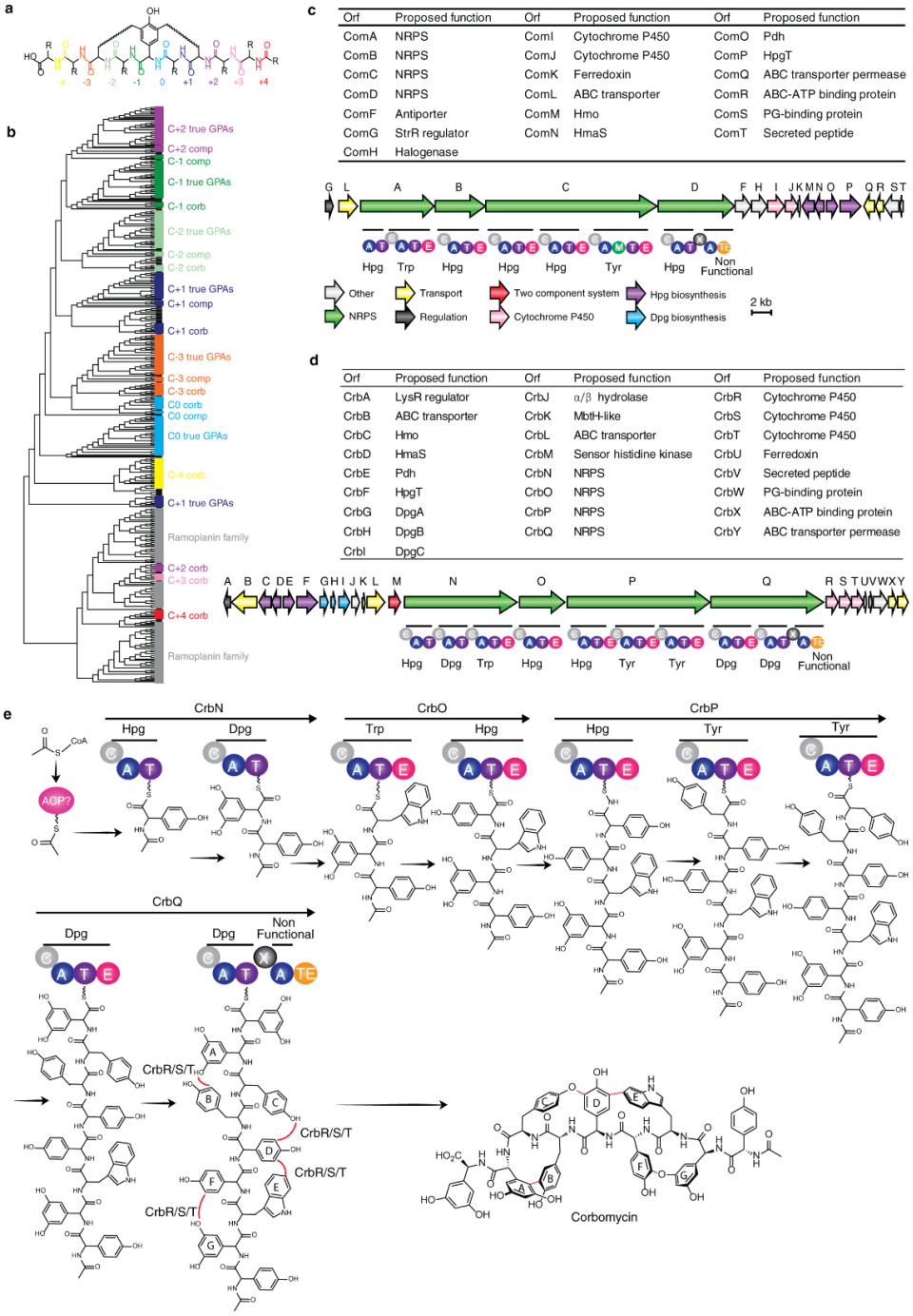
52. Schaub, R. E. & Dillard, J. P. Digestion of peptidoglycan and analysis of soluble fragments. *Bio-protocol* **7**, 2438 (2017).

53. Tiyanont, K. *et al.* Imaging peptidoglycan biosynthesis in *Bacillus subtilis* with fluorescent antibiotics. *Proc. Natl. Acad. Sci.* **103**, 11033–11038 (2006).

54. Yan, A. *et al.* Transformation of the anticancer drug doxorubicin in the human gut microbiome. *ACS Infect. Dis.* **4**, 68–76 (2018).

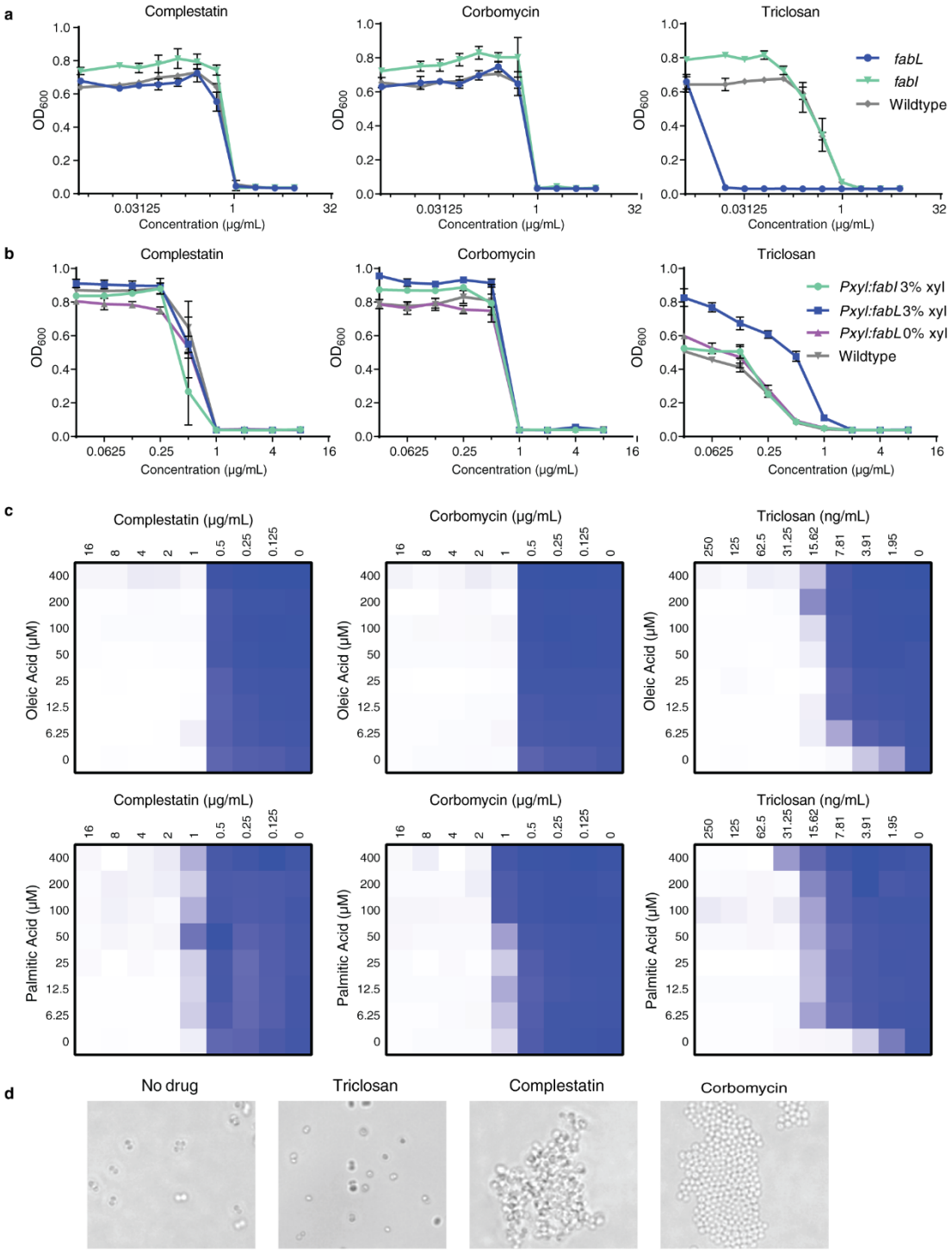
55. Verbist, L. The antimicrobial activity of fusidic add. *J. Antimicrob. Chemother.* **25**, (1990).

EXTENDED DATA



Extended data figure 1. *crb* and *com* biosynthetic gene clusters.
(A) Labeling scheme for condensation domains position with respect to the centrally encoded 4-hydroxypenylglycine. (B) Full maximum-likelihood phylogenetic tree of condensation domains with labels and colors according to scheme in panel A. Fig 1a is an

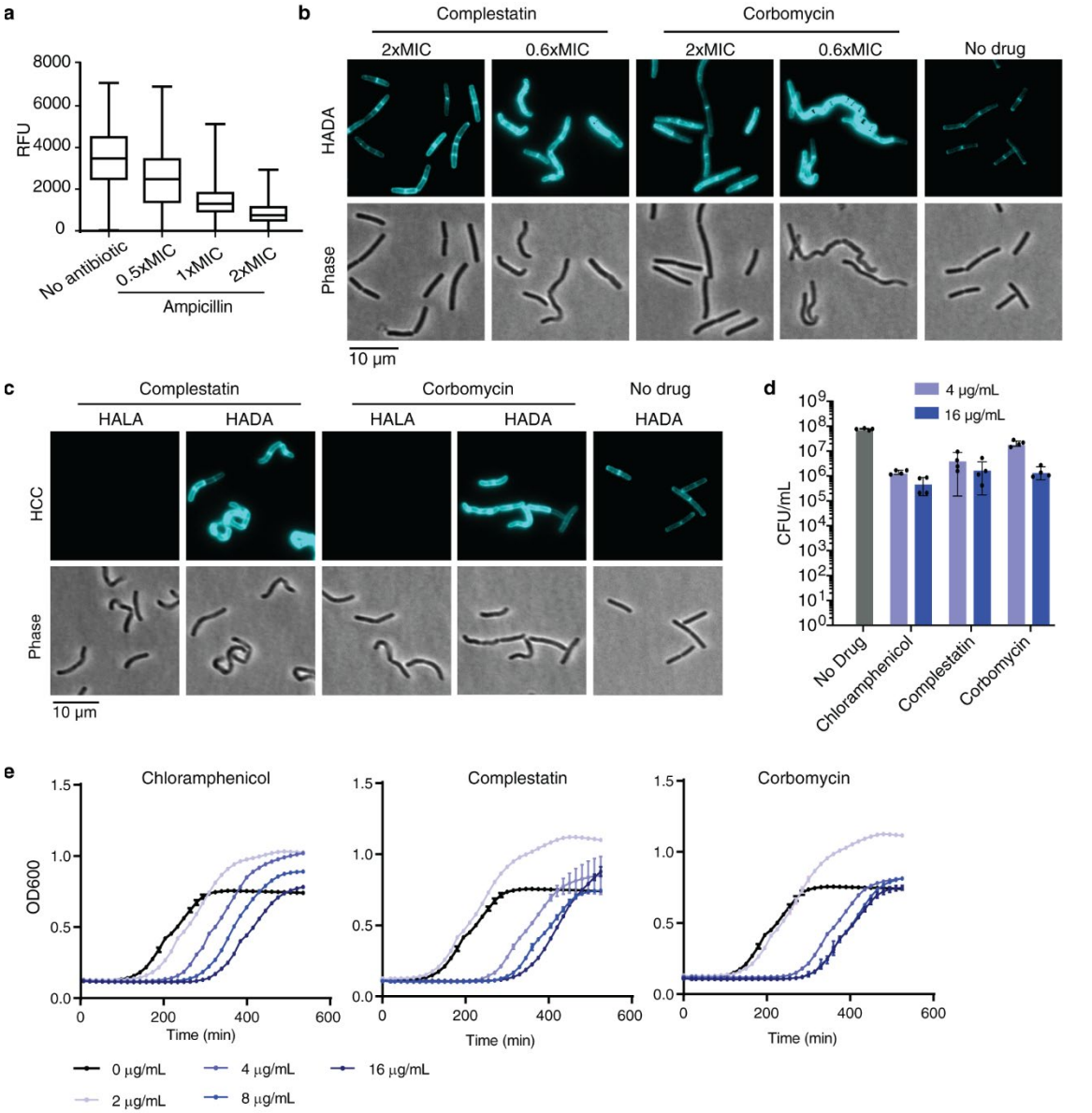
enlargement of the C-2 region. Tree is shown as a rectangular cladogram with midpoint root. Labels indicate well supported clades (> 0.89 SH-like support values) of condensation domains for the true GPAs, complestatin (comp)-type BGCs and corbomycin (corb)-type BGCs. (C) Schematic and gene annotations of WAC01325 complestatin's *com* BGC, and (D) WAC01529 corbomycin's *crb* BGC. Gene names in (C) are based on the previously published complestatin BGC in *S. lavendulae*[10], whose gene synteny matches exactly except for *comE,* an MbtH-like protein that appears fused the C terminus of *comD* in WAC01325's sequence. This may be due to sequencing error or incorrect orf calling, which we performed using Prodigal as part of the antiSMASH workflow. (E) Proposed biosynthetic pathway for corbomycin. A full description is provided in the supplemental discussion.



Extended data figure 2. FabI and FabL are not the target of complestatin or corbomycin.

Dose response curves on *B. subtilis* growth against complestatin, corbomycin and triclosan. (A) *Bacillus subtilis* has two enoyl-ACP reductase orthologs, FabI and FabL.

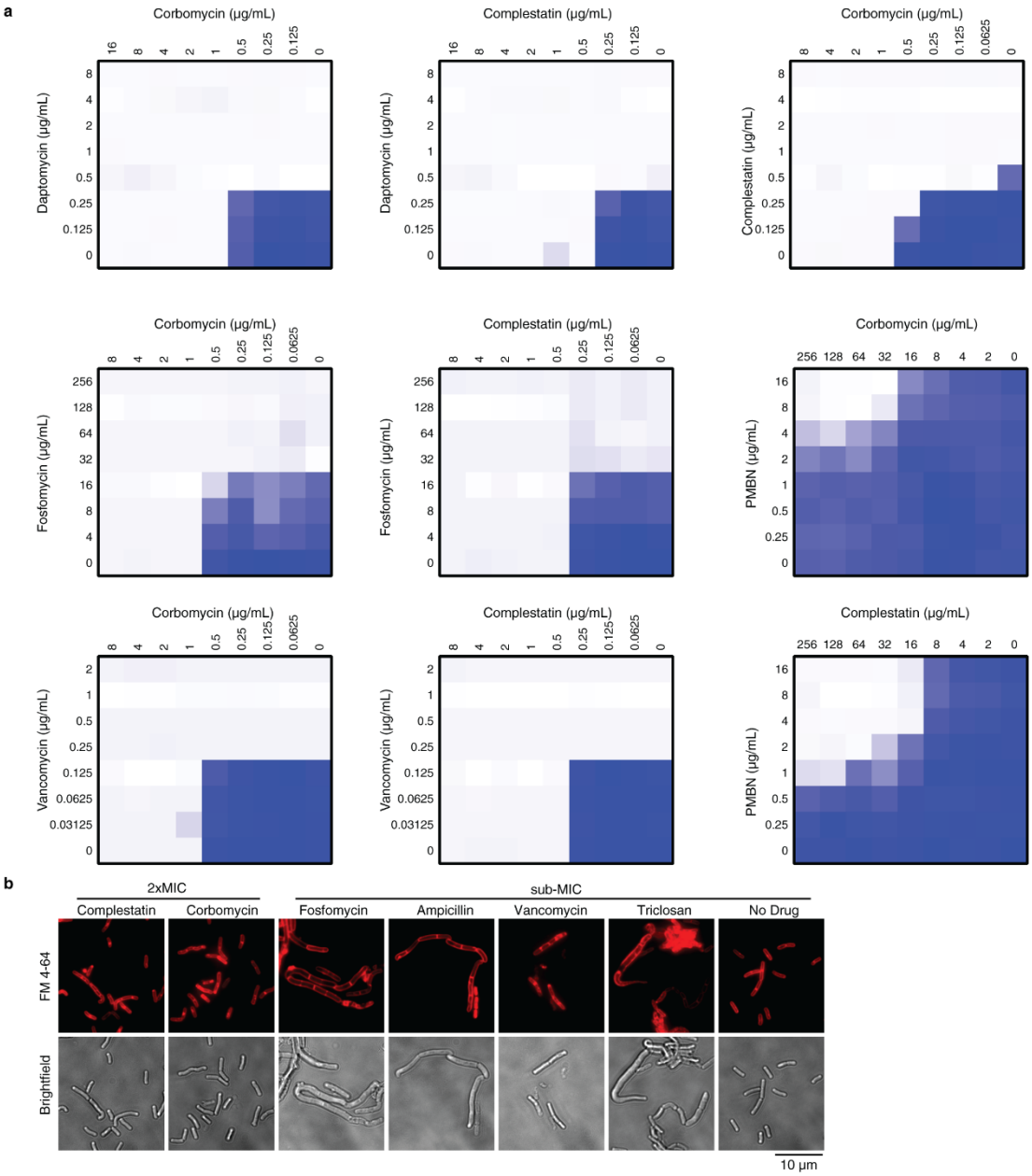
fabI and *fabL* gene deletions (BKK library strains BKK11720 and BKK08650, respectively[41]) do not affect complestatin or corbomycin susceptibility, but sensitize cells to triclosan, a strong *fabI* inhibitor, when *fabL* is deleted. (B) *fabI* and *fabL* overexpression does not affect complestatin or corbomycin susceptibility but protects cells from triclosan when *fabL* is overexpressed. Genes were expressed from the pSWEET integrative plasmid under the control of a xylose inducible promoter[46]. Average and standard deviation of biological triplicates is shown for A-B. (C) Supplementing exogenous fatty acid was previously reported to modestly reduce complestatin susceptibility[15], so was repeated here using 2D checkerboards of fatty acids against complestatin, corbomycin or triclosan in *S. aureus* ATCC29213. There is no robust reduction in susceptibility to complestatin or corbomycin in contrast to triclosan. Color is scaled to OD_{600} with white representing no growth and dark blue representing growth. (D) Brightfield microscopy of *S. aureus* ATCC 29213 grown in 0.6xMIC of triclosan (0.004 μg/mL), complestatin (0.6 μg/mL) or corbomycin (0.6 μg/mL) was used to examine the effect of these antibiotics on phenotype. While complestatin and corbomycin cause a distinct clumping phenotype, triclosan shows no gross morphological defect. All experiments (A-D) were performed on two independent occasions with similar results.



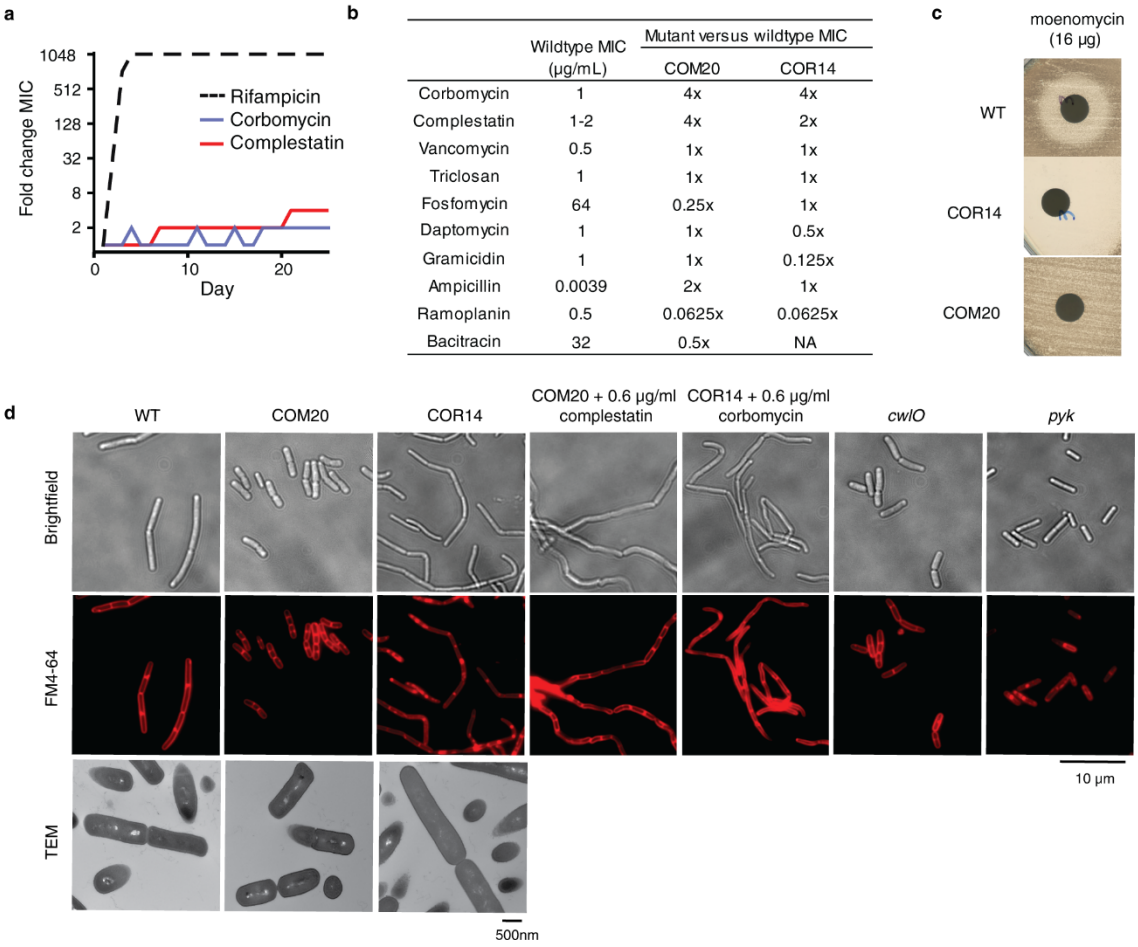
Extended data figure 3. PG biosynthesis is not inhibited by complestatin or corbomycin

(A-B) Incorporation of fluorescent HCC-amino-D-alanine (HADA) into PG of *B. subtilis* treated with complestatin or corbomycin for 5 min prior to continued exposure with HADA for 1 h, or with ampicillin for 1 min prior to FDAA incubation for 5 min. For (A) samples left to right, individual cells measured n = 243, 242, 299, 271. Whiskers show min-max of dataset, box shows upper and lower quartiles, line shows median. Results in (B) are quantified in Figure 2D of main text. (C) Incorporation of HCC-amino-L-alanine (HALA) or HADA, as in panel A. Cells are treated with 0.6xMIC and HALA incorporation is shown to confirm that incorporation represents transpeptidase activity. (D) Enumeration of colony forming units (CFU) after treatment of *B. subtilis* 168 for 6 hr shows that complestatin and corbomycin are bacteriostatic, similar to the control

antibiotic chloramphenicol. Mean and standard deviation CFU values are shown for four independent experimental replicates. (E) Washing and subculturing antibiotic treated cells as prepared in (D) similarly shows successful outgrowth indicating a bacteriostatic antibiotic. Mean and standard deviation is shown for three biological replicates. All experiments were performed twice with similar results.



Extended data figure 4. Complestatin and corbomycin have a novel MOA
(A) 2D checkerboards in *E. coli* BW25113 (polymyxcin B nonapeptide; PMBN) or *B. subtilis* 168 (remainder). For complestatin versus corbomycin checkerboard, fractional inhibitory concentration index = 0.75. Color is scaled to OD_{600} with white representing no growth and dark blue representing growth. (B) Phenotype of *B. subtilis* 168 grown in 2xMIC (8 μg/mL) complestatin or corbomycin for 2 hr, or treated with sub-MIC fosfomycin (100 μg/mL), ampicillin (0.2 μg/mL), vancomycin (1 μg/mL) and triclosan (0.5 μg/mL) until growth to midlog phase. All experiments were performed on two independent occasions with similar results.



b

	Wildtype MIC (µg/mL)	Mutant versus wildtype MIC	
		COM20	COR14
Corbomycin	1	4x	4x
Complestatin	1-2	4x	2x
Vancomycin	0.5	1x	1x
Triclosan	1	1x	1x
Fosfomycin	64	0.25x	1x
Daptomycin	1	1x	0.5x
Gramicidin	1	1x	0.125x
Ampicillin	0.0039	2x	1x
Ramoplanin	0.5	0.0625x	0.0625x
Bacitracin	32	0.5x	NA

e

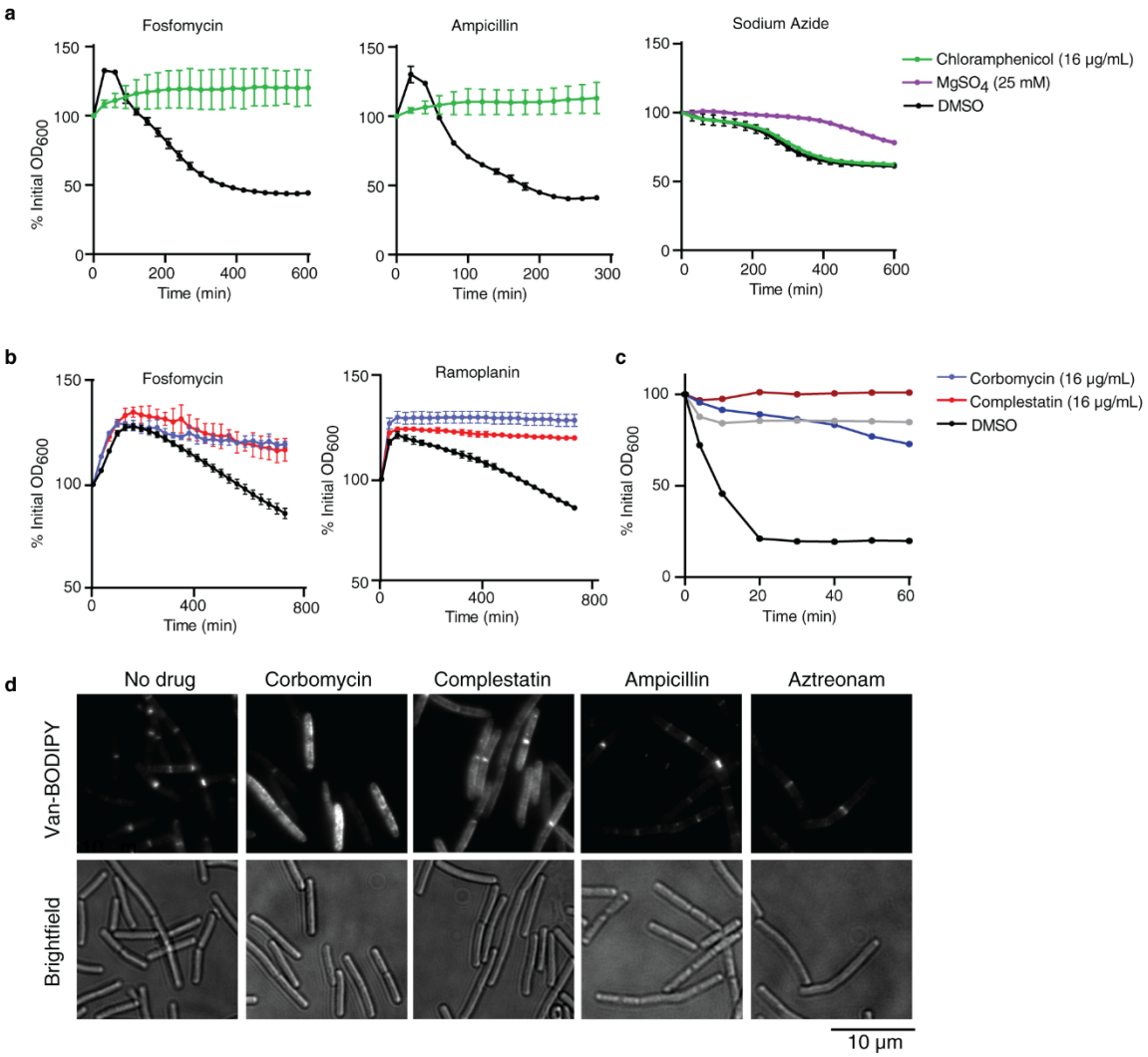
	Mutant versus wildtype MIC			
Gene	Corbomycin	Complestatin	Ampicillin	Ramoplanin
bofC	1x	1x	ND	ND
ftsH	1x	1x	1x	1x
pyk	2x	1x	1x	0.5x
*mbl**	1x	1x	0.25x	0.25x
ykaA	1x	1x	ND	ND
ymdB	2x	2x	2x	0.25x
cwlO	1-2x	1-2x	1x	0.5x
ftsX	1x	1x	1x	1x

f

	MIC (µg/mL)	
B. subtilis	Complestatin	Corbomycin
Wildtype	1-2	1
COM20	4-8	2
pSWEET-cwlO	1-2	1
pSWEET-cwlO Q408*	1-2	1
cwlO pSWEET-cwlO	2	1-2
cwlO pSWEET-cwlO Q408*	2	1-2

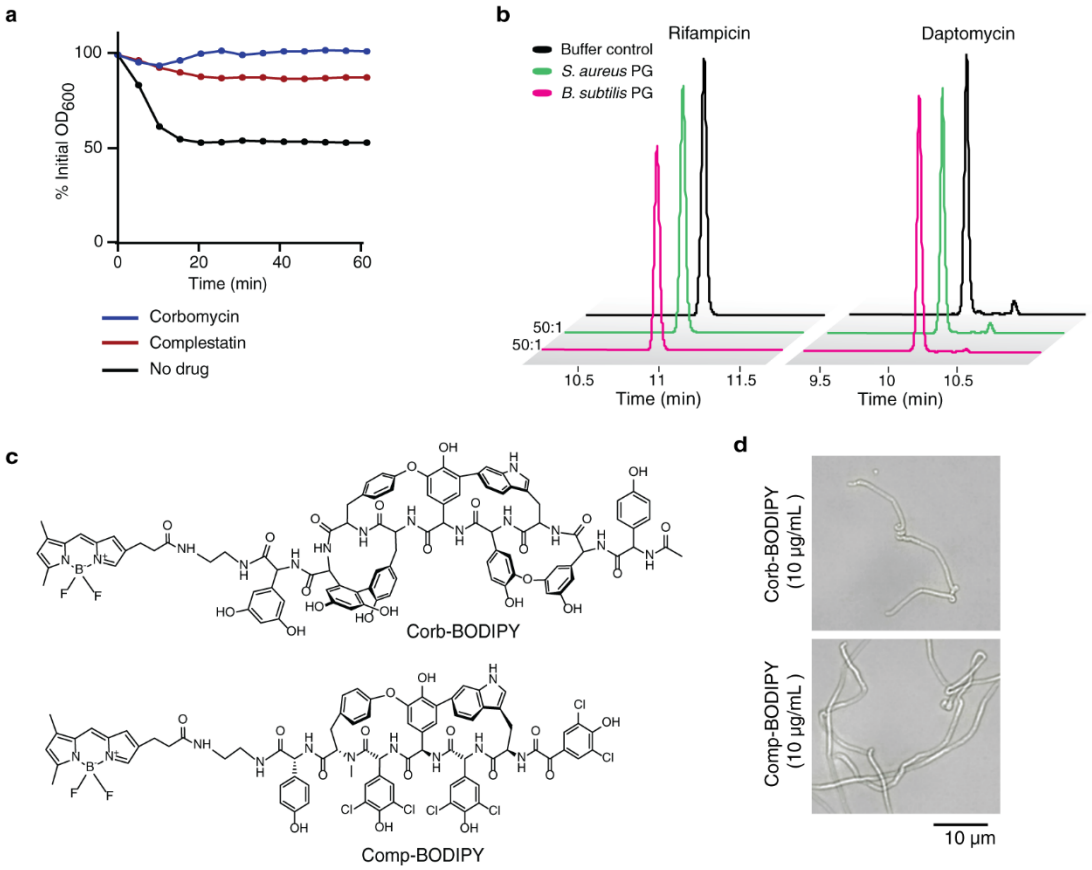
Extended data figure 5. Mutants raised with complestatin or corbomycin
(A) Results of serially passaging *S. aureus* in subMIC antibiotic over 25 days. Passaging was performed on two independent lines with similar results (B) MICs of strains COM14 and COR20 against a variety of antibiotics, with respect to *B. subtilis* 168. The Δ*mbl* strain was not viable in 40 µg/mL $ZnCl_2$ used for bacitracin, so MIC determination was not possible. (C) Susceptibility of *B. subtilis* strains to moenomycin was assessed by Kirby-Bauer assay on Mueller-Hinton agar, which results in a zone of sub-lethal growth inhibition in the wildtype strain. (D) Brightfield, fluorescence and transmission electron microscopy of resistant mutants COM20 and COR14 in comparison to wildtype and

independently generated mutants in relevant genes. (E) Fold MIC increase as compared to wildtype *B. subtilis* 168 of various independently generated mutants. All knockouts were obtained from a non-essential gene knockout library (see extended data table 2). ND- not determined. * *mbl* mutant MIC measured in the presence of 25 mM $MgCl_2$. (F) The effects of CwlO knockout and overexpression, either full length or the truncated version identified in COM20, on antibiotic susceptibility was determined. MICs were measured in the presence of 3% xylose to induce expression from P_{xyl}. Results are described in supplemental discussion. All experiments were performed on at least two independent occasions with similar results.

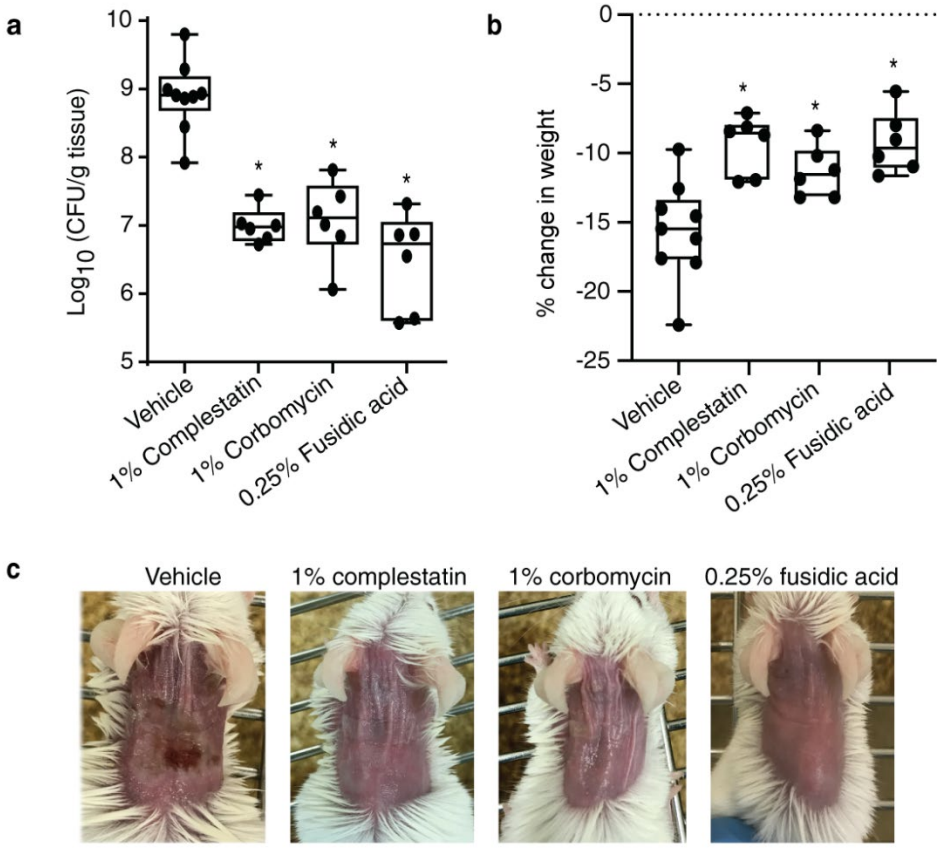


Extended data figure 6. Complestatin and corbomycin block autolysins i*n vitro* and *in vivo.*

(A) Effect on chloramphenicol on *B. subtilis* lysis. Cells were treated with ampicillin (100 µg/mL), fosfomycin (50 µg/mL), or sodium azide (75 mM) to induce lysis, in addition to chloramphenicol, $MgSO_4$ or solvent control. Excess Mg^{2+} inhibits azide induced lysis by a mechanism not well understood but thought to be partially through modulating autolysin activity[50,51]. (B) Complestatin and corbomycin antagonism of *S. aureus* lysis induced by fosfomycin (50 µg/mL) or ramoplanin (50 µg/mL). Average values and standard deviation from biological triplicates are plotted for A-B. (C) Activity of hen egg-white lysozyme on PG incubated with 400 µg/mL complestatin or corbomycin. The assay was performed in technical triplicate with similar results. Legend refers to both (B) and (C). (D) Vancomycin-BODIPY FL (Van-BODIPY) staining shows nascent PG containing D-Ala-D-Ala in cells treated with 10xMIC antibiotic for 30 min. All experiments were performed twice with similar results.



Extended data figure 7. Tools used to measure PG binding.
(A) Activity of CwlO on PG preincubated with 100 µg/mL antibiotic, then washed and resuspended in fresh buffer. The assay performed in technical triplicate with similar results. (B) HPLC chromatograms of antibiotic left unbound after incubation with either *B. subtilis* or *S. aureus* PG. Additional negative control drugs rifampicin (abs 330 nm) and daptomycin (abs 280 nm) were tested in conditions identical to those described in figure 4a of the main text. Incubation with PG does not reduce peak height, indicating the absence of non-specific binding. (C) Chemical structures of corbomycin-BODIPY FL (Corb-BODIPY) and complestatin-BODIPY FL (Comp-BODIPY). (D) BODIPY derivatives of corbomycin and complestatin induce the same twisted cell chaining as parent compounds, so retain the same mechanism of action. All experiments were performed on two independent occasions with comparable results.



Extended data figure 8. Complestatin and corbomycin are effective in a murine skin infection model.

Topical application of antibiotic in petroleum jelly (% w/w) treated MRSA in a skin infection model using neutropenic mice. Mice were assessed 33 h post-infection and compared to vehicle control (petroleum jelly containing 10% DMSO). (A) The colony-forming units (CFU)/g tissue was determined. Significance is tested by one-way ANOVA on ranks with Kruskal-Wallis test, *P* values vs. vehicle as follows: 1% complestatin = 0.0062, 1% corbomycin = 0.0157, 0.25% fusidic acid = 0.0003. (B) Percent weight change post infection compared to pre infection for each mouse was measured. Significance is tested by two-sided unpaired Mann-Whitney t-test, *P*-values vs. vehicle as follows: 1% complestatin = 0.0016, 1% corbomycin = 0.016, 0.25% fusidic acid = 0.0028. For panels A-B, * denotes significance, and whiskers show min-max of dataset, box shows upper and lower quartiles, line shows median. (C) Visual wound inspection of representative mice in each treatment group. Each treatment condition in A-C was tested with replicate mice (n= 6 complestatin, corbomycin, fusidic acid; n = 9 vehicle) divided between two independent experiments.

Extended Data Table 1. MICs of Complestatin and Corbomycin. Each MIC was measured in technical triplicate and on at least two independent occasions with the same result.

MIC (μg/mL)	Comp	Corb	Vanco	Resistance Phenotype
Escherichia coli BW25113	>256	>256	256	
E. coli BW25113 *ΔbamB ΔtolC*	64	32	8	
Pseudomonas aeruginosa PA01	>128	>128	>128	
Acinetobacter baumannii ATCC 17978	64	64	>128	
Enterobacter aerogenes ATCC 13048	>128	>128	>128	
Neisseria gonorrhoeae ATCC 49226	0.5	0.25	8	
Neisseria gonorrhoeae WHOX	0.5	0.5	8	Penicillin G, Cefixime, Ceftriaxone, Ciprofloxacin, Tetracycline
Neisseria gonorrhoeae WHOZ	2	0.5	8	Penicillin G, Cefixime, Ceftriaxone, Ciprofloxacin, Azithromycin Tetracycline
B. subtilis 168	1	1	0.25	
B. subtilis EB1451 (*ΔtagO*)	1	1	0.25	
B. subtilis COM20	4	4	0.25	Complestatin raised mutant day 20
B. subtilis COM25	4	4	0.25	Complestatin raised mutant day 25
B. subtilis COR14	4	2	0.25	Corbomycin raised mutant day 14
B. subtilis COR25	4	2	0.25	Corbomycin raised mutant day 25
S. aureus ATCC 29213	2	1	1	
S. aureus ATCC 33591	2	2	1	MRSA
S. aureus USA300	2	1	1	MRSA
S. aureus C0017	1	1	4	Fluoroquinolones, Macrolides, VISA
S. aureus C0018	1	0.5	4	Ciprofloxacin, Macrolides, VISA
S. aureus C1424	0.5	0.5	8	Ciprofloxacin, Daptomycin, VISA
S. aureus GDW2295	1	1	8	Daptomycin, VISA
S. aureus Mu50	0.5	1	8	VISA
S. aureus Mu3	0.5	1	2	hVISA
Enterococcus faecium ATCC 19434	64	4	2	
Enterococcus faecium ATCC 700221	64	4	>256	Vancomycin and Teicoplanin (VREA)
Enterococcus faecalis ATCC 29212	32	4	2	
Enterococcus faecalis ATCC 51299	32	4	32	Vancomycin (VREB)
Micrococcus luteus GDW1580	1	0.5	1	
Mycobacterium smegmatis MC^2155	128	256	1	
Streptomyces venezuelae ATCC 10712	8-16	8-16	0.25	
Streptomyces coelicolor M1154	16-32	16	256	Vancomycin
Streptomyces sp. WAC01529	32	32	1	
Streptomyces sp. WAC01325	32	16	0.25	
Saccharomyces cerevisiae ATCC 201389	32	>256		
Candida albicans ATCC 90028	>256	>256		
HEK cells (IC_{50})	>256	232		

Extended Data Table 2. Strains used in this study.

Strains	Strain	Description	Reference
Pseudomonas aeurginosa	PAO1	wildtype	
Escherichia coli	BW25113	wildtype	
	BW25113 Δ*bamB* Δ*tolC*	hyperpermeable, efflux deficient	38
Acinetobacter baumannii	ATCC 17978	wildtype	
Neisseria gonorrhoeae	ATCC 49226	wildtype	
	WHOX	Penicillin G, Cefixime, Ceftriaxone, Ciprofloxacin, Tetracycline resistant	39
	WHOZ	Penicillin G, Cefixime, Ceftriaxone, Ciprofloxacin, Azithromycin Tetracycline resistant	39
Enterobacter aerogenes	ATCC 13048	wildtype	
Staphylococcus aureus	ATCC 29213	wildtype	
	ATCC 33591	MRSA	
	USA300	MRSA	
	Mu50	VISA	
	Mu3	hVISA	
	GDW2295		Wright lab strain collection
	C0017	Fluoroquinolones, Macrolide resistant, VISA	Wright Clinical Collection
	C0018	Ciprofloxacin, Macrolide resistant, VISA	Wright Clinical Collection
	C1464	VISA, Daptomycin resistant	Wright Clinical Collection
Enterococcus faecium	ATCC 19434	wildtype	
	ATCC 700221	VREA	
Enterococcus faecalis	ATCC 29212	wildtype	
	ATCC 51299	VREB	
Micrococcus luteus	GDW1580	wildtype	Wright lab strain collection
Mycobacteria smegmatis	Mc^2155	wildtype	
Bacillus subtilis	168	trpC2, wildtype	
	EB1385	P_{ywac}:*luxABCDE, erm*^{R}	17
	BKK11720	168 Δ*fabI::kan*^{r}	Bacillus genetic stock center (BGSC) and 40
	BKK08650	168 Δ*fabL::kan*^{r}	BGSC
	BKK12850	168 Δ*ykaA::kan*^{r}	BGSC
	BKK27750	168 Δ*bofC::kan*^{r}	BGSC
	BKK34800	168 Δ*cwlO::kan*^{r}	BGSC
	BKE34800	168 Δ*cwlO::erm*^{r}	BGSC
	BKE36410	168 Δ*mbl::erm*^{r}	BGSC
	BKK29810	168 Δ*pyk::kan*^{r}	BGSC
	BKK16970	168 Δ*ymdB::kan*^{r}	BGSC
	BKK00690	168 Δ*ftsH::kan*^{r}	BGSC
	BKK35250	168 Δ*ftsX::kan*^{r}	BGSC
	pSWEET-fabL	168 *amyE::xylR* P_{xyl}-*fabL cat86*	This work
	pSWEET-fabI	168 *amyE::xylR* P_{xyl}-*fabI cat86*	This work
	pSWEET-cwlO	168 *amyE::xylR* P_{xyl}-*cwlO cat86*	This work
	pSWEET-cwlO Q408*	168 *amyE::xylR* P_{xyl}-*cwlO*$_{1-407}$ *cat86*	This work
	Δ*cwlO* pSWEET-cwlO	168 Δ*cwlO::erm*^{r} *amyE::xylR Pxyl-cwlO cat86*	This work
	Δ*cwlO* pSWEET-cwlO Q408*	168 Δ*cwlO::erm*^{r} *amyE::xylR Pxyl-cwlO*$_{1-407}$ *cat86*	This work

SUPPLEMENTARY INFORMATION

Supplementary Discussion

Structural characterization of corbomycin

Corbomycin was obtained as a white amorphous solid. To elucidate its structure, the compound was subjected to 1D and 2D nuclear magnetic resonance (NMR) as well as HR-ESI-MS (Supplementary Figure 1-6 and Supplementary table 1). The ESI-MS molecular ion cluster at m/z $[M+H]^+$ 1507.4434 (cal 1507.4432, Supplementary figure. 1c), together with 1D NMR data (Supplementary Figure 2-3) suggested the molecular formula was $C_{79}H_{66}N_{10}O_{22}$. In the ^{1}H-and ^{13}C- NMR spectra of corbomycin (Supplementary Figure 2-3, Supplementary Table 1), the signals of αCHs (δ_H 4.87, 5.14, 4.65, 4.25, 5.34, 5.74, 3.91, 5.05 and 5.39 ppm, and δ_C 58.3, 56.1, 52.1, 54.2, 55.2, 55.8, 56.9, 58.8 and 56.0 ppm, respectively) together with the amide carbon (δ_C 172.5, 169.1, 169.5, 168.7, 166.8, 170.0, 170.9, 168.6 and 170.7 ppm, respectively) indicated the presence of nine amino acid residues. These amino acid residues were determined as two tyrosine, a tryptophan, three 4-hydroxyphenylglycines and three 3,5-dihydroxyphenylglycines according to its long-rang heteronuclear (HMBC) and ^{1}H-^{1}H COSY correlations (Supplementary Figure 1,4,5). The linear sequence of the amino acid residues were established on the HMBC correlations from the protons of αCHs and amides to the amide carbons (Supplementary Figure 6). In the HMBC spectrum, the correlation between proton C3 and carbon E3 suggested carbon C-4 and E-3 was connected with an oxygen atom. The correlations from protons G5 and G7 to carbon E5 indicted the connection between carbon E5 and G6. The HMBC correlations between

proton D2 and carbon B6 suggested carbon D2 connected to carbon B6. The HMBC correlations between proton F6 and carbon H3 suggested carbon F-5 and H-3 was connected with an oxygen atom (Supplementary Figure 6).

Biosynthetic gene cluster and pathway for corbomycin

We propose a biosynthetic pathway for corbomycin similar to classical GPAs[1] (Extended data fig. 1), with a few notable features. Biosynthesis of the precursor Hpg from prephenate involves orthologs of Hmo, HmaS, Pdh and HpgT encoded by *crbC-F,* respectively. Dpg biosynthesis from malonyl-CoA uses orthologs of DpgA, DpgB and DpgC encoded by *crbG-I,* respectively. Interestingly, while Dpg biosynthesis also usually involves DpgD, which acts redundantly with DpgA as a enoyl-coA hydratase[2], the *crb* cluster lacks an ortholog of this gene.

Biosynthesis of the nonapeptide backbone of corbomycin is encoded in a collinear fashion by CrbN-Q. CrbN begins with a starter condensation domain which is likely responsible for transferring an activated acetyl group to the N-terminal Hpg. The *crb* cluster does not encode an acyl-carrier protein (ACP) usually responsible for activation of acetyl-coA or fatty acyl-coAs, so it is possible that this function is encoded elsewhere in the genome or that the starter C domain is capable of direct formation of an amide bond between acetyl-coA and an amino acid moiety, as has been reported in unique cases previously[3]. Adenylation (A) domain specificity was hypothesized *in silico* using AntiSMASH and phylogenetic relationships between *crb* A-domains and GPAs with known structures, and matches that experimentally observed. The A-domain in module 10 is predicted be non-functional from sequence-based analysis, and similar non-functional

A-domains are observed in other type V GPA BGCs including complestatin (Extended data fig. 1a) and kistamicin[4].

Corbomycin contains four crosslinks (A-B aryl, C-O-D phenolic, D-E aryl, F-O-G phenolic) but only three cytochrome p450s, *crbR-T.* Kistamicin, a related type V glycopeptides, uses a bifunctional cytochrome p450 $OxyC_{kis}$ to introduce two crosslinks on its peptide backbone[4], and a similar case may occur here. The X domain encoded in module 10 is involved in recruitment of cytochrome p450s to the NRPS machinery.

Non-biosynthetic genes encoded in *crb* include many of those typically seen in GPA BGCs such as transcriptional activators (*crbA*), ABC transporters for export of the produced metabolite (*crbB* and *crbL*), an α/β hydrolase for removal of misprimed amino acids (*crbJ*)[5], an MbtH-like protein required for the activity of some NRPS machineries including all known GPAs[6], and a ferredoxin for regenerating cytochrome p450s (*crbU*). *crbM* is a sensor histidine kinase similar to VanS, but notably lacking its cognate response regulator nor the requirement for the typical function of VanRS in activating *vanHAX* expression[1]. *crbV-Y* encodes an operon of unknown function or relevance to corbomycin production, but interestingly includes a putative PG-binding domain and has orthologs *comQ-T* in the complestatin BGC.

The complestatin and corbomycin BGCs contain no obvious self-resistance determinant, and indeed producer and non-producer *Streptomyces* spp. have similar MICs (Extended Data Table 1). It is possible that temporal control restricts antibiotic production to stationary phase when autolysins are no longer required for growth, preventing self-inhibition in producers.

Analysis of resistant mutants

In the complestatin raised line, mutations in the two isolates were identified in genes including *ykaA,* a gene of unknown function*,* and *ymdB,* a phosphodiesterase (Fig. 3b). Disruption of *ymdB* results in increased expression of the σ^D regulon including genes for increased motility, a phenotype we observed in COM20, and several autolysins[7]. We also observed shorter, slightly wider cells characteristic of the mutations we identified in *cwlO* or *ftsX* (Extended data fig. 5d)[8]. CwlO is an autolysin responsible for cleavage of PG stem peptides on the lateral cell wall, while FtsX is a transmembrane protein required for CwlO localization to the cell membrane[8,9].

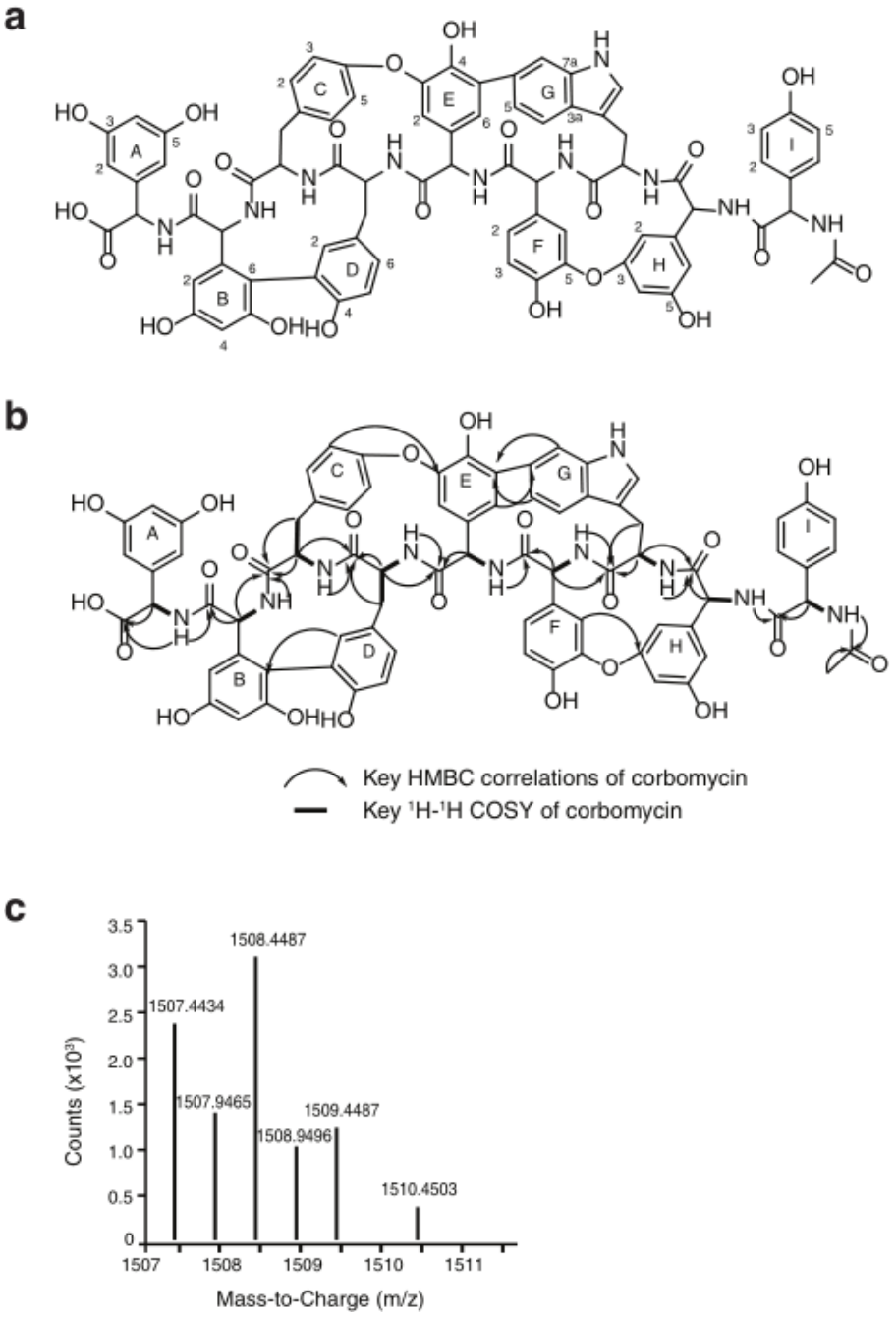
In COM25, mutations were also identified in *ftsH,* a metalloprotease[10], and *pyk,* a pyruvate kinase[11]. Interestingly, mutations in these same two genes were also found in the corbomycin raised isolates (Fig. 3b). *ftsH* mutants are defective in Z ring formation[10], while *pyk* suppresses the lethality of FtsZ depletion[11].

Isolates from the corbomycin raised line, COR14 and COR25, were identical by sequencing. In addition to *pyk* and *ftsH,* we identified mutations in *mbl,* an actin homolog[8], and *bofC,* a general stress protein (Fig 3b). The role of the identified Mbl R54C mutation in disrupting protein function is unclear, as Mbl null mutants are not known to be viable in the absence of excess Mg^{2+}, but notably, Mbl is required for proper CwlO function[8].

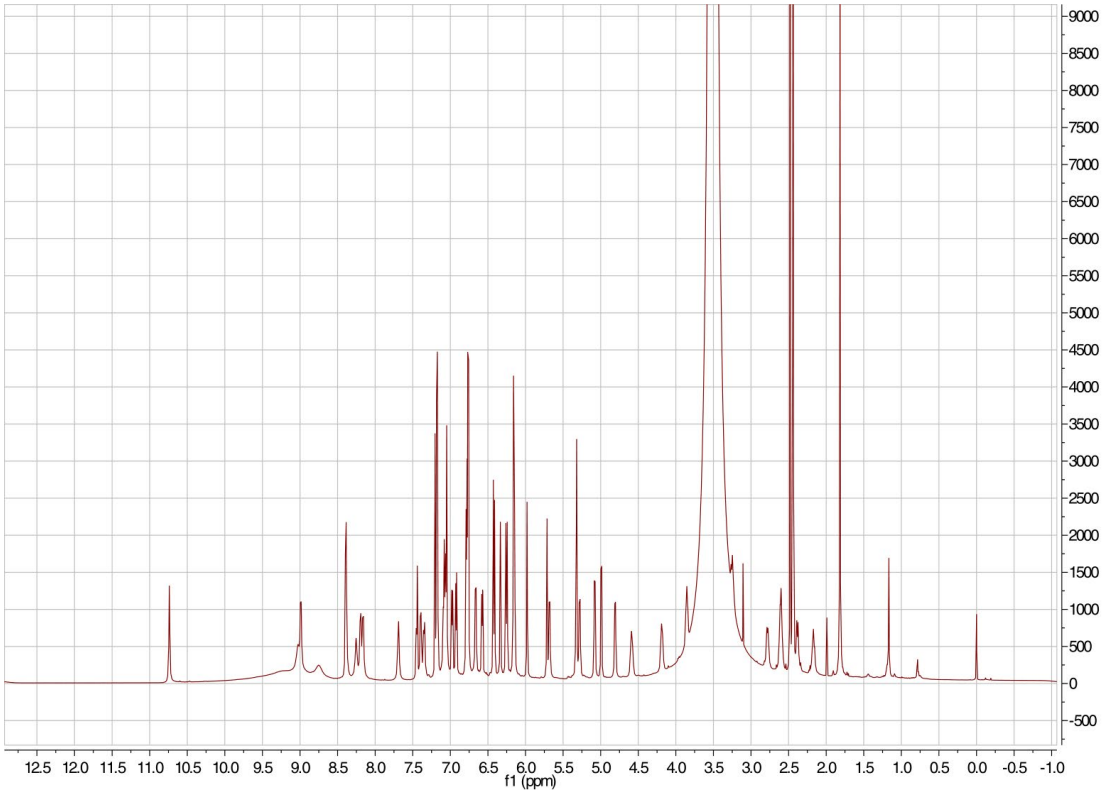
With evidence of autolysin inhibition in mind, we revisited our *B. subtilis* resistant mutants, which paradoxically included mutations to the autolysin CwlO or its regulator FtsX. Deletion of *cwlO* in a wildtype or Δ*ymdB* background conferred ≤2-fold reduced

susceptibility to corbomycin or complestatin, and we ruled out a gain of function role of the truncated CwlO protein by overexpression in *B. subtilis* (Extended data fig. 5e,f). In the presence of sub-MIC antibiotic, COR14 and COM20 forms characteristic twisted chains of cells (Extended data fig. 5d), suggesting that they are still subject to autolysin inhibition and that their mechanism of resistance is indirect. FtsX mutants fail to localize CwlO to the cell membrane[8]. It is possible that disruption of CwlO itself adjusts the activity or localization of other autolysins through regulatory feedback mechanisms, for example through the WalRK regulon, which may increase the expression of other autolysins (e.g. LytE). Absence of CwlO function may also alter PG composition in a way that reduces the motif recognized by corbomycin and complestatin, or even enriches for this motif in order to titrate excess antibiotic.

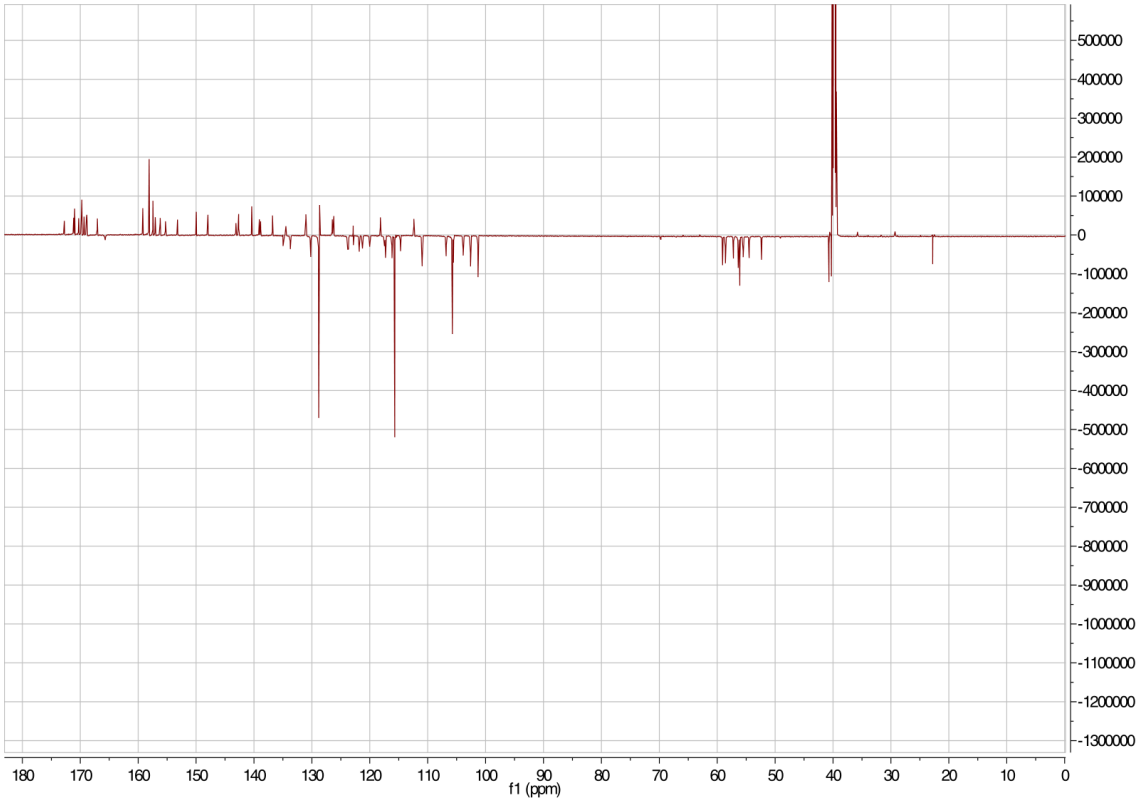
Supplementary Tables and Figures



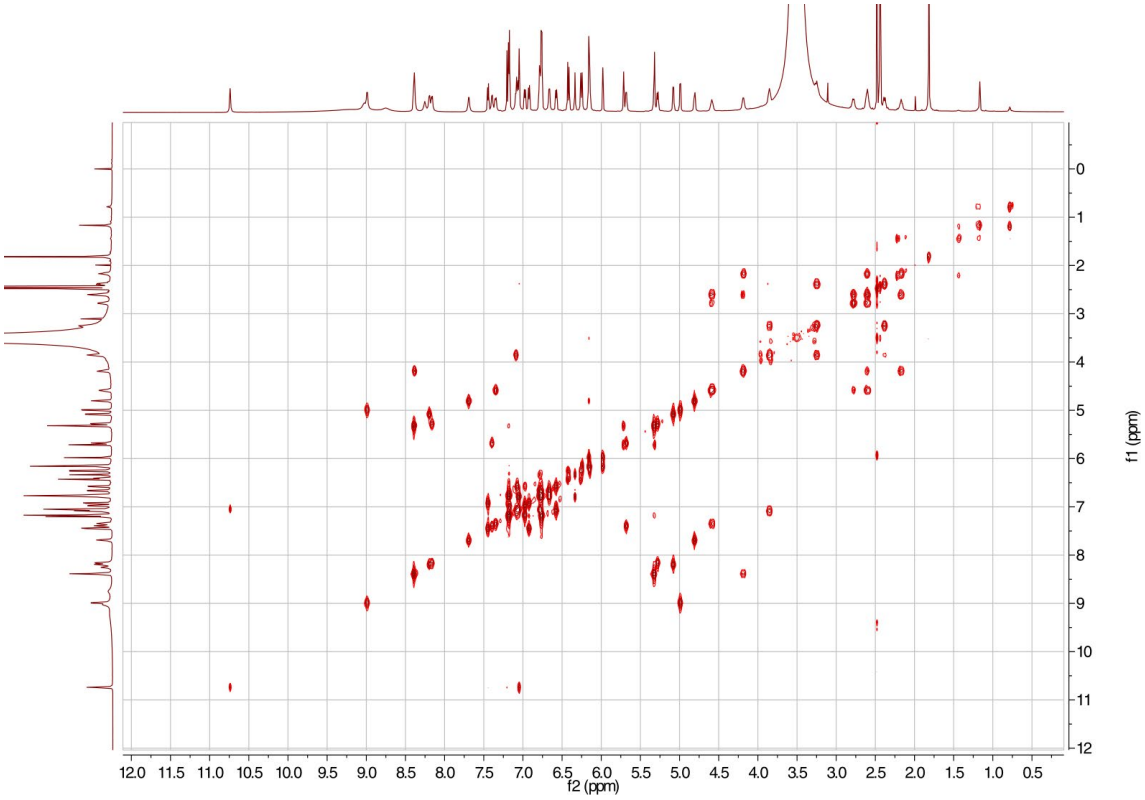
Supplementary Figure 1. Structural key to NMR spectra for corbomycin.
(A) Positional assignments and (B) key correlations of corbomycin relevant to NMR spectra. (C) High resolution ESI-MS of corbomycin showing observed $[M+H]^+$ 1507.4434 (expected = 1507.4432, error = 0.132 ppm).



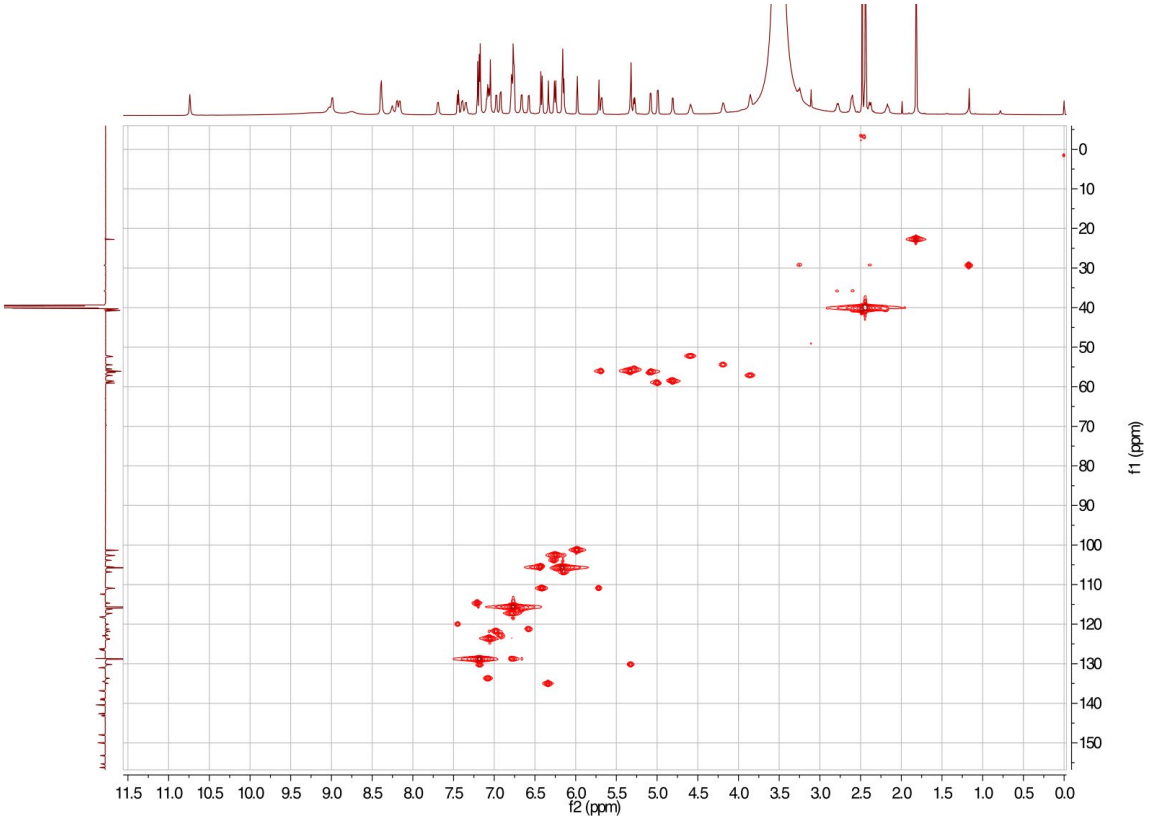
Supplementary Figure 2. ^{1}H NMR spectra for corbomycin.



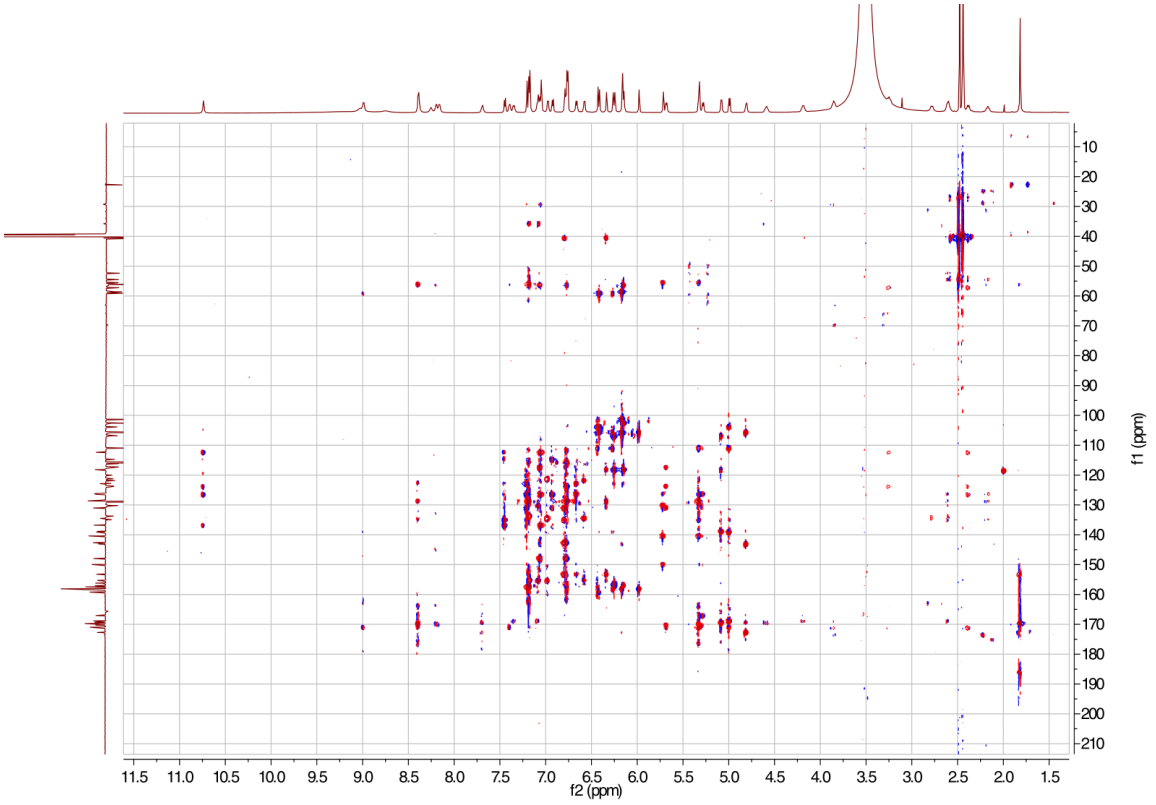
Supplementary Figure 3. ^{13}C NMR spectra for corbomycin



Supplementary Figure 4. ^{1}H-^{1}H COSY NMR spectra for corbomycin



Supplementary Figure 5. HSQC NMR spectra for corbomycin



Supplementary Figure 6. HMBC NMR spectra for corbomycin

Supplementary Table 1. ^{1}H- and ^{13}C-NMR data for corbomycin.

No.	δ_H	δ_C	No.	δ_H	δ_C
A-1		142.8	E-C=O		166.8
A-2, 6	6.22 (2H, br s)	105.5	E-NH	8.22 (d, *J* = 14 Hz)	
A-3, 5		157.8	F-1		130.7
A-4	6.04 (br s)	101.0	F-2	7.12 (m, overlapped)	123.42
A-αCH	4.87 (d, *J* = 7.0 Hz)	58.3	F-3	6.84 (m, overlapped)	117.0
A-C=O		172.5	F-4		147.7
A-NH	7.75 (d, *J* = 7.0 Hz)		F-5		142.4
B-1		138.6	F-6	6.82 (m, overlapped)	117.2
B-2	6.21 (br s)	106.5	F-αCH	5.74 (br d, *J* =7.0 Hz)	55.8
B-3		156.8	F-C=O		170.0
B-4	6.30 (br s)	102.3	F-NH	7.46 (d, *J* = 7.0 Hz)	
B-5		155.9	G-1NH	10.80 (br s)	
B-6		117.9	G-2	7.11 (m, overlapped)	123.5
B-αCH	5.14 (d, *J* = 7.0 Hz)	56.1	G-3		112.1
B-C=O		169.1	G-3a		126.2
B-NH	8.26 (d, *J* = 7.0 Hz)		G-4	7.51 (d, *J* = 7.0 Hz)	119.7
C-1		134.3	G-5	6.98 (d, *J* = 7.0 Hz)	122.56
C-2	7.24 (m, overlapped)	129.9	G-6		134.74
C-3	7.04 (br d, *J* = 7.0 Hz)	121.5	G-7	7.26 (s)	114.4
C-4		155.0	G-7a		136.5
C-5	6.63 (br d, *J* = 7.0 Hz)	121.0	G-αCH	3.91 (dd, *J* = 7.0, 14.0 Hz)	56.9
C-6	7.14 (m, overlapped)	133.4	G-βCH_2	2.44 (dd, *J* = 7.0, 14.0 Hz)	29.0
				3.31 (dd, *J* = 14.0, 14.0 Hz)	
C-αCH	4.65 (br dd, *J* =14, 7.0 Hz)	52.1	G-C=O		170.9
C-βCH_2	2.84 (br d, *J* =14Hz)	35.5	G-NH	7.14 (m, overlapped)	
	2.66 (m, overlapped)				
C-C=O		169.5	H-1		138.8
C-NH	7.41 (d, *J* = 14.0 Hz)		H-2	6.32 (s)	103.6
D-1		126.05	H-3		157.8
D-2	6.40 (br s)	134.70	H-4	6.49 (s)	105.3
D-3		122.56	H-5		158.9
D-4		152.9	H-6	6.47 (s)	110.7
D-5	6.72 (d, *J* = 7.0 Hz)	115.9	H-αCH	5.05 (d, *J* = 7.0 Hz)	58.8
D-6	6.84 (m, overlapped)	128.5	H-C=O		168.6
D-αCH	4.25 (br dd, *J* =14, 7.0 Hz)	54.2	H-NH	9.05 (d, *J* = 7.0 Hz)	
D-βCH_2	2.66 (m, overlapped)	40.3	I-1		128.4
	2.23 (t, *J* = 14 Hz)				
D-C=O		168.7	I-2, 6	7.24 (2H, d, *J* = 7.0 Hz)	128.5
D-NH	8.45 (d, *J* = 7.0 Hz)		I-3, 5	6.83 (2H, m, overlapped)	115.4
E-1		125.9	I-4		157.2
E-2	5.77 (br s)	110.7	I-αCH	5.39 (d, *J* = 7.0 Hz)	56.0
E-3		149.7	I-C=O		170.7
E-4		140.1	I-NH	8.45 (d, *J* = 7.0 Hz)	
E-5		130.8	I-Me-C=O	1.88 (s)	22.5
E-6	5.38 (br s)	129.9	I-Me-C=O		169.4
E-αCH	5.34 (d, *J* = 14 Hz)	55.2			

Supplementary information references

1. Yim, G., Thaker, M. N., Koteva, K. & Wright, G. Glycopeptide antibiotic biosynthesis. *J. Antibiot. (Tokyo).* **67**, 31–41 (2014).

2. Chen, H., Tseng, C. C., Hubbard, B. K. & Walsh, C. T. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14901–6 (2001).

3. Kraas, F. I., Helmetag, V., Wittmann, M., Strieker, M. & Marahiel, M. A. Functional Dissection of Surfactin Synthetase Initiation Module Reveals Insights into the Mechanism of Lipoinitiation. *Chem. Biol.* **17**, 872–880 (2010).

4. Greule, A. *et al.* Kistamicin biosynthesis reveals the biosynthetic requirements for production of highly crosslinked glycopeptide antibiotics. *Nat. Commun.* **10**, 2613 (2019).

5. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci.* **99**, 14083–14088 (2002).

6. Herbst, D. A., Boll, B., Zocher, G., Stehle, T. & Heide, L. Structural basis of the interaction of mbth-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.* **288**, 1991–2003 (2013).

7. Diethmaier, C. *et al.* A novel factor controlling bistability in *Bacillus subtilis*: the YmdB protein affects flagellin expression and biofilm formation. *J. Bacteriol.* **193**, 5997–6007 (2011).

8. Domínguez-Cuevas, P., Porcelli, I., Daniel, R. A. & Errington, J. Differentiated roles for MreB-actin isologues and autolytic enzymes in *Bacillus subtilis* morphogenesis. *Mol. Microbiol.* **89**, 1084–1098 (2013).

9. Meisner, J. *et al.* FtsEX is required for CwlO peptidoglycan hydrolase activity during cell wall elongation in *Bacillus subtilis*. *Mol. Microbiol.* **89**, 1069–83 (2013).

10. Yepes, A. *et al.* The biofilm formation defect of a *Bacillus subtilis* flotillin-defective mutant involves the protease FtsH. *Mol. Microbiol.* **86**, 457–471 (2012).

11. Monahan, L. G., Hajduk, I. V., Blaber, S. P., Charles, I. G. & Harry, E. J. Coordinating bacterial cell division with nutrient availability: a role for glycolysis. *MBio* **5**, e00935-14 (2014).

Chapter IV- Target-directed genome mining identifies ClpP associated natural products

PREFACE

The work presented in this chapter is in preparation as of September 17, 2020:

Culp, E.J., Hobson, C., Pawlowski, A.C., Wright, G.D. Target-directed genome mining identifies ClpP associated natural products. *In preparation.*

E.J.C. and G.D.W. conceived the study and designed experiments. C.H. performed NMR structural elucidation. A.C.P performed genome mining analysis. E.J.C performed all other experiments. E.J.C. and G.D.W. prepared the manuscript.

ABSTRACT

Intracellular proteolytic complexes play an important role in modelling the proteome in both bacteria and eukaryotes. ClpP is the protease subunit of one such highly conserved proteolytic complex that associates with AAA+ ATPases to select and degrade damaged or specifically targeted protein substrates. Its high conservation and vital biological functions make ClpP a highly attractive drug target. Several synthetic compounds have been developed to target ClpP, however few natural product modulators of ClpP activity have been identified. Here we describe a genome mining strategy to identify novel natural products targeting ClpP from a rich source of bioactive metabolites, the bacterial order Actinomycetales. Informed by the knowledge that antibiotic resistance genes sometimes take the form of a duplicated target gene located within the encoded antibiotic's biosynthetic gene cluster, we identify a family of bacterial pyrrolizidine alkaloids whose gene clusters always contain *clpP*. Characterizing one of these gene clusters, *cac* from *Streptomyces cattleya,* we use heterologous expression to activate the cryptic cluster and produce a series of novel azabicyclenes. We also reconstitute the proteolytic activity of *S. cattleya'*s ClpPs *in vitro* and show that the cluster associated ClpPs within *cac,* Cac16/Cac17, form a complex that functions analogously to housekeeping ClpPs and may be able to complement their roles. Finally, we probe the biological function of azabicyclenes and show that while they do not directly inhibit ClpP *in vitro*, they do disrupt the *in vivo* function of housekeeping ClpPs to which protection is provided by Cac16/Cac17. Our findings provide the first evidence of the biological

function of this widespread family of pyrrolizidine alkaloids and highlights the role of bacterial secondary metabolites in regulating the proteome at the post-translational level.

INTRODUCTION

The Caseinolytic protease (ClpP) complex is a central regulator of protein homeostasis in both bacteria[1,2] and eukaryotic mitochondria[3]. As a highly conserved complex, it has been extensively characterized for its role in the degradation of mistranslated proteins also well as specifically targeted proteins for general regulation of the proteome[1–3]. To achieve these controlled functions, ClpP's activity is tightly regulated by formation of a barrel-shaped tetradecameric complex of two stacked heptameric rings that shelter fourteen internal serine protease catalytic sites[4]. Access requires passage through axial openings of the barrel too small for native proteins to pass. To gain access, substrates are selected and unfolded by an Hsp100 family AAA+ superfamily ATPase, including ClpA, ClpX and ClpC, that form heptameric rings and dock on the axial openings[1,2]. By interacting with different adaptor proteins and AAA+ ATPases, specific substrates can be selected.

Pharmacologically targeting ClpP lends itself to the development of both anticancer as well as antibacterial therapeutics. For example, a number of synthetic β-lactones and phenyl-esters have been developed that inhibit bacterial or mitochondrial ClpP[5–9]. While ClpP is dispensable in most bacterial pathogens, its inhibition can be used as an anti-virulence strategy[6,10], and as it is essential in *Mycobacteria tuberculosis,* can also sometimes act as a traditional antibiotic[5]. In the context of anticancer drugs, ClpP appears to be more important to cancer cells' function than normal cells[11]. For example, ClpP

inhibition can selectively kill acute myeloid leukemia cells versus normal hematopoietic cells[12].

Rather than inhibition, inappropriate activation of proteolysis represents another strategy to disrupt ClpP activity. Acyldepsipeptide antibiotics (ADEPs), produced by *Streptomyces hawaiiensis* NRRL15010[13], were the first compounds discovered with this mechanism and have potent antibacterial activity against species even where ClpP is dispensible[14]. By mimicking AAA+ ATPase binding, ADEPs induce widening of the ClpP tetradecamer's axial pores allowing inappropriate substrates to access the catalytic sites and be degraded[15,16]. However, ADEP appears to be rare in nature as *S. hawaiiensis* NRRL15010 remains the only organism in GeneBank containing the ADEP biosynthetic gene cluster (BGC). Indeed, despite its broad conservation and essentiality in actinomycetes, other than a handful of natural products (NPs) targeting the AAA+ ATPase ClpC1[17–20], few NPs targeting ClpP have been described. We therefore undertook a genome mining approach to identify novel ClpP targeting NPs.

To avoid self-intoxication, BGCs contain resistance genes against their encoded antibiotic. When this resistance gene takes the form of an antibiotic-insensitive copy of the targeted protein, searching for BGCs containing that target gene can specifically lead to the identification of new antibiotics. So called "target-directed genome mining" has been previously applied to discover thiolactomycins as inhibitors of fatty acid synthases FabB/F[21,22]. Giving precedence that this form of resistance is possible for a ClpP targeting compound, *clpP* is essential in actinomycetes, and the ADEP BGC encodes a resistant copy of ClpP[23]. We therefore applied target-directed genome mining to identify novel

NPs targeting ClpP. Using this approach, we uncover a family of ClpP associated clusters and investigate their impact on ClpP function.

RESULTS

Target directed genome mining for ClpP directed natural products

To identify ClpP associated clusters, we used the collection of Actinobacterial ClpP proteins as queries to extract 5928 *clpP* genes from RefSeq (Fig. 1a). We included genomic regions 50 kb upstream and downstream of each *clpP* homolog and identified biosynthetic clusters in 1096 of these regions using AntiSMASH[24]. These BGCs were then filtered again for those that contained ClpP within cluster boundaries, as identified through tblastn of *Streptomyces coelicolor* ClpP1, leaving 145 candidate hits.

We next sought to differentiate cluster-associated ClpPs from housekeeping copies coincidentally in proximity to a BGC. Complicating this search, *Streptomyces* spp. usually have at least three copies of *clpP,* and up to five. These five *clpPs* are encoded as three independent transcriptional units, *clpP1clpP2, clpP3clpP4,* and *clpP5*. The *clpP3clpP4* pair is absent in strains with only three paralogs[25]. Each paralog has a distinctive genetic context: *clpP1clpP2* is associated with the AAA+ ATPase *clpX, clpP3clpP4* is associated with the transcriptional regulator *popR,* and *clpP5* is associated with a distinctive gene encoding a protein of unknown function. By filtering out housekeeping ClpPs using these markers, we landed on a final prioritized list of ten BGCs associated with ClpPs (Fig. 1a, Supplementary Table 1).

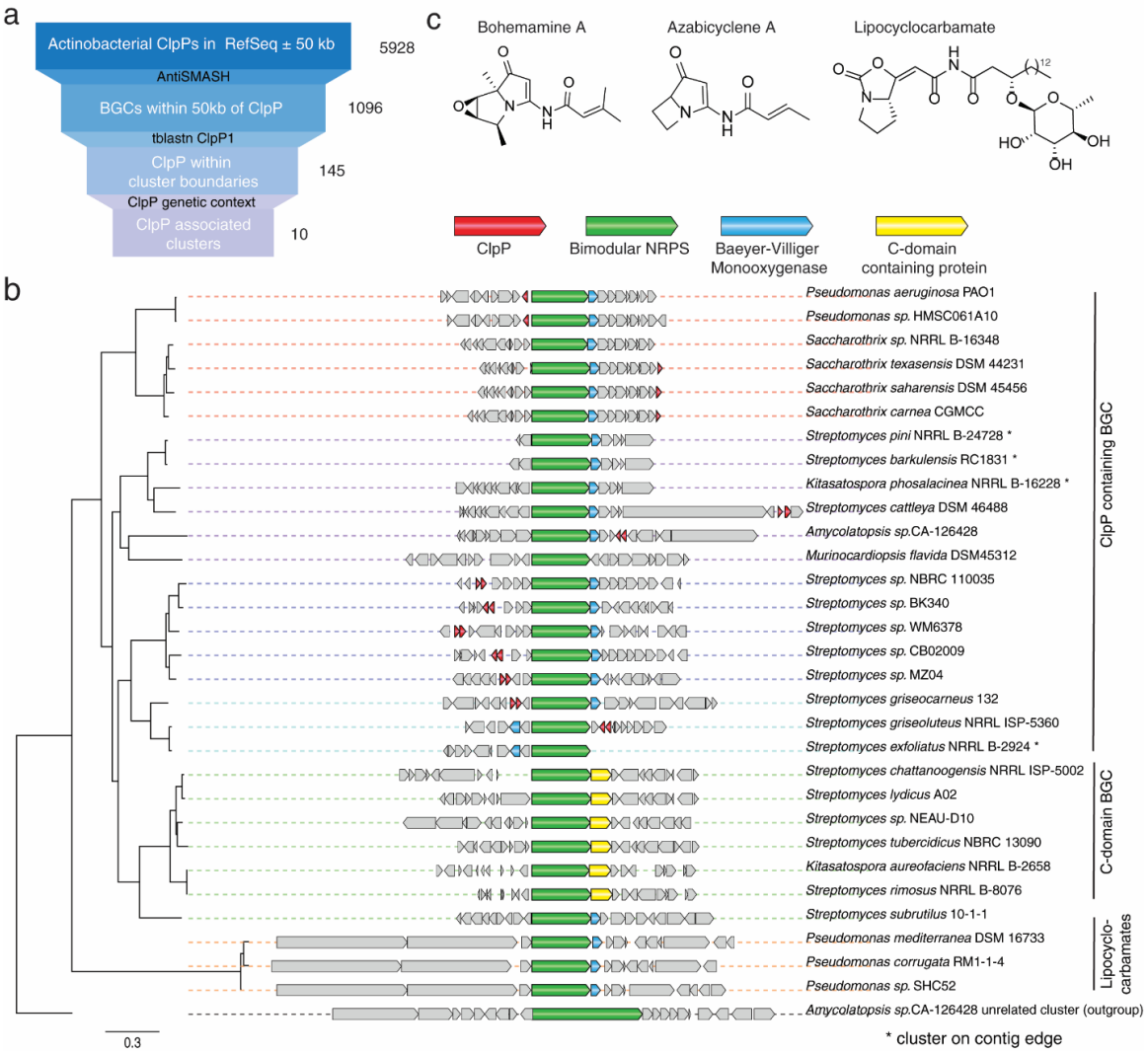


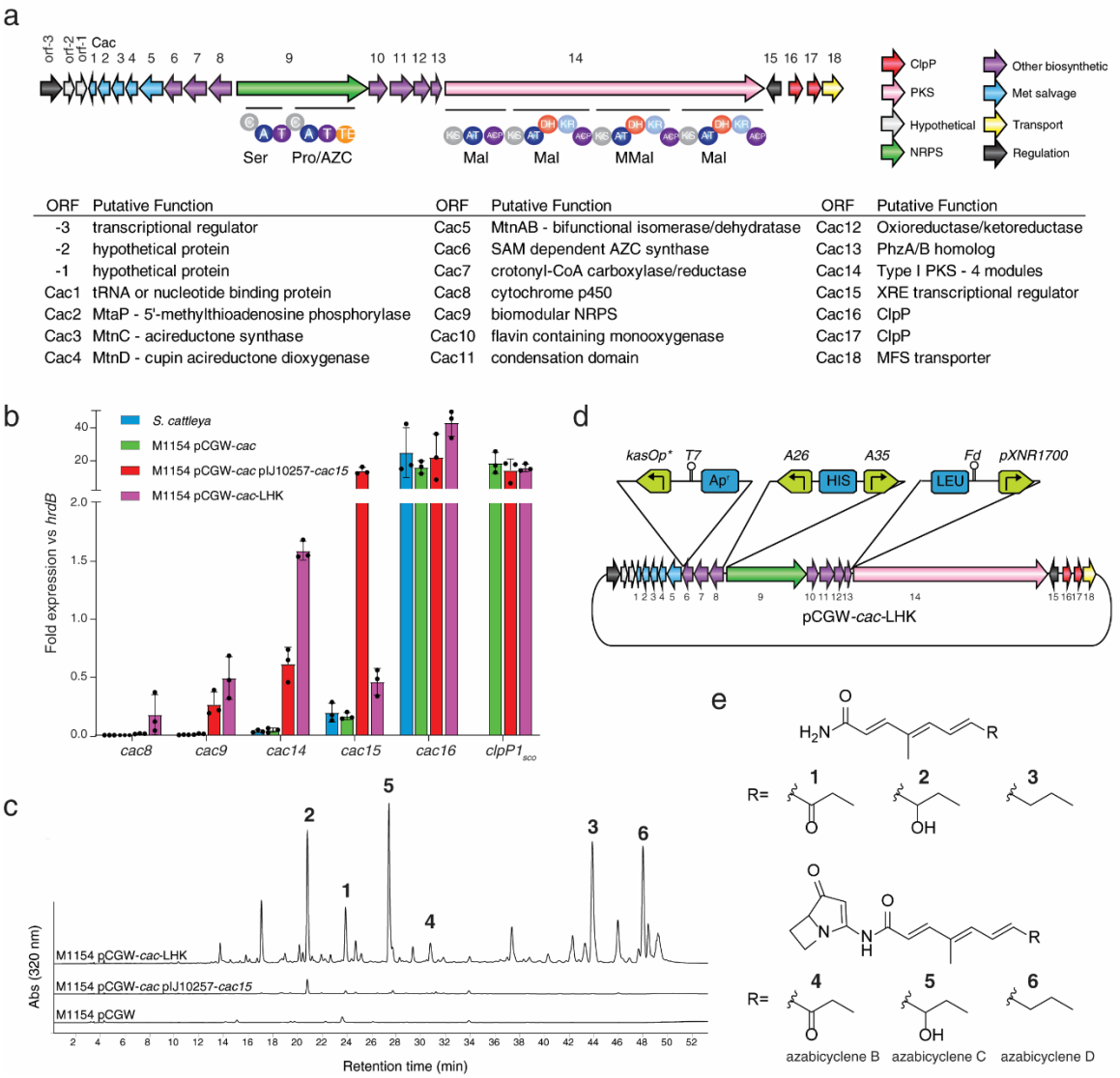
Figure 1. Genome mining for ClpP associated BGCs.
(A) Summary of target-directed genome mining approach used to identify ClpP-associated clusters from genomes available in GenBank. Numbers to the right of the workflow indicate the number of hits at each step. (B) Phylogenetic tree and cluster alignment built using CORASON[26] on 30 clusters identified containing a bimodular NRPS. An unrelated BGC identified in *Amycolatopsis* sp. CA-126428 was used to root the tree. Three key groups of clusters are labelled on the right, and characteristic genes are coloured as indicated. (C) Structures of three previously identified metabolites produced by biomodular NRPS containing clusters. Bohemamine A is produced by *S.* sp. CB02009[29], azabicyclene A is produced by *P. aeruginosa* PAO1[31], and lipocyclocarbamate is produced by *Pseudomonas* sp. SHC52[28], though other strains in (B) with homologous BGCs may also produce these compounds.

Interestingly, six of our ten final prioritized BGCs contained a bimodular non-ribosomal peptide synthetase (NRPS) predicted to activate a proline and serine. Using this bimodular NRPS as a BLAST query, we collected several more related BGCs from GenBank, many of which apparently lacked associated ClpPs (Supplementary Table 2). To visualize the relationship between these clusters and their association with ClpP, we used the tool CORASON (CORe Analysis of Syntenic Orthologs to prioritize Natural product biosynthetic gene clusters)[26]. Given a query gene, in this case the bimodular NRPS, CORASON identifies nearby homologous genes among a set of BGCs and uses them to build a multi-locus phylogeny. This tool effectively revealed three distinct groups among our collection of BGCs (Fig. 1b). In the first group, the bimodular NRPS is adjacent to a Baeyer-Villiger flavin-containing monooxygenase (FMO) in a Pseudomonal strain, with members of this group encoding the known phospholipase inhibitor lipocyclocarbamate (also called brabantamide, Fig. 1c)[27,28]. In the second group, the bimodular NRPS is adjacent to a condensation domain containing protein, with no product having been characterized. The third group contains BGCs all associated with an FMO and ClpP, with members responsible for the production of a family of bacterial pyrrolizidine alkaloids (Fig. 1c). The biosynthesis of some of these compounds has been characterized such as for the bohemamines produced by *Streptomyces sp.* CB02009[29] and azabicyclenes (also called azetidomonamide B) produced by *Pseudomonas aeruginosa* PAO1[30,31]. However, to the best of our knowledge, the association of these BGC's with ClpP has not been previously noted. We therefore set out to further explore this widespread family of clusters and investigate their association with ClpP.

Heterologous expression of a ClpP associated cluster from *Streptomyces cattleya*

In exploring the unknown members of this family of ClpP-associated BGCs, we became interested in a cluster from *Streptomyces cattleya* DSM 46488 that in addition to the bimodular NRPS, also contains a type I PKS (Fig. 2a, Supplementary Table 3). We named the cluster *cac* for ClpP Associated Cluster, and homologous clusters in *Streptomyces pini* PL19 NRRL B-24728, *Streptomyces barkulensis* RC1831 and *Kitasatospora phosalacinea* NRRL B-16228 were used to define the cluster borders (Supplementary Fig. 1). While *S. cattleya* is a highly studied organism with at least seven characterized secondary metabolites, the product of *cac* has never been described. Indeed, *cac* was transcriptionally silent under several culture conditions tested (Fig. 2b). We found *S. cattleya* difficult to genetically manipulate and so chose to use heterologous expression as a strategy to activate *cac*. Using transformation-assisted recombination (TAR) cloning[21] to directly capture the 44 kb cluster (pCGW-*cac*), we shuttled it into our heterologous host *S. coelicolor* M1154.

To transcriptionally activate the cluster, we first attempted overexpression of the cluster situated XRE family transcriptional regulator, Cac15. While overexpression of *cac15* activated several transcriptional units in *cac*, it only weakly activated the putative operon from *cac1-8* (Fig. 2b). Correspondingly, we did not observe production of any new metabolites from this engineered strain (Fig. 2c). Therefore, we chose to refactor *cac* by inserting non-native promoters in front of four putative transcriptional units (Fig. 2d pCGW-*cac*-LHK). By successfully activating the full BGC, this refactored construct supported the production of numerous novel metabolites, as detected by HPLC (Fig. 2c).



ORF	Putative Function	ORF	Putative Function	ORF	Putative Function
-3	transcriptional regulator	Cac5	MtnAB - bifunctional isomerase/dehydratase	Cac12	Oxioreductase/ketoreductase
-2	hypothetical protein	Cac6	SAM dependent AZC synthase	Cac13	PhzA/B homolog
-1	hypothetical protein	Cac7	crotonyl-CoA carboxylase/reductase	Cac14	Type I PKS - 4 modules
Cac1	tRNA or nucleotide binding protein	Cac8	cytochrome p450	Cac15	XRE transcriptional regulator
Cac2	MtaP - 5'-methylthioadenosine phosphorylase	Cac9	biomodular NRPS	Cac16	ClpP
Cac3	MtnC - acireductone synthase	Cac10	flavin containing monooxygenase	Cac17	ClpP
Cac4	MtnD - cupin acireductone dioxygenase	Cac11	condensation domain	Cac18	MFS transporter

Figure 2. Heterologous expression of a ClpP associated cluster from *S. cattleya.* Schematic of *cac* identified in *S. cattleya* DSM 46488 and putative ORF annotations. (B) RT-PCR results compare the expression of *cac* in *S. cattleya* and heterologous producer *S. coelicolor* M1154 with various engineered constructs. RT-PCR was performed in biological triplicate and technical duplicate and normalized to the housekeeping sigma factor *hrdB.* Bars represent the mean with error bars showing standard deviation and dots individual measurements from biological replicates. Note the change in scale for the broken Y axis. (c) HPLC chromatogram of butanolic extracts from *S. coelicolor* M1154 strains carrying the *cac* cluster shows production of novel metabolites supported by the refactored construct pCGW-*cac*-LHK. Fermentations were performed in triplicate and a representative chromatogram is shown. (D) Schematic shows the design of the refactored construct pCGW-*cac*-LHK. (E) Structure of azabicyclenes family compounds or hydrolyzed acyl tails, with numbers corresponding to peaks shown in (C).

We purified and solved the structure of major peaks by NMR and/or LC-MS/MS (Supplementary Figures 2-22, Supplementary Tables 4-7). The major product of the BGC was a novel azabicyclene with a hydroxylated decatriene acyl tail (Fig. 2e, **5**). We also identified congeners lacking the hydroxy group (Fig. 2e, **6**), oxidized to a ketone (Fig. 2e, **4**), or hydrolyzed products of their respective acyl tails (Fig. 2e, compounds **1-3**). We named these compounds azabicyclenes B (**4**), C (**5**) and D (**6**), following the nomenclature of the compound azabicyclene (which we will hereto refer to as azabicyclene A) produced by the *aze* BGC in *P. aeruginosa* PAO1[30,31].

The biosynthesis azabicyclenes B-D is predicted to be similar to bohemamines and azabicyclene A[29,31], with the bimodular NRPS Cac9 and FMO Cac10 being essential for generation of the bicyclic core (Fig. 3). First, Cac9 selects L-azetidine-2-carboxylic acid (L-AZC) and L-serine to generate a dipeptide. These unique bimodular NRPSs dehydrate L-serine to dehydroalanine, then release and cyclize the product to generate a predicted 4,6-bicyclic intermediate[27–29]. The free tautomeric quinone would then be acted upon by the characteristic FMO Cac10, which catalyzes a Baeyer-Villiger oxidative ring expansion and rearrangement to the final azabicyclene scaffold[27,28]. Notably, Cac1-5 encodes a methionine salvage pathway for the regeneration of S-adenosylmethionine (SAM), which is used as a substrate for the AZC synthase Cac6, homologous to AzeJ in the *aze* BGC.

Biosynthesis of the acyl tail matches the prediction from the type I PKS Cac14 using ethylmalonyl-CoA as a starter unit, which could be produced by the crotonyl-CoA carboxylase/reductase Cac7 (Fig. 3). Interestingly, azabicyclene C is modified at the

alpha rather than beta position relative to the conjugated unsaturated system, meaning that the hydroxy group must not be part of the starter unit accepted by the PKS assembly line. Instead, we predict that Cac8, a cytochrome p450, is likely responsible for hydroxylating azabicyclene D and that this hydroxy group becomes further oxidized in azabicyclene B.

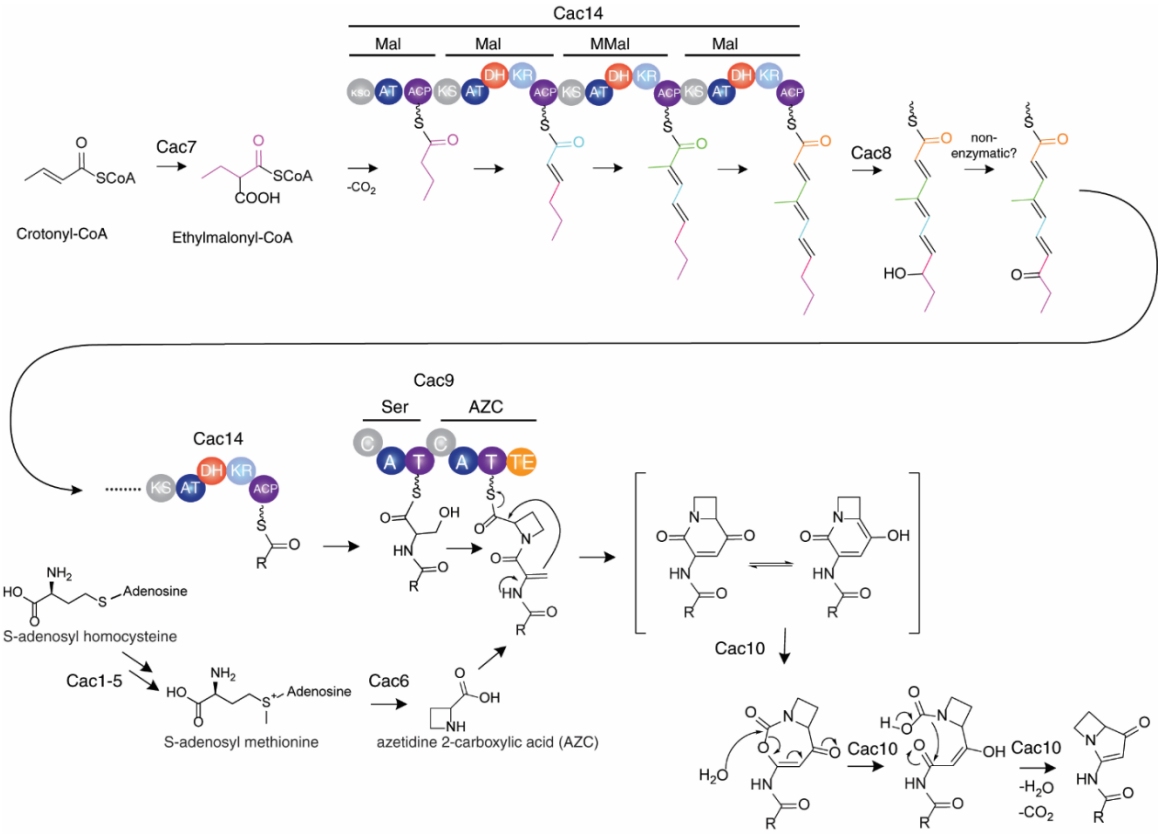


Figure 3. Proposed biosynthesis of azabicyclenes.

Cluster-associated ClpPs are functional proteases

Having identified the novel metabolites produced by *cac,* we next sought to investigate the function of the cluster-associated ClpPs. All the ClpP-associated clusters that were identified in *Streptomyces* spp. encode two *clpP* paralogs in an operon, which we refer to as *cac16* and *cac17*. This is the commonly observed arrangement for housekeeping ClpPs, ClpP1 and ClpP2, in actinobacteria. Comparing these ClpP homologs in *S. cattleya,* $Cac16_{scatt}$ shares 70% amino acid identity with $ClpP1_{scatt}$, and $Cac17_{scatt}$ shares 66% with $ClpP2_{scatt}$. Phylogenetic analysis of ClpP homologs in a collection of *Streptomyces* genomes shows that cluster-associated *clpPs* are paralogs of the main housekeeping ClpPs in *Streptomyces*, *clpP1* and *clpP2,* respectively, rather than secondary copies such as *clpP3, clpP4* and *clpP5* (Fig. 4a). Furthermore, the binding site motif for ClgR, a transcriptional activator controlling the *clpP1* operon, is also found in the *cac16* promoter region (Supplementary Fig. 23)[32]. These observations suggest that Cac16 and Cac17 might form a functional proteolytic complex capable of complementing the function of ClpP1/P2.

To further characterize the function of cluster-associated ClpPs, we reconstituted the activity of *S. cattleya* ClpP1/P2 and Cac16/17 proteolytic complexes *in vitro*. The structure and function of asymmetric hetero ClpP complexes *in vitro* has been previously shown for several species including *Mycobacteria smegmatis* and *M. tuberculosis*[33,34], but never a member of the family Actinomycetaceae. Mycobacterial ClpP activity *in vitro* requires both ClpP isoforms and is activated synergistically by a combination of agonist peptides, such as benzyloxycarbonyl-leu-leucinal, and interaction with its cognate AAA+

ATPase or mimic, i.e. ADEP[33]. Similarly, we observed that while individual subunits had no activity, mixtures of ClpP1$_{scatt}$/P2 $_{scatt}$ or Cac16/17 with ADEP were able to cleave short peptide substrates (Fig. 4b, Supplementary Table 8). ADEP was required for activity, while a variety of peptide agonists had no stimulatory effect on ClpP1$_{scatt}$/P2 $_{scatt}$ and in fact inhibited Cac16/17.

ClpP proteins are produced as propeptides and require N-terminal processing to form active peptidases[34], and has been observed natively in *Streptomyces lividans*[35]. Using whole protein MS, we measured and predicted N terminal processing sites of *S. cattleya* ClpPs (Supplementary Table 9). We observed no evidence of ClpP1$_{scatt}$ processing, while ClpP2$_{scatt}$ was cleaved only in the presence of ClpP1$_{scatt}$ or Cac16, indicating cross-processing. Similarly, we observed cross-processing of cluster-associated ClpPs, this time of both Cac16, by either ClpP2$_{scatt}$ or Cac17, and Cac17, by either ClpP1$_{scatt}$ or Cac16. Processing did not require the presence of agonist peptide or ADEP. These results correspond closely with *M. smegmatis* ClpP characterization, where cross-processing of ClpP2 but not ClpP1 occurs, and does not require additional stimulatory agents[34].

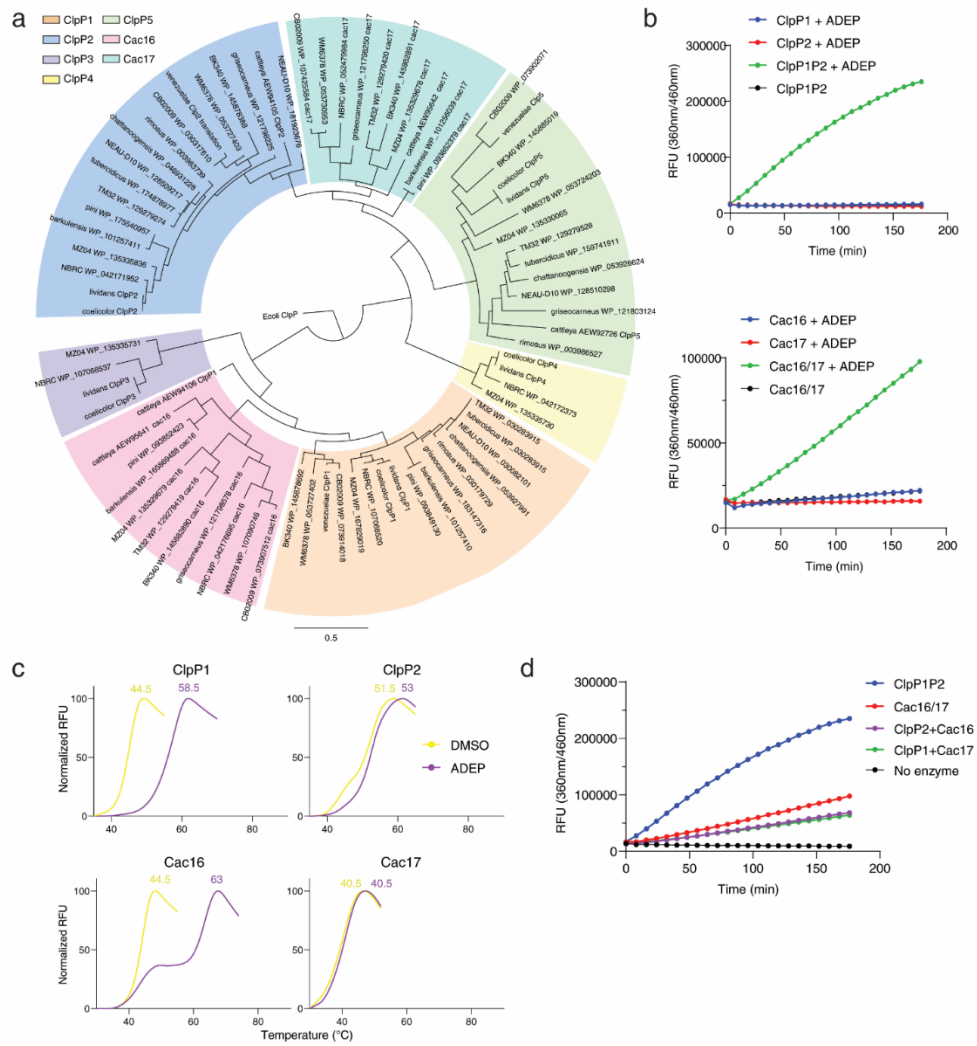


Figure 4. Cac16/17 forms a functional proteolytic complex.
(A) Phylogenetic tree of *Streptomyces* ClpPs from 15 strains containing a ClpP associated cluster and 3 model species. Node labels contain the strain/species identifier and GenBank protein accession or ClpP label, if known. Cluster-associated ClpPs homologous to Cac16 and Cac17 are labelled respectively. Colours represent cladal organization of different ClpP paralogs, noting the neighbouring relationships between ClpP1 and Cac16, and ClpP2 and Cac17. *E. coli* ClpP is used as an outgroup. (B) $ClpP1_{scatt}/P2_{scatt}$ (top) and Cac16/17 (bottom) are active heterocomplexes *in vitro* using Suc-LLVY-AMC as a substrate. (C) Thermal shift assays of *S. cattleya* ClpPs in the presence or absence of ADEP. Numbers indicate melting temperature. Assays were performed in technical duplicate with representative curve shown. (D) Protease assays, as in panel B, show formation of functional $ClpP1_{scatt}$/Cac17 and Cac16/$ClpP2_{scatt}$ complexes *in vitro.* For panels B and D, mean of technical triplicate reactions with error bars representing standard deviation (not visible at this scale) are shown.

To further compare the function and regulation of *S. cattleya* housekeeping and cluster-associated ClpPs, we measured ADEP binding to each subunit using thermal shift assays (Fig. 4c). ADEP induces a dramatic conformational change in ClpP heptamers/tetradecamers that stabilizes the protein and results in a 10-20°C shift in melting temperature[36]. ADEP was shown to bind and stabilize $ClpP1_{scatt}$ and Cac16, but not $ClpP2_{scatt}$ or Cac17. This contrasts with Mycobacterial ClpPs, where the cognate AAA+ ATPase, or ADEP as a mimic, interacts only with the ClpP2 subunit[34,37]. In the context of *S. cattleya,* these results suggest analogous function of $ClpP1_{scatt}$/$ClpP2_{scatt}$ and Cac16/Cac17 complexes.

Finally, we tested whether Cac16/$ClpP2_{scatt}$ and $ClpP1_{scatt}$/Cac17 could form functional mixed complexes. Indeed, we observed that these mixed complexes were capable of proteolysis (Fig. 4d). Notably, Cac16/Cac17 complexes were ~12x less active than $ClpP1_{scatt}$/$P2_{scatt}$ *in vitro*, and so correspondingly Cac16/$ClpP2_{scatt}$ and $ClpP1_{scatt}$/Cac17 complexes were also less active than $ClpP1_{scatt}$/$P2_{scatt}$ alone. Whether the reduced Cac16/17 activity is physiologically relevant or an artefact of an artificial *in vitro* system is unknown. Nonetheless, our ability to successfully demonstrate that Cac16 and Cac17 form functional and analogous complexes to ClpP1 and ClpP2 in terms of peptide cleavage, propeptide processing and ADEP binding suggests that they may be able to functionally complement housekeeping functions *in vivo.*

Characterization of activity of azabicyclenes

We next investigated the effect of azabicyclenes on the function of ClpP *in vitro* and *in vivo.* Azabicyclene C has no growth inhibitory activity against a range of bacterial

and mammalian cells tested, while azabicyclene D is weakly cytotoxic (IC_{50} = 60 µM) and antibacterial (MIC = 64-128 µg/mL)(Fig. 5a). Co-overexpression of *cac16* and *cac17* in their native operon in *S. coelicolor* or *Streptomyces venezuelae* did not change susceptibility to azabicyclene D, and since it is also able to inhibit growth of *Bacillus subtilis* Δ*clpP,* the mechanism of this toxicity is likely unrelated to ClpP. Low yield of azabicyclene B prevented us from testing its activity.

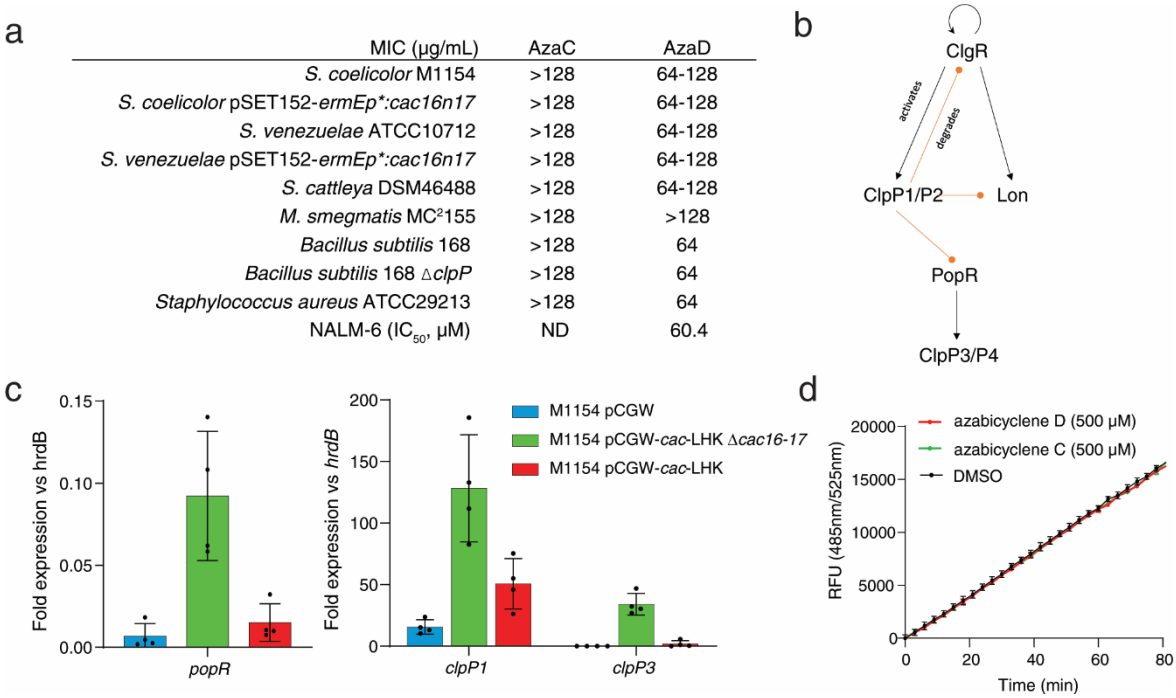


Figure 5. Biological effects of azabicyclenes.
(A) Susceptibility testing of various bacteria and mammalian cells against azabicyclene C and D. (B) Schematic of ClpP regulation by the ClgR regulon in *S. coelicolor.* ClgR is a transcriptional activator that regulates the *clpP1clpP2* operon, *lon* as well as its own expression[32]. Subsequently, ClpP1/P2/X degrades Lon, ClgR, as well as the transcriptional activator for the *clpP3clpP4* operon, PopR. (C) Transcriptional effects of *cac16-17* deletion from *cac* in the heterologous producer *S. coelicolor* M1154, as measured by RT-PCR. Bars represent the mean with error bars showing standard deviation and dots show individual measurements from biological replicates. (D) Effect of azabicyclene on *in vitro* function of ClpP1 $_{scatt}$/P2$_{scatt}$. ClpP1/P2 activity was reconstituted using ADEP and followed by cleavage of (S-LLVY)2-rhodamine 110. Mean of triplicate reactions is shown with error bars representing standard deviation.

Despite not being traditional antibiotics, we wondered whether azabicyclenes might alter the function of ClpPs to remodel the proteome in a way that is not necessarily lethal. To this end, we deleted *cac16-cac17* from our construct for heterologous expression and measured evidence of ClpP disruption through transcriptional profiling. By using our heterologous producer directly, we could measure the combined effect of all azabicyclenes and avoid missing potentially active derivatives in this preliminary experiment.

We performed RT-PCR on genes whose expression is related to ClpP1/P2 function (Fig. 5b) in our heterologous producer, comparing *S. coelicolor* harbouring empty vector (pCGW), the refactored *cac* construct (pCGW-*cac*-LHK), or the refactored *cac* construct missing *cac16-17* (pCGW-*cac*-LHK-Δ*cac16-17*). In comparison to *S. coelicolor* pCGW and pCGW-*cac*-LHK, deletion of *cac16-17* resulted in 2.5-fold upregulation of *clpP1,* 6-fold upregulation of *popR* and strikingly 19-fold upregulation of *clpP3* (Fig. 5c). In *S. coelicolor*, there are five ClpP isoforms and ClpP1/P2 is the main housekeeping complex. In conjunction with ClpX, it targets a number of characterized substrates for degradation including the Lon protease, its own transcriptional activator ClgR, and the activator of *clpP3clpP4,* PopR (Fig. 5b). Since ClpP1/P2/X normally degrades PopR, the *clpP3clpP4* operon is not expressed, but can be activated upon ClpP1P2 disruption resulting accumulation of PopR[35]. Therefore, induction of *clpP1, popR* and *clpP3* provide evidence for ClpP1/P2 disruption in *S. coelicolor* pCAP-*cac*-LHK-Δ*cac16-17.* The lack of induction in the pCAP-*cac*-LHK strain harbouring *cac16-cac17* suggests that Cac16/17 may form a functional complex resistant to disruption.

Finally, we tested whether we could detect *in vitro* inhibition of ClpPs using our reconstituted ClpP1$_{scatt}$/P2$_{scatt}$ or Cac16/17 complexes (Fig. 5d). However, purified azabicyclene C and azabicyclene D had no effect. Notably, bohemamines, which we purified from *Streptomyces sp.* NBRC110035, also had no effect on ClpP *in vitro* (results not shown). Instead, it is possible that azabicyclenes act indirectly on ClpP function, for example through their cognate AAA+ ATPases, in a way not detectable in our assay utilizing a short peptide substrate and ADEP as an AAA+ ATPase mimic. Further studies are required to investigate these hypotheses.

DISCUSSION

By applying target-directed genome mining, here we describe the association of a widespread family of BGCs encoding pyrrolizidine alkaloids with the highly conserved and often essential protease ClpP. While our target-directed genome mining approach is similar to the idea employed by the Antibiotic Resistant Target Seeker (ARTS)[22], the later focuses on a single genome/group of genomes, and using a reference 'core' genome, identifies any duplicated essential gene within these genomes' BGCs. Conversely, our approach focuses on a single target, ClpP, and searches all genomes in RefSeq. Furthermore, given the multiple and variable number of *clpP* copies per genome, specific knowledge of genetic context was vital to differentiating housekeeping from cluster-associated paralogs. High redundancy is a common feature of actinomycete genomes and should be kept in mind when designing future pipelines for detecting target duplication within BGCs during genome mining.

The biosynthesis of several pyrrolizidine alkaloids, including lipocyclocarbamates[27,28], bohemamines[29], legonmycins[38] and azabicyclene A[30,31], have been described previously, though many questions still remain. For example, the enzymatic basis that results in biosynthesis of the differing 5,5 ring systems between lipocyclocarbamates and ClpP-associated BGCs in this family remains to be determined. Swapping the FMO alone, which is putatively responsible for reactions catalyzed at the branch point of these two pathways, is not sufficient to redirect biosynthesis towards a lipocyclocarbamate type scaffold[39]. In the case of the biosynthesis of azabicyclenes described in this study, the putative role of Cac8 in hydroxylation of the acyl tail remains to be confirmed, while the roles of Cac11 and Cac12, which are conserved in related clusters, and Cac13, which is unique to *cac,* remain elusive.

Given the widespread nature of these bacterial pyrrolizidine alkaloids, from *Streptomyces* to *Xenorhabdus* to *Pseudomonas*, and their conserved association with ClpP, these metabolites undoubtedly play an important role in bacterial physiology. Few distinctive phenotypes have been previously noted for these compounds, though some are cytotoxic against chromic myeloid leukaemic cells[40], while activation of azabicyclene biosynthesis in *P. aeruginosa* decreases virulence[30]. We also note the lack of a strong phenotype produced by our novel azabicyclenes, but by probing the ClpP-associated regulatory networks, provide the first evidence that they disrupt ClpP1/P2 function and that cluster-associated ClpPs reverse this disruption. Through future studies to probe this phenotype, we hope to uncover the details of azabicyclene function. It will also be interesting to see if Cac16/17 can fully complement the function of ClpP1/P2, which

includes characterization of their association with cognate AAA+ ATPases as well as substrate specificity.

Our results suggest that ClpP disruption in *Streptomyces* spp. by azabicyclenes is not lethal to cells. The reason for ClpP essentiality in actinobacteria is not fully understood, and at least in *M. tuberculosis* likely has to do with the essential degradation of multiple proteins including the regulator WhiB1, targeted by ClpC1, and *ssrA*-tagged proteins, targeted by ClpX[41]. In addition to the essential genes *clpC1* and *clpX, Streptomyces* spp. also usually encode two other ClpC proteins that have not been characterized[32]. It is possible that altering association with these various AAA+ ATPases could remodel the proteome without being lethal. Alternatively, it has been shown in *M. smegmatis* that ClpP1 and ClpP2 isoforms have overlapping but distinct substrate preferences, and so could perhaps provide another level of regulation[34]. The intricate control of Clp proteolytic complexes through conformational changes and association with cognate proteins leaves several avenues within the realm of possibility.

Numerous antibiotics remodel the transcriptional and translational landscape of the cell at sub-MIC levels[42]. Selective inhibition of a subset of genes or proteins can even be achieved by antibiotics that directly block RNA polymerase (e.g. rifamycin[43]) or the ribosome (e.g. macrolides[44]). Meanwhile, non-antimicrobial signalling molecules interact with global transcriptional regulators to affect change (e.g. γ-butyrolactones in *Streptomyces*[45], acyl-homoserine-lactones in *Pseudomonas*[46]). These are just a few examples of how secondary metabolites are appreciated to be regulators of transcription and translation, and play an important role during infection or survival in the

environment. Post-translational regulation is likely to be no less important in coordinating cellular activities, and so it should be expected that specialized metabolites have also evolved to modulate this level of regulation. Fully elucidating the function of azabicyclenes and related compounds may be the first step towards understanding post-translational control of the proteome through ClpP.

METHODS

Strains and Culture Conditions

All primers, gBlocks and strains used in this study are listed in Supplementary Tables 10-12, respectively. *Streptomyces* strains were routinely grown at 30°C, 250 rpm, in Tryptone-Soya-Broth (TSB; BD Biosciences) for starter cultures and genomic DNA preparation, SM-$MgCl_2$ (2% each D-mannitol, soya flour, agar, 10 mM $MgCl_2$) for sporulation and conjugation, and Bennett's media (1% potato starch, 0.2% casamino acids, 0.18% yeast extract, 0.02% KCl, 0.02% $MgSO_4 \cdot 7H_2O$, 0.024% $NaNO_3$, 0.0004% $FeSO_4 \cdot 7H_2O$) for fermentation. Cloning was performed in *Escherichia coli* Top10 using standard conditions and routinely grown in LB media (Bioshop).

All pCGW derived constructs were introduced into *S. coelicolor* M1154 via triparental conjugation using *E. coli* ET12567/pCGW-XXX and *E. coli* ET12567/pR9406 using standard protocols[47]. pSET152-*ermE*p*:*cac16-17* and pIJ10257-*cac15* were introduced via biparental conjugation using *E. coli* ET12567/pUZ8002 *neo::bla.* Antibiotics were supplemented as necessary (50 µg/mL kanamycin, 100 µg/mL ampicillin, 50 µg/mL apramycin, 35 µg/mL chloramphenicol, 25 µg/mL nalidixic acid, 100 µg/mL or 50 µg/mL hygromycin B for *E. coli* and *Streptomyces* spp., respectively).

MICs were assessed using standard microbroth dilution protocols. *Streptomyces* spp. were assayed in TSB, *M. smegmatis* in 7H9 media and all other strains in Mueller-Hinton Broth (BD Biosciences).

In silico detection and analysis of ClpP associated clusters

First, we collected all ClpP homologs from Actinobacteria in PFAM (PF00574) in order to assemble a complete and diverse set of relevant ClpP-like proteins. This collection of ClpP protein sequences were used to query prokaryotic RefSeq proteins using blastp, and the top five hits for each query were taken. Proteins were then mapped to RefSeq genomes with the identical protein report function. Using a custom python script, 50 kb upstream and downstream of each *clpP* hit was collected and BGCs were identified using AntiSMASH[24]. Further filtering was performed in Geneious V8.9.1. Clusters containing ClpP homologs were identified by tblastn of *S. coelicolor* ClpP1 (NP_626855.1). Housekeeping ClpPs were identified by association with the following proteins using tblastn: ClpX (NP_626853.1), PopR (NP_631335.1, WP_067791847.1, WP_028564420.1), ClpP5-associated hypothetical protein (NP_625527.1). Eight BGCs where *clpP* was located on the contig edge were also discounted.

27 BGCs containing the biomodular NRPS for CORASON analysis were collected from GenBank by blastp using $Cac9_{scatt}$ as a query and an arbitrary amino acid identity cut-off of 54% (Supplementary Table 2). 11/17 ClpP-associated BGCs collected in this way were either uploaded to GenBank after our initial genome mining or were fragmented, explaining why they were not identified by our previous analysis. Five *Pseudomonas* spp. BGCs with amino acid identity <54% were also included for

comparison. CORASON was run on contigs containing these 30 BGC using Cac9 (NRPS) as the query gene, *cac* from *S. cattleya* as the reference BGC, and a bit score cut-off of 1000.

Bioinformatic analysis of ClpPs

For phylogenetic analysis of ClpP, homologs were identified in the genomes of *Streptomyces* strains by blastp. A protein alignment was generated using MUSCLE and used to construct the phylogenetic tree with FastTree 2.1.11. *E. coli* ClpP was used to root the tree.

Cloning cac using TAR

The pCAP03 capture vector was a gift from Dr. Bradley Moore's Lab[21], and modified to use the 'oriV-ori2-repE-sopABC' single copy origin of replication from pBAC-lacZ to generate pCGW[48]. pCGW was maintained in *E. coli* EPI300, which controls the expression of *trfA* required for high copy amplification from *oriV* with an arabinose inducible promoter. When necessary for miniprep and plasmid mapping, high copy number was induced using 1 mM arabinose. The vector backbone contains all the necessary elements for propagation and selection in *E. coli* (pUC *ori* and Kan^R), yeast (ARSH4/CEN6 and TRP1 auxotrophic marker) and *Streptomyces* (φC31 integration, *oriT* and Kan^R).

50 bp homology arms flanking *cac* (*orf-3* to *cac18*) were concatenated with an MssI site in between, and 18bp overlaps with the pCGW backbone were added to either end. The gBlock (Integrated DNA Technologies; IDT) was cloned into pCGW linearized

with NdeI/XhoI by Gibson assembly (Supplementary Table 11). Prior to TAR cloning, capture vector was digested with MssI to release the hooks.

To prepare HMW genomic DNA for TAR, *S. cattleya* DSM 46488 was grown in TSB to midlog phase. Cells were treated with lysozyme, followed by lysis with Proteinase K and 1% SDS. Phenol/chloroform and ethanol precipitation was used to extract and clean up the DNA. The resulting nucleic acid pellet was RNaseA treated and again ethanol precipitated. Approximately 15 μg of purified genomic DNA was digested with FastDigest restriction enzymes that would cut surrounding, but not within, the gene cluster of interest (XhoI and NdeI). The resulting DNA was again cleaned by ethanol precipitation and dissolved in TE buffer.

Saccharomyces cerevisiae VL648N sphereoplasting and transformation was carried out as previously described[49]. 22 colonies were restreaked on SD-Trp plates and screened by colony PCR for the desired insertion using diagnostic primers located inside *cac,* giving 1 positive colony (Supplementary Table 10). DNA was extracted from the positive yeast colony by zymolyase treatment followed by phenol/chloroform extraction and ethanol precipitation. This DNA was transformed into *E. coli* EPI300 cells by electroporation and plasmids were selected with kanamycin. *E. coli* EPI300 was induced for high copy number using 1 mM arabinose, and the resulting construct, pCGW-*cac,* was confirmed by restriction mapping.

Refactoring pCGW-cac

pCGW-*cac* was refactored using a combination of yeast recombination and λ-red recombineering in *E. coli.* All primers and plasmids are listed in Supplementary Tables

10 and 12. First, Leu2 and His3 auxotrophic selection cassettes were constructed to contain orthogonal promoters and terminators using pSASS series plasmids, a gift from Dr. Mike Tyers (University of Montreal). The *leu2* open reading frame was PCR amplified from pRS316 and inserted into pSASS5, and *his3* was PCR amplified from pYAC10 (Mike Tyers) and inserted into pSASS4, using Gibson assembly. Selection cassettes containing the marker along with yeast promoter and terminator sequences were subsequently PCR amplified from pSASS5-*leu2* and pSASS4-*his3.* For replacement of the *cac14* promoter, XNR_1700p was PCR amplified from *Streptomyces albus* gDNA[50], and stitched to the *leu2* selection cassette with 500 bp homology arms to *cac* on either side through overlap extension PCR. The resulting refactoring cassette was co-transformed with pCGW-*cac* into *Saccharomyces cerevisiae* SASy31/SASy35 by standard lithium acetate/single-stranded carrier DNA/PEG mediated transformation[51]. Recombinants were selected on SD-trp-leu plates, screened by colony PCR, and the resulting construct, pCGW-*cac*-L was recovered and confirmed in *E. coli* EPI300 similar to initial TAR cloning. Replacement of *cac8* and *cac9* promoters were achieved similarly using the His3 selection marker sandwiched by synthetic *Streptomyces* promoters, A26 and A35, incorporated into primer sequences[52]. This time, *S. cerevisiae* SASy31 transformants were selected using SD-trp-leu-his plates, and the resulting construct, pCGW-*cac*-LH, was moved to *E. coli* EPI300.

After refactoring *cac8, cac9* and *cac14,* RT-PCR revealed that *cac5* and the putative operon including *cac4, cac3, cac2* and *cac1* was still lowly transcribed (data not shown). Since there were few selection markers left for use in *S. cerevisiae,* we chose to

use *E. coli* λ-Red recombineering. The promoter *kasO*p* and T7 terminator was designed adjacent to the apramycin resistance gene *aac*(3)IV flanked by PmeI/HpaI restriction sites to allow for removal of *aac*(3)IV after recombination. 39 bp homology arms were added to each end to direct recombination. The recombineering cassette was synthesized as a gBlock by IDT and recombineering was carried out using standard protocols[53]. pCGW-*cac*-LH was transformed into *E. coli* BW25113/pKD46, selected using kanamycin and ampicillin, and grown at 30°C to maintain pKD46. The strain was grown in SOB without $MgSO_4$ overnight at 30°C, then subcultured with the addition of 10 mM arabinose to induce expression of *red* genes from pKD46. After reaching OD_{600} = 0.6, the cells were recovered by centrifugation, washed twice with ice-cold 10% glycerol, and electroporated with 1 μg of linear refactoring cassette. Successfully recombinants were selected with apramycin and grown at 37°C to promote loss of pKD46. The resulting construct, pCGW-*cac*-LHK, was extracted from *E. coli* BW25113, transformed into *E. coli* EPI300, and verified by restriction mapping. Finally, *aac*(3)IV was removed by digestion with HpaI and PmeI followed by intramolecular blunt end ligation with T4 DNA ligase, generating pCGW-*cac*-LHK-hpa.

Generation of pCGW-cac-Δcac16-17

An in-frame deletion of *cac16-cac17,* leaving downstream *cac18* intact, was created using λ-Red recombineering as described above and primers listed in Supplementary Table 10. A refactoring cassette with *aac*(3)IV, HpaI/PmeI restriction sites and 40 bp homology arms was PCR amplified and transformed into *E. coli* BW25113/pKD46/pCGW-*cac*-LHK-hpa. The *aac*(3)IV cassette was again removed by

HpaI and PmeI digestion followed by blunt end ligation, generating pCGW-*cac*-LHK-Δ*cac16-17.*

Cac15 and Cac16-17 overexpression in Streptomyces

Coding sequences were PCR amplified from *S. cattleya* gDNA using primers in Supplementary Table 10. To overexpress the putative transcriptional activator *cac15,* the gene was cloned into pIJ10257 harbouring the strong constitutive promoter *ermE*p* using Gibson assembly. To co-overexpress *cac16-17,* the native operon was cloned downstream of *ermE*p* in a pSET152 backbone using Gibson assembly.

RT-PCR measurement of cac and clpP regulon expression

Strains for analysis were inoculated 1:100 from a saturated TSB seed culture into 60 mL Bennett's media in a 250 mL baffled flask, and pellets were taken at 24 hr. Cells were lysed by bead beating mycelium with 4 mm glass beads in 5 mL TRIzol reagent (Invitrogen), and RNA was extracted using the manufacturer's recommendations. RNA from the resulting aqueous phase was extracted a second time using acid phenol/chloroform, then combined with a half volume of anhydrous ethanol and finally purified using PureLink RNA Mini Kit (Invitrogen). Maxima H Minus First Strand cDNA synthesis kit with dsDNase (Thermo Scientific) was used for cDNA synthesis, and PowerUp SYBR Green master mix (Applied Biosystems) was used for RT-PCR quantification on a BioRad CFX96 real time system. Primers targeting genes of interest (Supplementary Table 10) were designed and 80-100% efficiency was verified prior to quantification. Analysis was performed on three or four independent fermentations and quantified in technical duplicate. Technical duplicates for each biological replicate were

averaged, then fold change expression for each replicate was calculated by normalizing to *hrdB* expression using the ΔCt method.

Fermentation and purification of azabicyclenes C and D

S. coelicolor pCGW-*cac*-LHK seed culture was grown in TSB with 50 μg/mL kanamycin for 3 days, then spent media removed and 50 mL worth of mycelium was inoculated into each of 25x600 mL Bennett's media in 3 L flasks. Fermentations were grown for 7 days at 30°C, 250 rpm. 13 L of spent media was harvested and extracted with 390 g HP-20 resin (Diaion). The resin was washed with 10% methanol (MeOH), then eluted with 100% MeOH and concentrated under vacuum. Dry material was extracted with ethyl acetate:MeOH (1:1) and dried onto 5 g silica gel (Sigma) under vacuum. Normal phase vacuum liquid chromatography was performed using the following stepwise gradient: 1. hexanes, 2. ethyl acetate, 3. ethyl acetate:MeOH (3:1), 4. ethyl acetate:MeOH (1:1). Fractions 2-4 were combined, dried, and using DMSO, applied to an 86 g reverse-phase CombiFlash ISCO (RediSep Rf C18, Teledyne) eluted with a linear gradient system (5-45% water/acetonitrile, 0.1% formic acid). Fractions containing azabicyclenes were pooled and further purified by semipreparative HPLC using a ZORBAX StableBond C18 column (Agilent, 9.4 x 150 mm, 5 μm, 3 mL/min, Buffer A water + 0.1% formic acid, Buffer B acetonitrile + 0.1% formic acid). Compound **2** was purified by the following gradient: 0-3 min 25.5% B, 3-13.5 min 25.5-26.3% B, 13.5-14 min 26.3-95% B, 14-18 min 95% B, 18-19 min 95-25.5% B. Azabicyclene C and Compound 1 were purified by the following gradient: 0-1 min 30% B, 1-3 min 30-38% B, 3-8 min 38-43% B, 8-21 min 43-44.1% B, 21-22 min 41.1-95% B, 22-24 min 95% B, 24-

25 min 95-30% B. Azabicyclene D (11.7 mg) was purified by the following gradient: 0-1 min 13% B, 1-5 min 13-20% B, 5-25 min 20-23.7% B, 25-26min 23.7-95% B, 26-30 min 95% B, 30-31 min 95-13% B. Care was taken to protect compounds from light.

For structural elucidation, compounds **1**, **2**, azabicyclene C (**5**) and azabicyclene D (**6**) were subjected to 1D and 2D NMR and MS/MS. See Supplementary Figures 2-21 and Supplementary Tables 4-7. 1D and 2D NMR experiments were performed using a Bruker AVIII 700 MHz instrument equipped with a cryoprobe in deuterated methanol. Chemical shifts are reported in parts per million relative to tetramethyl silane using the residual solvent signals at 3.31 ppm in proton NMR and 49.0 ppm in carbon NMR as an internal signals. High resolution-electrospray ionization-MS and MS/MS was acquired using an Agilent 6546 LC/Q-TOF in positive ion mode. The structure of azabicyclene B (**4**) was predicted from LC-MS/MS fragmentation (Supplementary Fig. 22).

Expression and purification of ClpPs

S. cattleya ClpP1 (SCATT_17350), ClpP2 (SCATT_17340), Cac16 (SCATT_32700) and Cac17 (SCATT_32710) were codon optimized for expression in *E. coli* and synthesized as gBlocks by IDT (Supplementary Table 11). Synthesized DNA was cloned into pET-28 digested with NcoI and NotI using Gibson assembly to be expressed with a C terminal 6xHis tag. Constructs were transformed into *E. coli* BL21(DE3) 1146D (Δ*clpP::*cm^r). For protein expression, overnight cultures were inoculated into 1 L LB with appropriate antibiotics in a 3 L flask, grown at 37°C until cultures reached OD_{600} = 0.6-0.8, then induced with 0.5 mM IPTG. Cultures were grown for 18 hr at 17°C before harvesting cell pellets by centrifugation.

Cell pellets were resuspended in 20 mL lysis buffer (20 mM Tris (pH 8), 300 mM KCl, 10 mM imidazole, 10% glycerol), treated with 10 mg/mL lysozyme and 5 µg/mL DNase, and lysed on ice by sonication. Clarified lysates were loaded onto equilibrated 2 mL Ni-NTA agarose (Qiagen), washed (20 mM Tris, pH 8, 300 mM KCl, 25 mM imidazole, 10% glycerol) and eluted (20 mM Tris, pH 8, 300 mM KCl, 250 mM imidazole, 10% glycerol). $ClpP1_{scatt}$ and $ClpP2_{scatt}$ were dialyzed into buffer A for further purification. Anion exchange was performed on a HiTrap Q HP 5mL column (GE Healthcare) (Buffer A: 20 mM Tris-HCl, pH 8, 10% glycerol, 1 mM DTT; Buffer B: 20 mM Tris-HCl, pH 8, 10% glycerol, 1 mM DTT, 1 M KCl). Buffer exchange and concentration into storage buffer (50 mM Tris-HCl, pH 8, 10% glycerol, 200 mM KCl) of fractions containing pure protein was performed using Amicon Ultra centrifugal filters with 50 K cutoff. We were unable to purify Cac16 and Cac17 by anion exchange, and so Ni-NTA elutions were dialyzed overnight and concentrated directly into storage buffer for use. Protein purity was >95% as assessed by SDS-PAGE analysis.

In vitro ClpP assays

In vitro reactions were carried out with the following reaction conditions: 2 µM each ClpP subunit (monomer), 50 µM ADEP, 200 µM substrate, ClpP reaction buffer (50 mM HEPES, pH 8, 100 mM KCl, 10% glycerol). Unless otherwise indicated, N-Succinyl-Leu-Leu-Val-Tyr-AMC (S-LLVY-AMC) or (Suc-LLVY)2-Rhodamine110 were used as substrates (Cedarlane). We observed fluorescence quenching of AMC fluorescence by azabicyclenes, and so (Suc-LLVY)2-Rhodamine110 was used for all assays containing these compounds. ADEP was purified from *S. hawaiiensis* NRRL15010

as previously described[23]. Where indicated, agonist peptides (Sigma) were tested at 500 μM. 100 μL reactions were initiated with addition of substrate, carried out at room temperature, and tracked by fluorescence excitation/emission 360 nm/460 nm (AMC) or 485 nm/525 nm (rhodamine 110).

Thermal shift assays and whole protein mass spectrometry

Melt curves were performed using the following conditions: 5 μM ClpP subunit (monomer), 5x SYPRO orange dye (5000x stock, ThermoFisher), 100 μM ADEP (or DMSO), in ClpP reaction buffer. 50 μL reactions were assayed on a BioRad CFX96 real time system, ramping from 35°C to 95°C in increments of 0.5°C for 10 seconds, reading SYPRO fluorescence after every increment. Melting temperatures were calculated using CFX Maestro software.

Whole protein MS was performed on an Agilent 6546 LC/Q-TOF in positive ion mode, with a ZORBAX StableBond 300 C3 column (Agilent, 3.0 x 150 mm, 3.5 μm). To assess cross-processing, subunits were mixed together in 1:1 molar ratios and incubated overnight at room temperature. 1 μL of 30-40 μM protein was injected for analysis and spectra were deconvoluted using mMass open source software. Processing sites were predicted using ExPASy FindPept.

REFERENCES

1. Culp, E. & Wright, G. D. Bacterial proteases, untapped antimicrobial drug targets. *J. Antibiot. (Tokyo).* **70**, 366–377 (2017).

2. Sauer, R. T. & Baker, T. A. AAA+ Proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80**, 587–612 (2011).

3. Glynn, S. E. Multifunctional mitochondrial AAA proteases. *Frontiers in Molecular Biosciences* **4**, (2017).

4. Alexopoulos, J. A., Guarné, A. & Ortega, J. ClpP : A structurally dynamic protease regulated by AAA + proteins. *J. Struct. Biol.* **179**, 202–210 (2012).

5. Compton, C. L., Schmitz, K. R., Sauer, R. T. & Sello, J. K. Antibacterial activity of and resistance to small molecule inhibitors of the clpp peptidase. *ACS Chem. Biol.* **8**, 2669–2677 (2013).

6. Ju, Y. *et al.* Discovery of Novel peptidomimetic boronate ClpP inhibitors with noncanonical enzyme mechanism as potent virulence blockers in vitro and in vivo. *J. Med. Chem.* **63**, 3104–3119 (2020).

7. Hackl, M. W. *et al.* Phenyl esters are potent inhibitors of Caseinolytic Protease P and reveal a stereogenic switch for deoligomerization. *J. Am. Chem. Soc.* **137**, 8475–83 (2015).

8. Böttcher, T. & Sieber, S. A. Structurally refined β-lactones as potent inhibitors of devastating bacterial virulence factors. *ChemBioChem* **10**, 663–666 (2009).

9. Lakemeyer, M. *et al.* Tailored peptide phenyl esters block ClpXP proteolysis by an unusual breakdown into a heptamer–hexamer assembly. *Angew. Chemie Int. Ed.* **58**, 7127–7132 (2019).

10. Böttcher, T. & Sieber, S. A. β-Lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of *Staphylococcus aureus*. *J. Am. Chem. Soc.* **130**, 14400–14401 (2008).

11. Seo, J. H. *et al.* The mitochondrial unfoldase-peptidase complex ClpXP controls bioenergetics stress and metastasis. *PLoS Biol.* **14**, (2016).

12. Cole, A. *et al.* Inhibition of the mitochondrial protease ClpP as a therapeutic strategy for human acute myeloid leukemia. *Cancer Cell* **27**, 864–876 (2015).

13. Michel, K. H. & Kastner, R. E. A54556 antibiotics and process for production thereof. US 4492650 A (1985).

14. Brötz-oesterhelt, H. & Sass, P. Bacterial caseinolytic proteases as novel targets for antibacterial treatment. *Int. J. Med. Microbiol.* **304**, 23–30 (2014).

15. Li, D. H. S. *et al.* Acyldepsipeptide antibiotics induce the formation of a structured axial channel in ClpP: A model for the ClpX/ClpA-bound state of ClpP. *Chem. Biol.* **17**, 959–969 (2010).

16. Lee, B. *et al.* Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat. Struct. Mol. Biol.* **17**, 471–478 (2010).

17. Gavrish, E. *et al.* Lassomycin, a ribosomally synthesized cyclic peptide, kills *Mycobacterium tuberculosis* by targeting the ATP-dependent protease ClpC1P1P2. *Chem. Biol.* **21**, 509–518 (2014).

18. Choules, M. *et al.* A rufomycin analogue is an anti-tuberculosis drug lead targeting CLPC1 with no cross resistance to ecumicin. *Planta Med.* **81**, CL2 (2015).

19. Gao, W. *et al.* Discovery and Characterization of the Tuberculosis Drug Lead Ecumicin. *Org. Lett.* **16**, 6044–6047 (2014).

20. Schmitt, E. K. *et al.* The natural product cyclomarin kills *Mycobacterium tuberculosis* by targeting the ClpC1 subunit of the caseinolytic protease. *Angew. Chemie* **123**, 6011–6013 (2011).

21. Tang, X. *et al.* Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10**, 2841–2849 (2015).

22. Alanjary, M. *et al.* The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res.* **45**, W42–W48 (2017).

23. Thomy, D. *et al.* The ADEP biosynthetic gene cluster in *Streptomyces hawaiiensis* NRRL 15010 reveals an accessory ClpP gene as a novel antibiotic resistance factor. *Appl. Environ. Microbiol.* **85**, e01292-19 (2019).

24. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).

25. Gominet, M., Seghezzi, N. & Mazodier, P. Acyl depsipeptide (ADEP) resistance in Streptomyces. *Microbiology* **157**, 2226–2234 (2011).

26. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).

27. Schmidt, Y. *et al.* Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in plant-associated *Pseudomonas*. *ChemBioChem*

15, 259–266 (2014).

28. Johnston, C. W., Zvanych, R., Khyzha, N. & Magarvey, N. A. Nonribosomal assembly of natural lipocyclocarbamate lipoprotein-associated phospholipase inhibitors. *ChemBioChem* **14**, 431–435 (2013).

29. Liu, L. *et al.* Activation and characterization of bohemamine biosynthetic gene cluster from *Streptomyces* sp. CB02009. *Org. Lett.* **22**, 4614–4619 (2020).

30. Hong, Z. *et al.* Azetidine-containing alkaloids produced by a quorum-sensing regulated nonribosomal peptide synthetase pathway in Pseudomonas aeruginosa. *Angew. Chemie* **131**, 3210–3214 (2019).

31. Patteson, J. B., Lescallette, A. R. & Li, B. Discovery and biosynthesis of azabicyclene, a conserved nonribosomal peptide in *Pseudomonas aeruginosa*. *Org. Lett.* **21**, 4955–4959 (2019).

32. Bellier, A. & Mazodier, P. ClgR, a novel regulator of *clp* and *lon* expression in *Streptomyces*. *Society* **186**, 3238–3248 (2004).

33. Schmitza, K. R., Carneyb, D. W., Sellob, J. K. & Sauera, R. T. Crystal structure of *Mycobacterium tuberculosis* ClpP1P2 suggests a model for peptidase activation by AAA+ partner binding and substrate delivery. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4587–E4595 (2014).

34. Nagpal, J. *et al.* Molecular and structural insights into an asymmetric proteolytic complex (ClpP1P2) from *Mycobacterium smegmatis*. *Sci. Rep.* **9**, 1–12 (2019).

35. Viala, J. & Mazodier, P. ClpP-dependent degradation of PopR allows tightly regulated expression of the *clpP3 clpP4* operon in *Streptomyces lividans*. *Mol Microbiol* **44**, 633–643 (2002).

36. Gersch, M. *et al.* AAA+ chaperones and acyldepsipeptides activate the ClpP protease via conformational control. *Nat. Commun.* **6**, 6320 (2015).

37. Leodolter, J., Warweg, J. & Weber-Ban, E. The *Mycobacterium tuberculosis* ClpP1P2 protease interacts asymmetrically with its ATPase partners ClpX and ClpC1. *PLoS One* **10**, (2015).

38. Huang, S. *et al.* Discovery of a single monooxygenase that catalyzes carbamate formation and ring contraction in the biosynthesis of the legonmycins. *Angew. Chemie* **127**, 12888–12892 (2015).

39. Schimming, O. *et al.* Structure, biosynthesis, and occurrence of bacterial pyrrolizidine alkaloids. *Angew. Chemie Int. Ed.* **54**, 12702–12705 (2015).

40. Hua, J.-F. *et al.* Jenamidines A to C: Unusual alkaloids from *Streptomyces* sp. with specific antiproliferative properties obtained by chemical screening. *J. Antibiot. (Tokyo).* **56**, 747–754 (2003).

41. Raju, R. M. *et al.* Post-translational regulation via Clp protease Is critical for survival of *Mycobacterium tuberculosis*. *PLoS Pathog.* **10**, e1003994 (2014).

42. Davies, J., Spiegelman, G. B. & Yim, G. The world of subinhibitory antibiotic concentrations. *Current Opinion in Microbiology* **9**, 445–453 (2006).

43. Yim, G., Spiegelman, G. B. & Davies, J. E. Separate mechanisms are involved in rifampicin upmodulated and downmodulated gene expression in *Salmonella typhimurium*. *Res. Microbiol.* **164**, 416–424 (2013).

44. Vázquez-Laslop, N. & Mankin, A. S. How macrolide antibiotics work. *Trends in Biochemical Sciences* **43**, 668–684 (2018).

45. Takano, E. Gamma-butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr. Opin. Microbiol.* **9**, 287–294 (2006).

46. Kostylev, M. *et al.* Evolution of the *Pseudomonas aeruginosa* quorum-sensing hierarchy. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7027–7032 (2019).

47. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (John Innes Foundation, 2000).

48. Xu, M. *et al.* GPAHex: A synthetic biology platform for Type IV-V glycopeptide antibiotic production and discovery. *Submit 03/09/2020 to Nat. Commun.*

49. Yamanaka, K. *et al.* Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci.* **111**, 1957–1962 (2014).

50. Luo, Y., Zhang, L., Barton, K. W. & Zhao, H. Systematic identification of a panel of strong constitutive promoters from *Streptomyces albus*. (2015). doi:10.1021/acssynbio.5b00016

51. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).

52. Ji, C.-H., Kim, J.-P. & Kang, H.-S. Library of synthetic *Streptomyces* regulatory sequences for use in promoter engineering of natural product biosynthetic gene clusters. *ACS Synth. Biol.* **7**, 1946–1955 (2018).

53. Gust, B., Kieser, T. & Chater, K. PCR targeting system in *Streptomyces coelicolor*

A3(2). (2002).

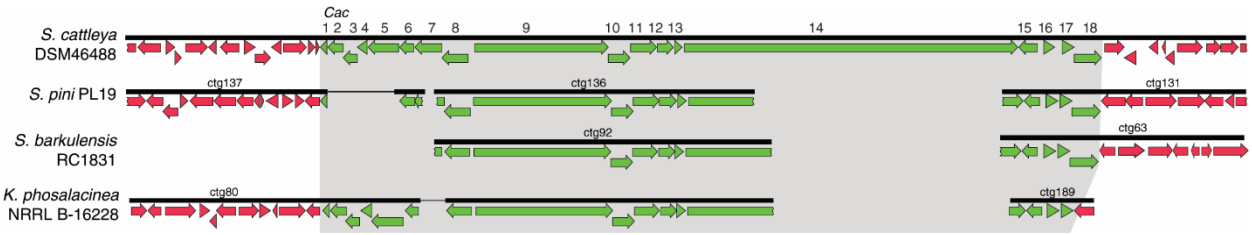
54. Jones, A. C. *et al.* Phage P1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PLoS One* **8**, e69319 (2013).

55. 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 (2000).

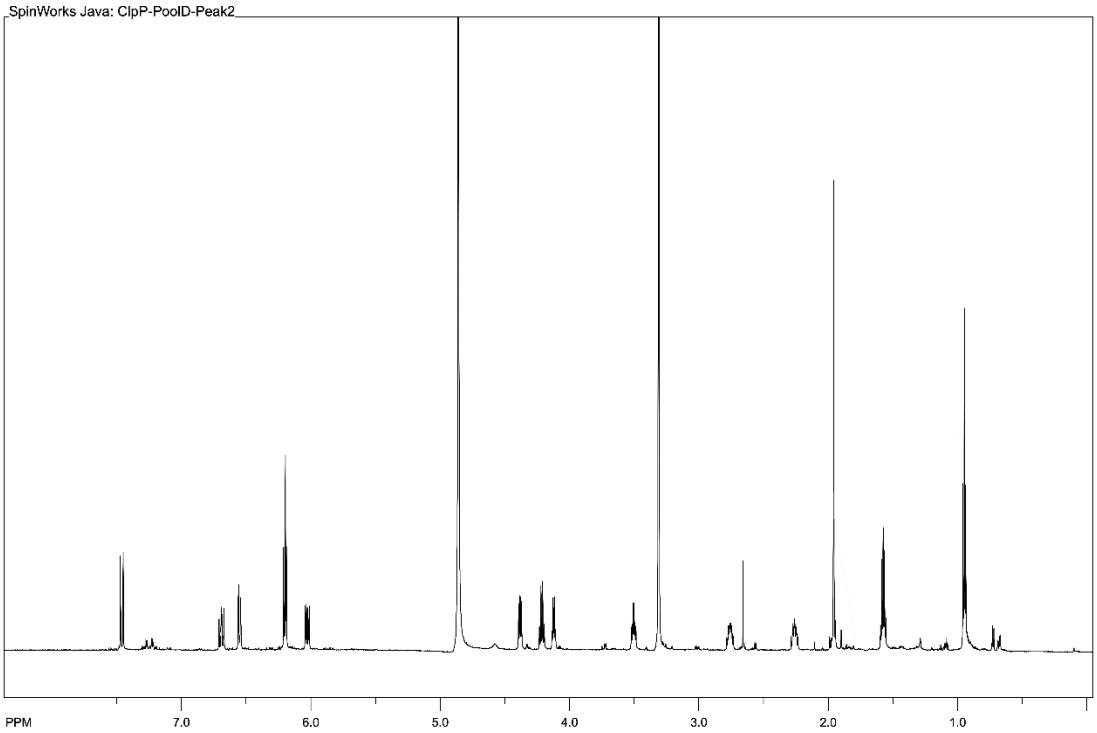
56. Gomez-Escribano, J. P. & Bibb, M. J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207–215 (2011).

57. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61–68 (1992).

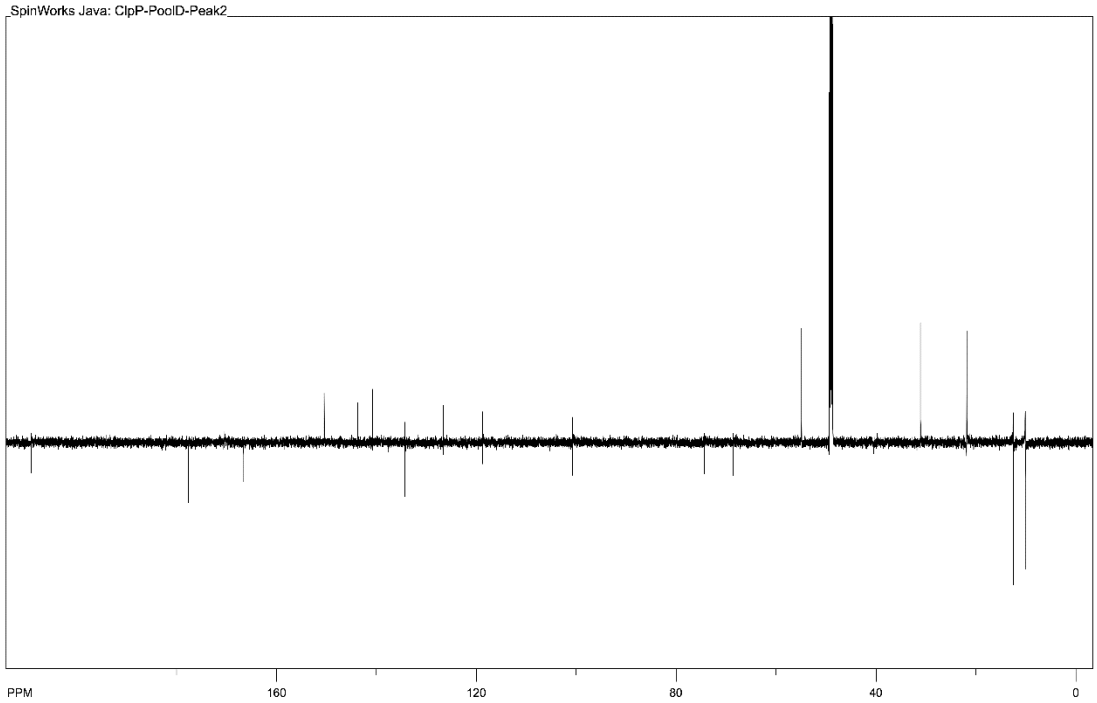
SUPPLEMENTARY INFORMATION



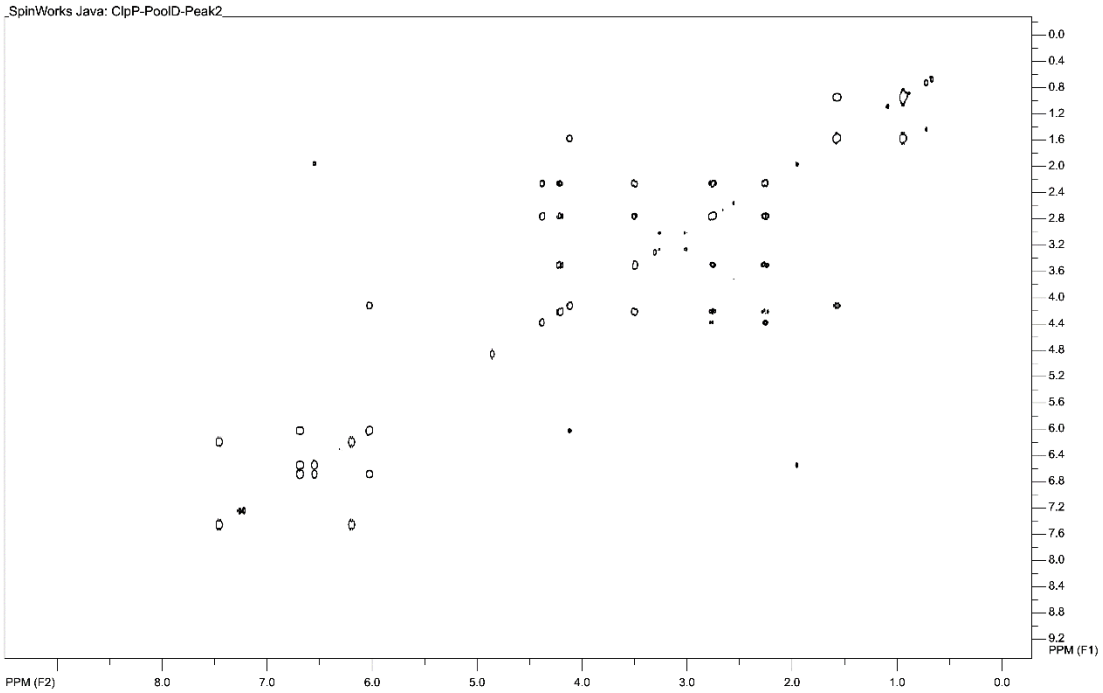
Supplementary Figure 1. Alignment of *cac* homologous clusters.
Alignment of homologous *cac* BGCs in *S. pini* PL19, *S. barkulensis* RC1831 and *K. phosalacinea* NRRL B-16228. Green arrows and grey shading highlight a region of conserved gene synteny. Red genes outside of this region are not homologous. Clusters in *S. pini, S. barkulensis* and *K. phosalacinea* are incomplete and broken over multiple contigs, as represented by black bars.



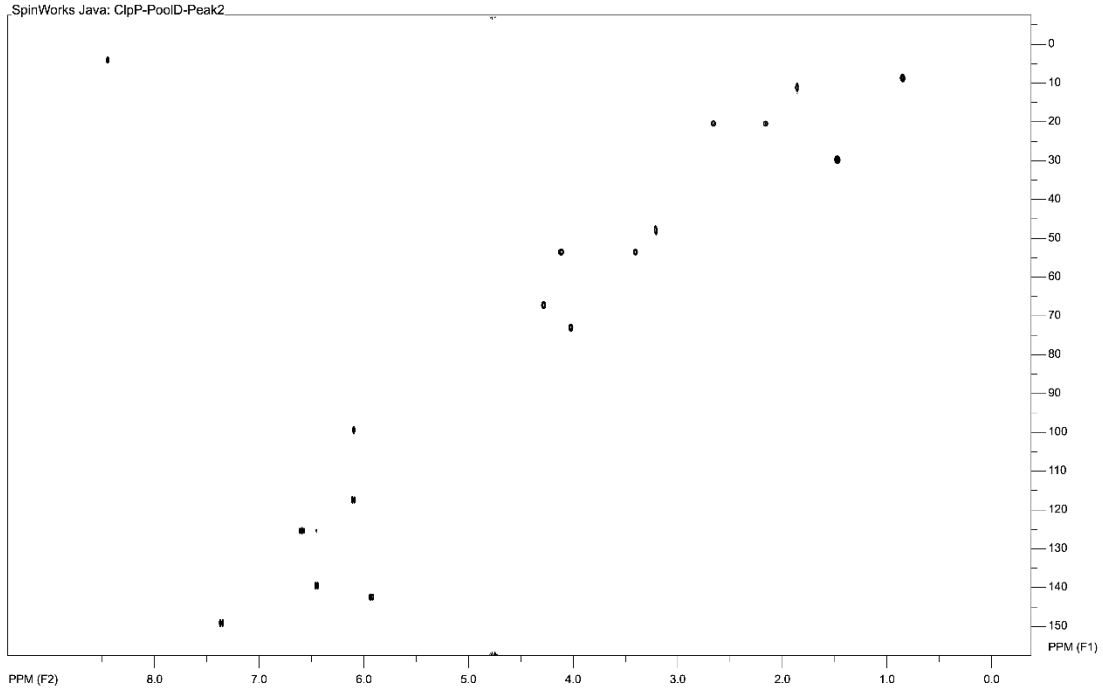
Supplementary Figure 2. The ^{1}H NMR spectrum of azabicyclene C (d^{4}-MeOD).



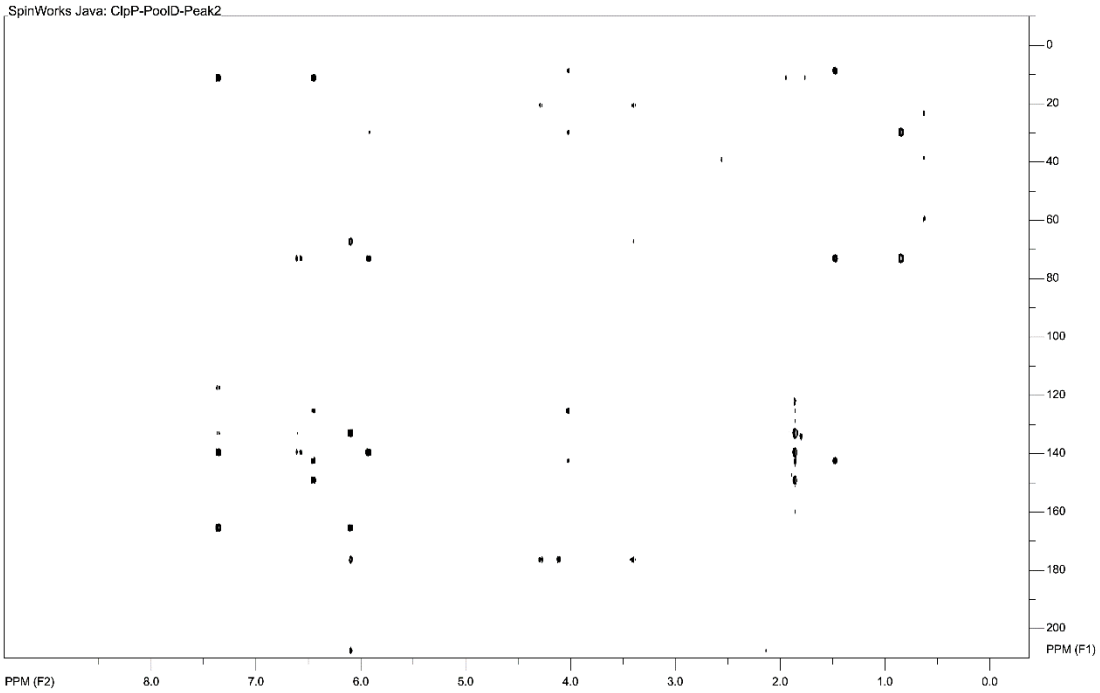
Supplementary Figure 3. The ^{13}C NMR spectrum of azabicyclene C (d^{4}-MeOD).



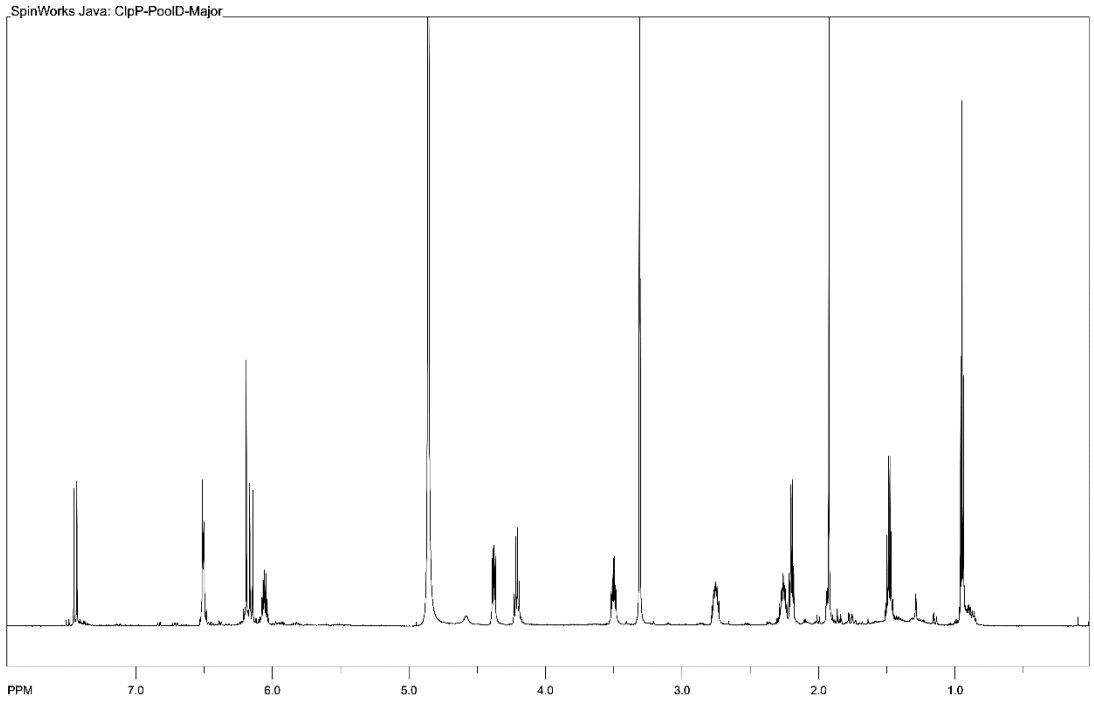
Supplementary Figure 4. The ^{1}H-^{1}H COSY NMR spectrum of azabicyclene C (d^{4}-MeOD).



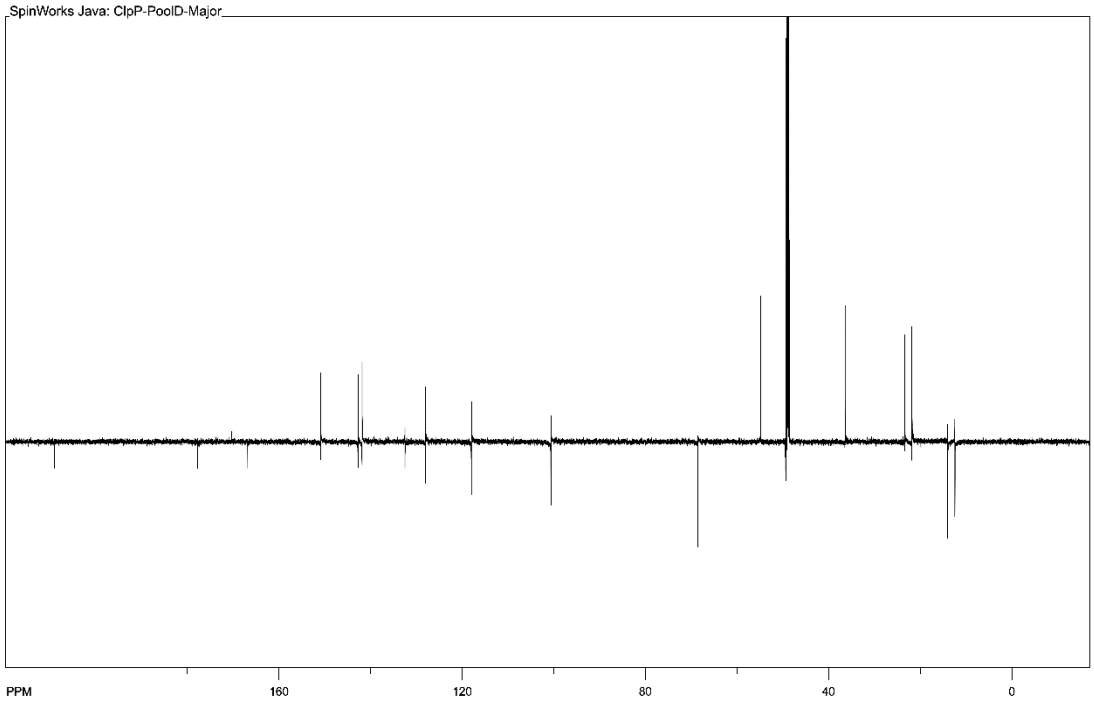
Supplementary Figure 5. The ^{1}H-^{13}C HSQC NMR spectrum of azabicyclene C (d^{4}-MeOD).



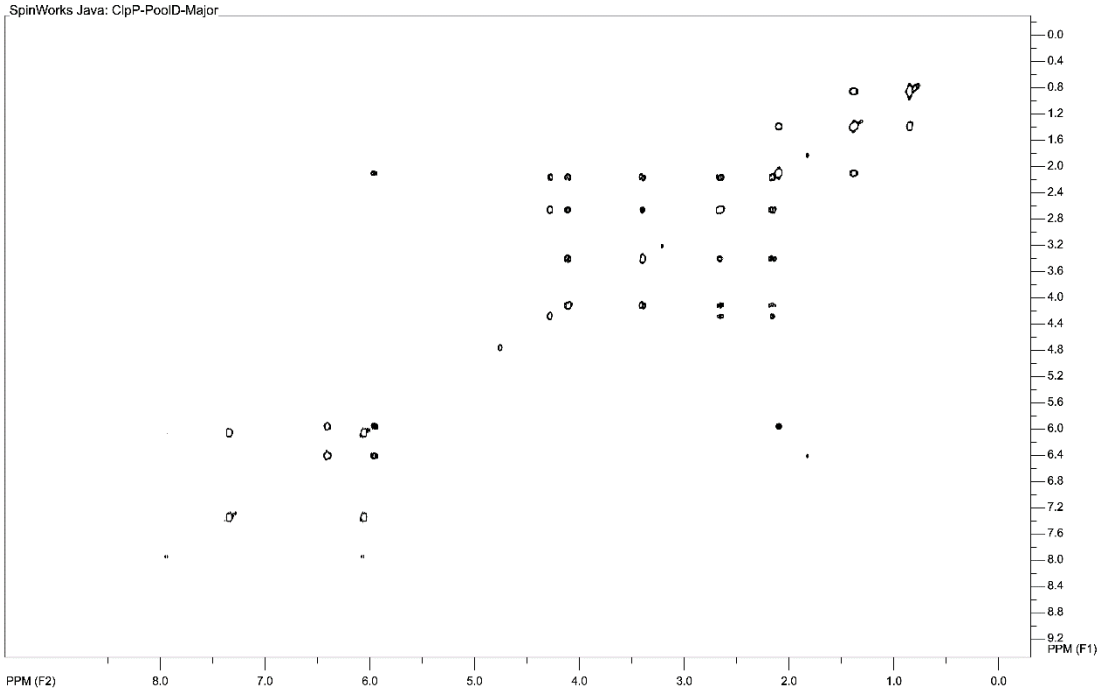
Supplementary Figure 6. The ^{1}H-^{13}C HMBC NMR spectrum of azabicyclene C (d^4-MeOD).



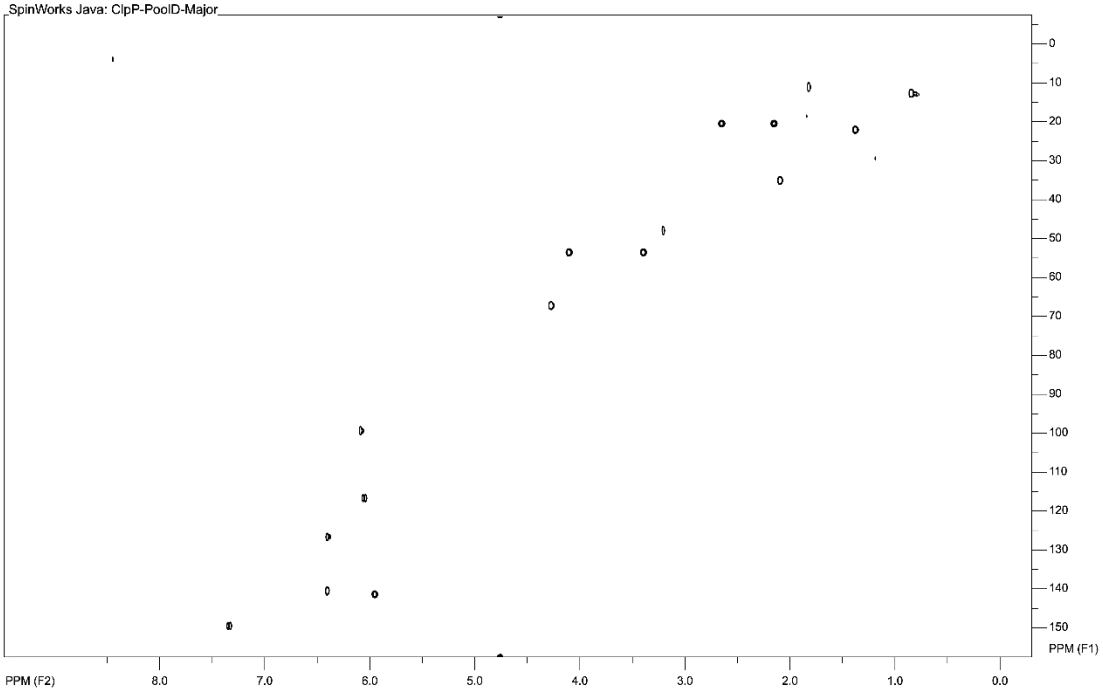
Supplementary Figure 7. The ^{1}H NMR spectrum of azabicyclene D (d^4-MeOD)



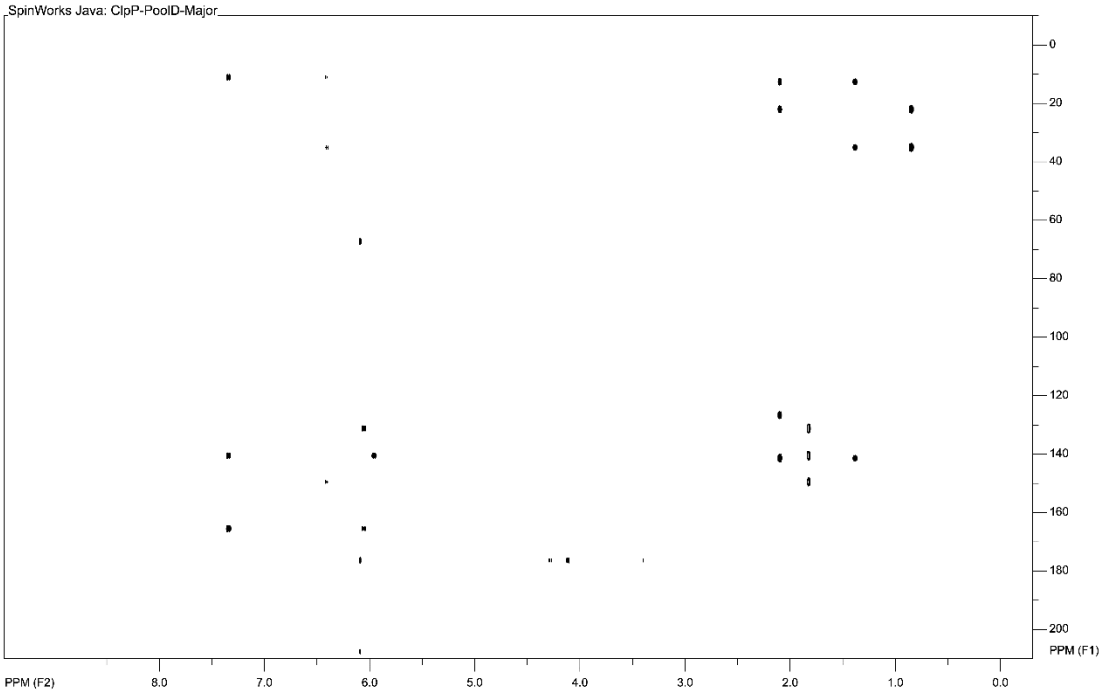
Supplementary Figure 8. The ^{13}C NMR spectrum of azabicyclene D (d^4-MeOD).



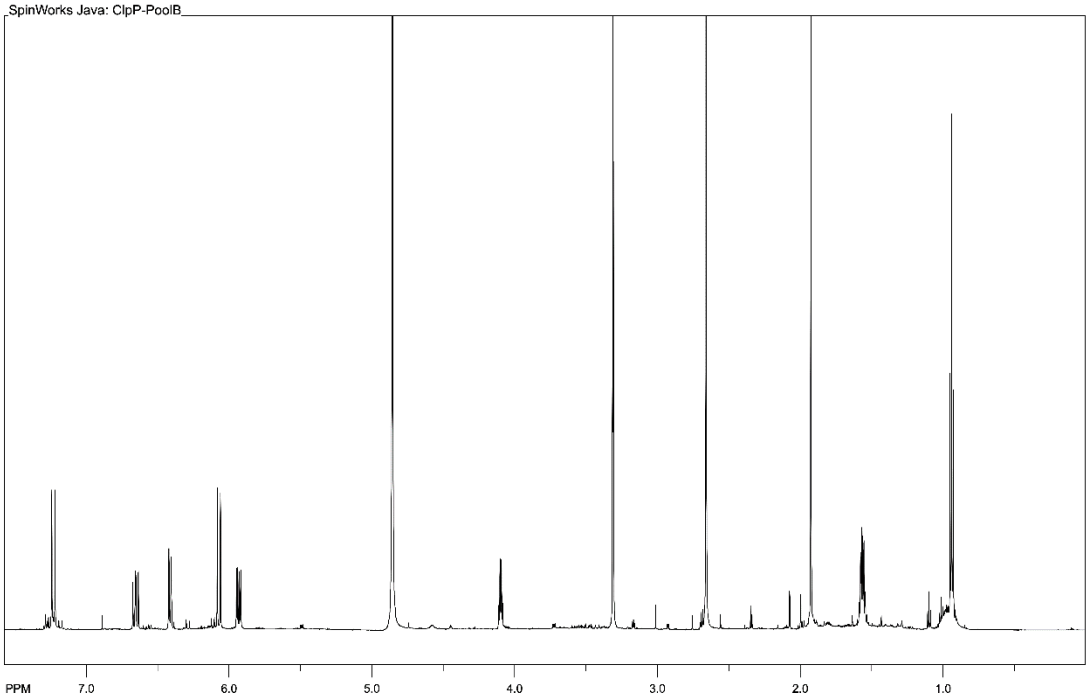
Supplementary Figure 9. The ^1H-^1H COSY NMR spectrum of azabicyclene D (d^4-MeOD).



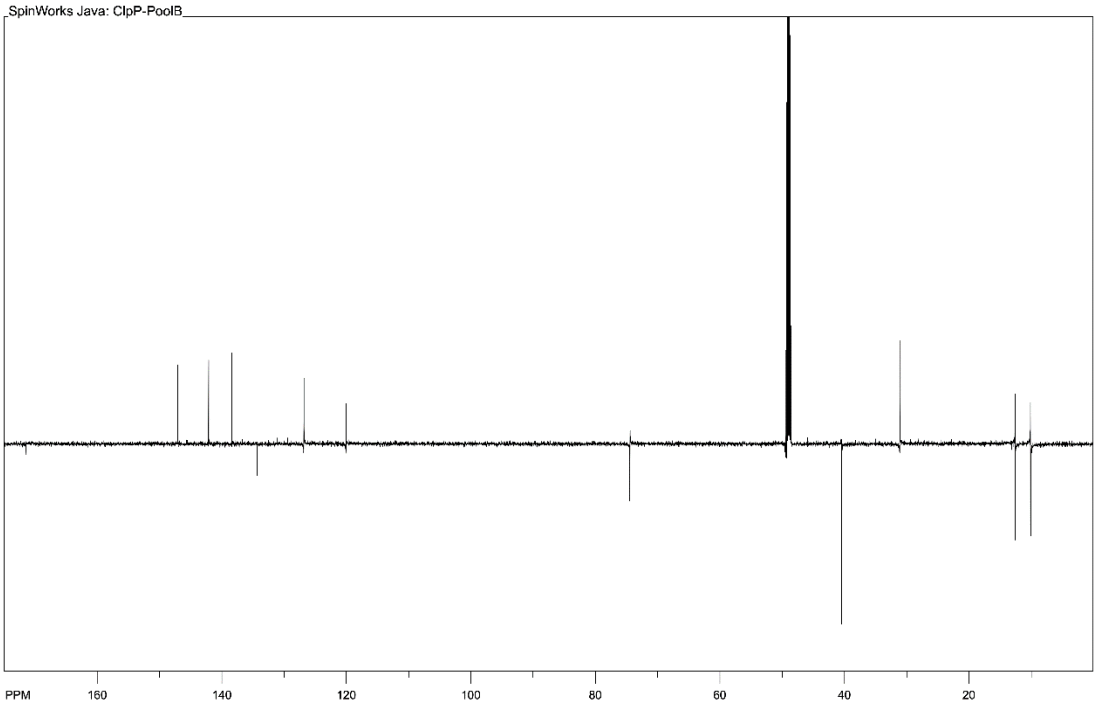
Supplementary Figure 10. The ^{1}H-^{13}C HSQC NMR spectrum of azabicyclene D (d^{4}-MeOD).



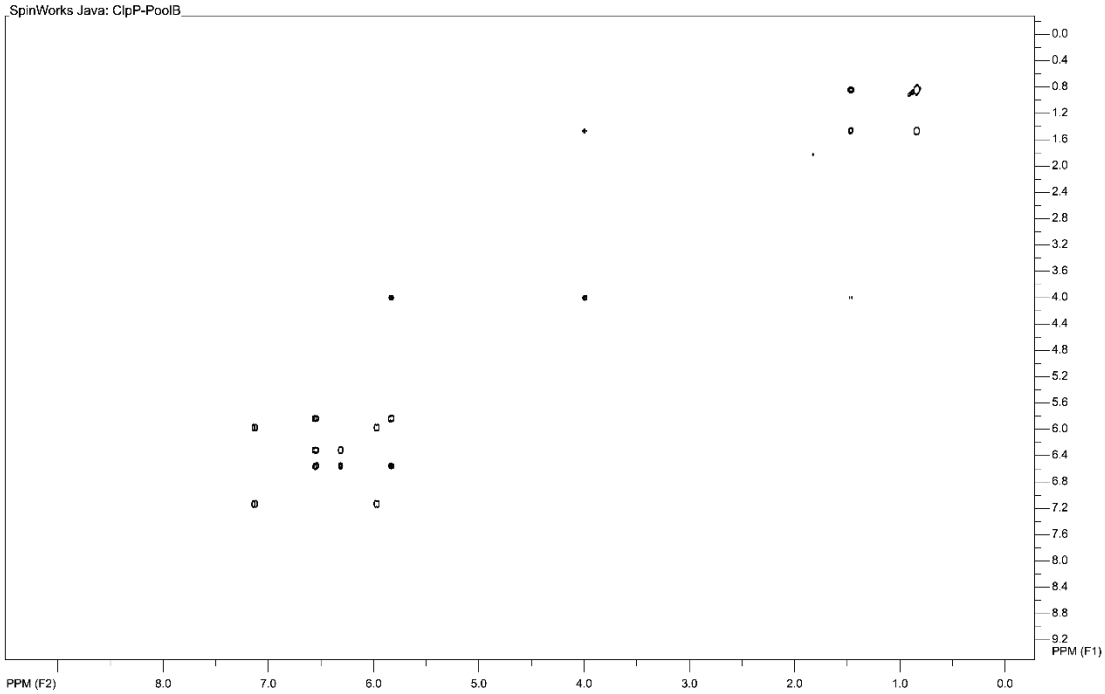
Supplementary Figure 11. The ^{1}H-^{13}C HMBC NMR spectrum of azabicyclene D (d^{4}-MeOD).



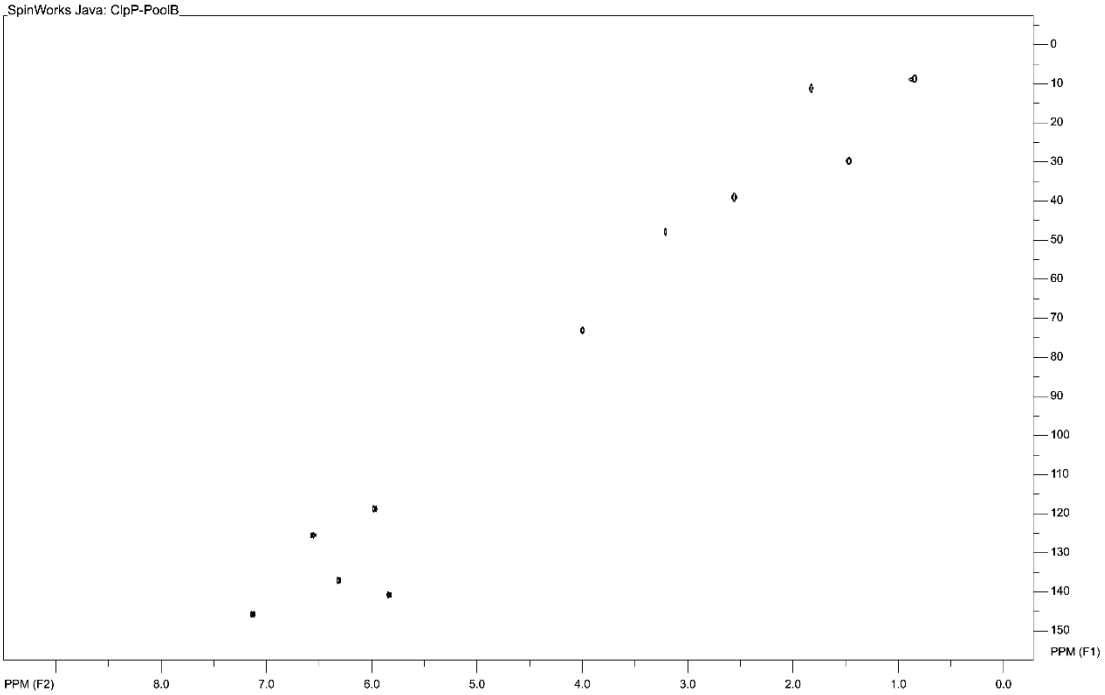
Supplementary Figure 12. The ^{1}H NMR spectrum of compound 2 (d^{4}-MeOD).



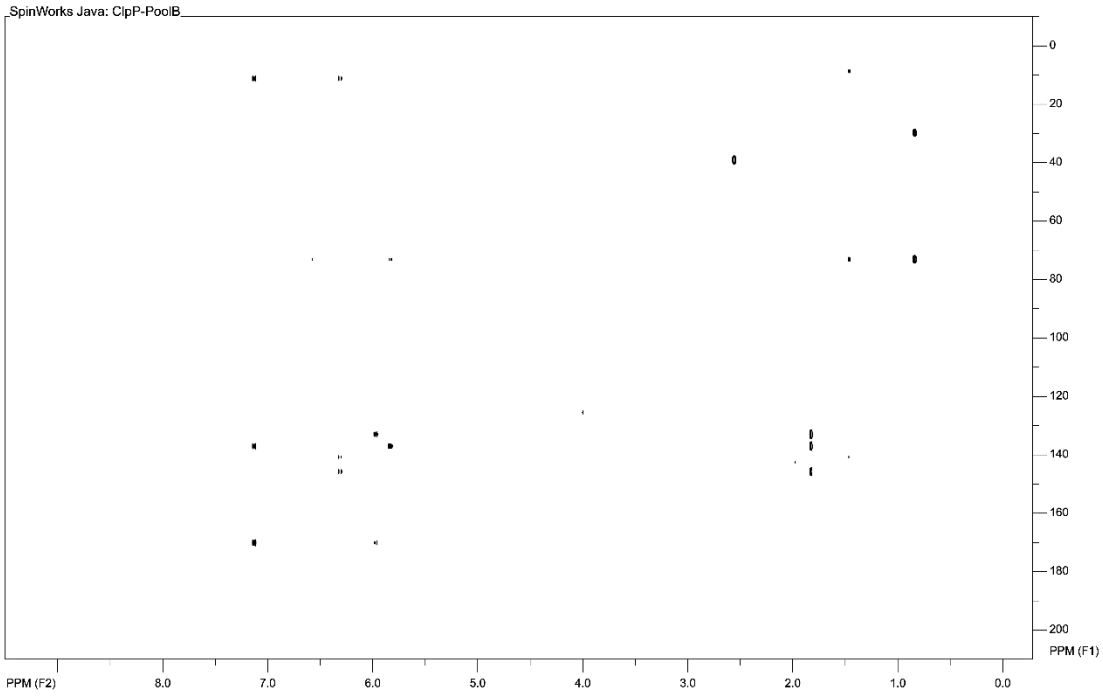
Supplementary Figure 13. The ^{13}C NMR spectrum of compound 2 (d^{4}-MeOD).



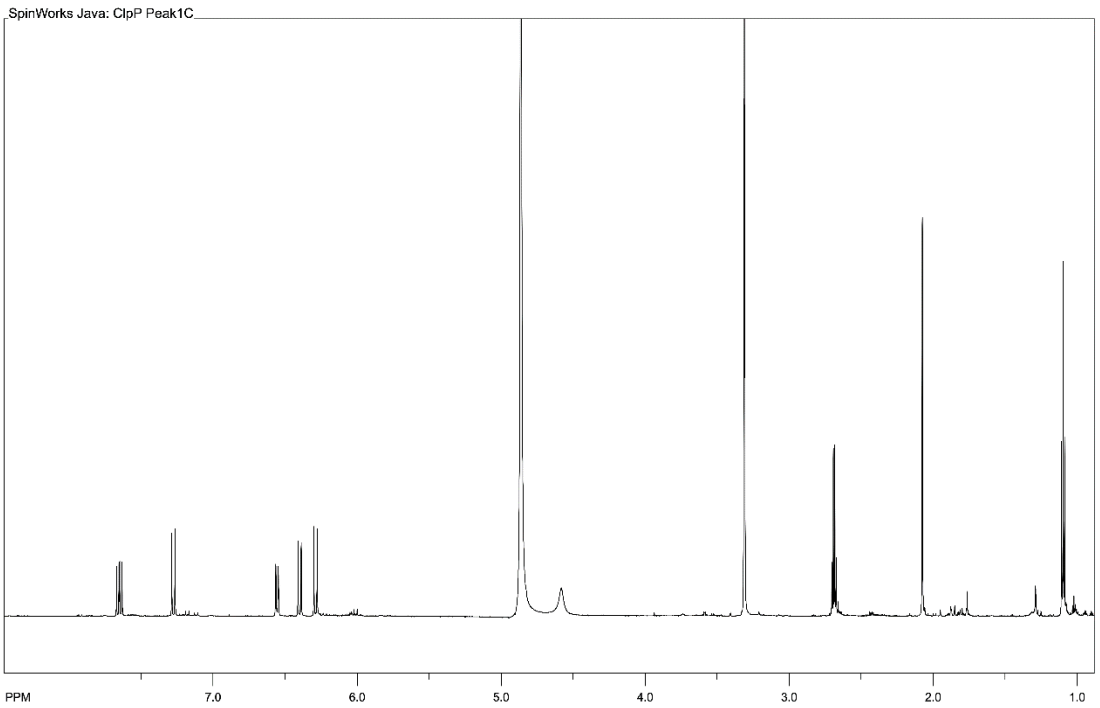
Supplementary Figure 14. The ^{1}H-^{1}H COSY NMR spectrum of compound 2 (d^{4}-MeOD).



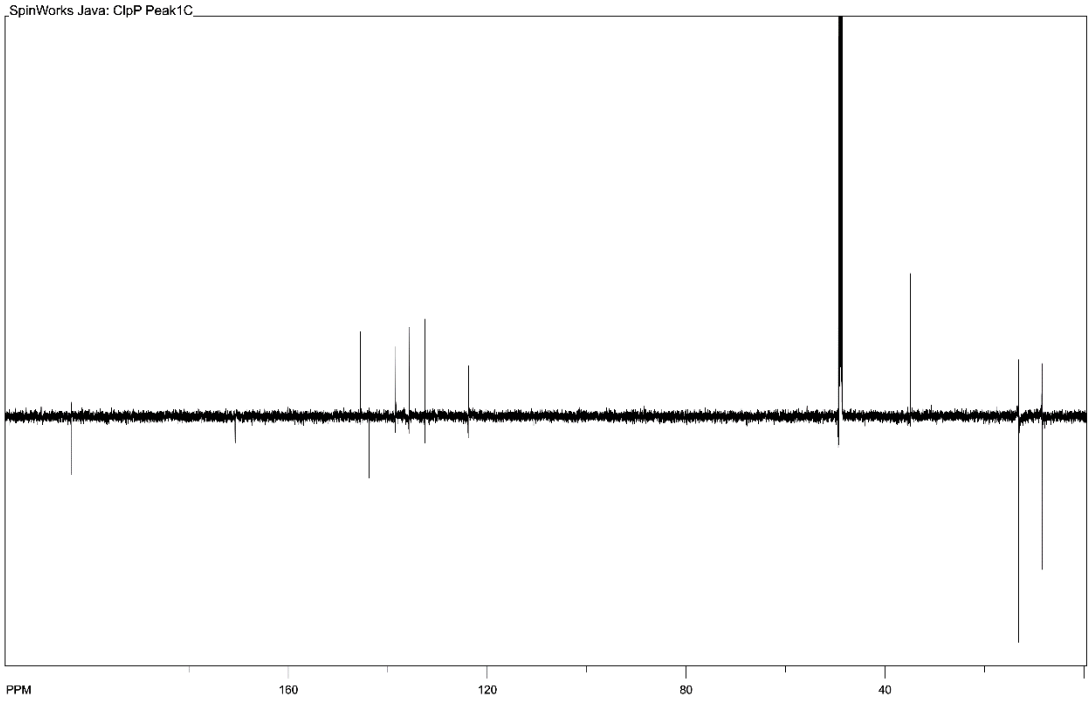
Supplementary Figure 15. The ^{1}H-^{13}C HSQC NMR spectrum of compound 2 (d^{4}-MeOD).



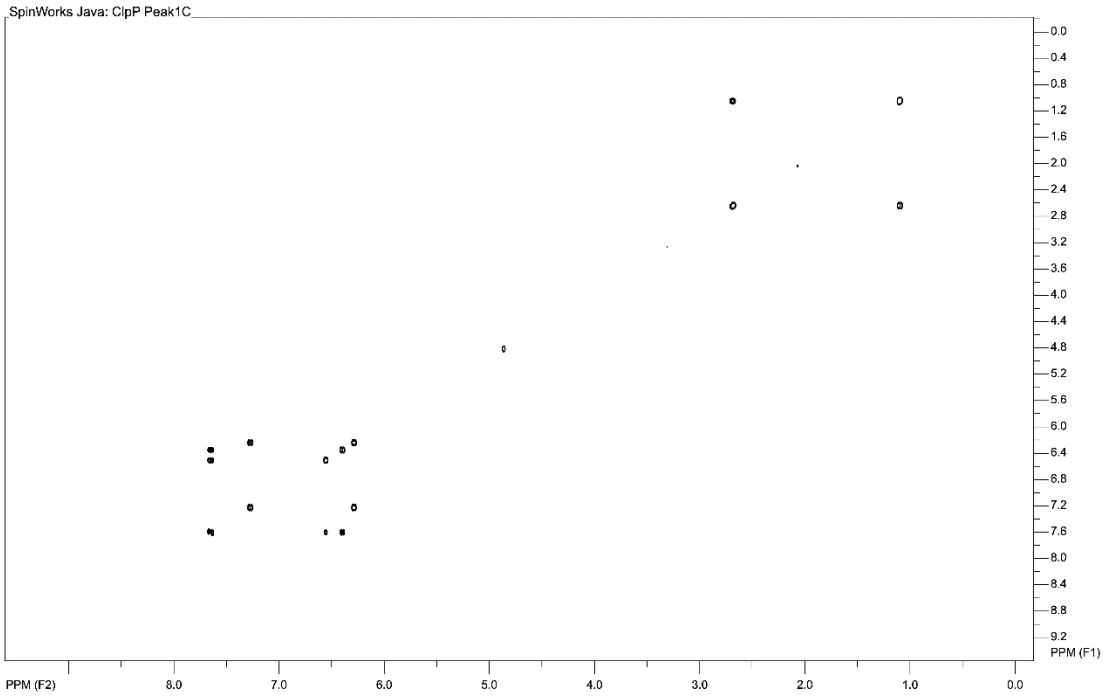
Supplementary Figure 16. The ^{1}H-^{13}C HMBC NMR spectrum of compound 2 (d^{4}-MeOD).



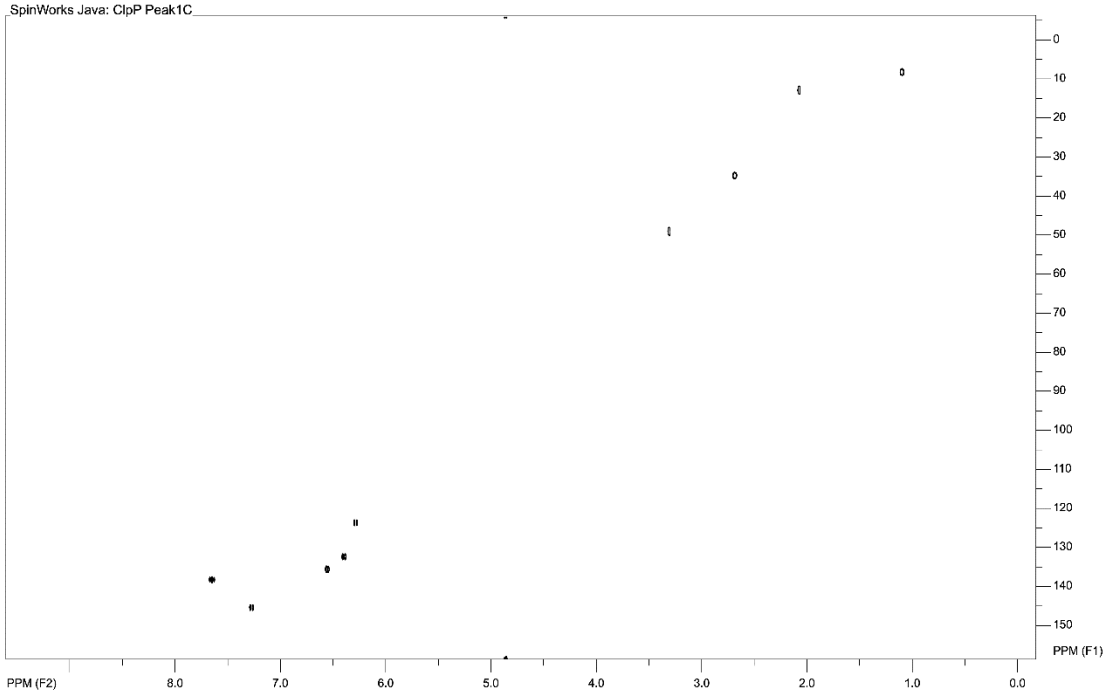
Supplementary Figure 17. The ^{1}H NMR spectrum of compound 1 (d^{4}-MeOD).



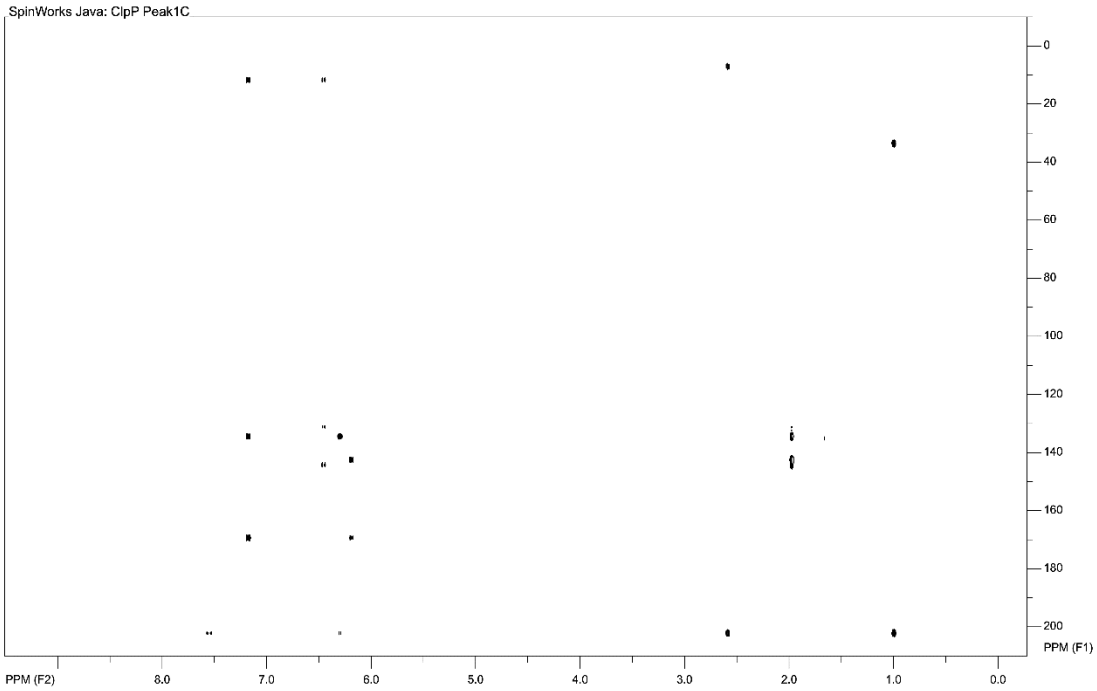
Supplementary Figure 18. The ^{13}C NMR spectrum of compound 1 (d^4-MeOD)



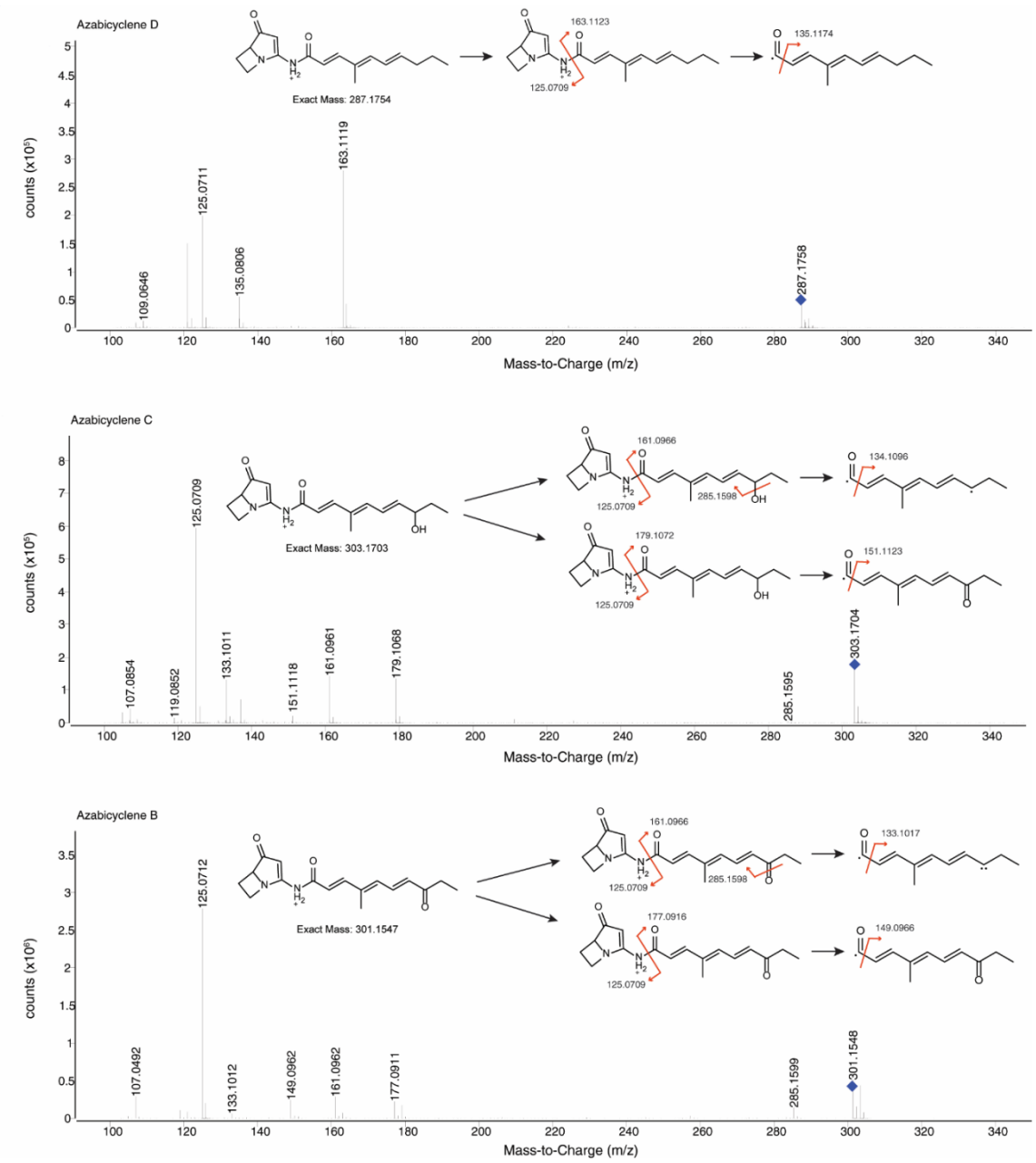
Supplementary Figure 19. The ^{1}H-^{1}H COSY NMR spectrum of compound 1 (d^4-MeOD)



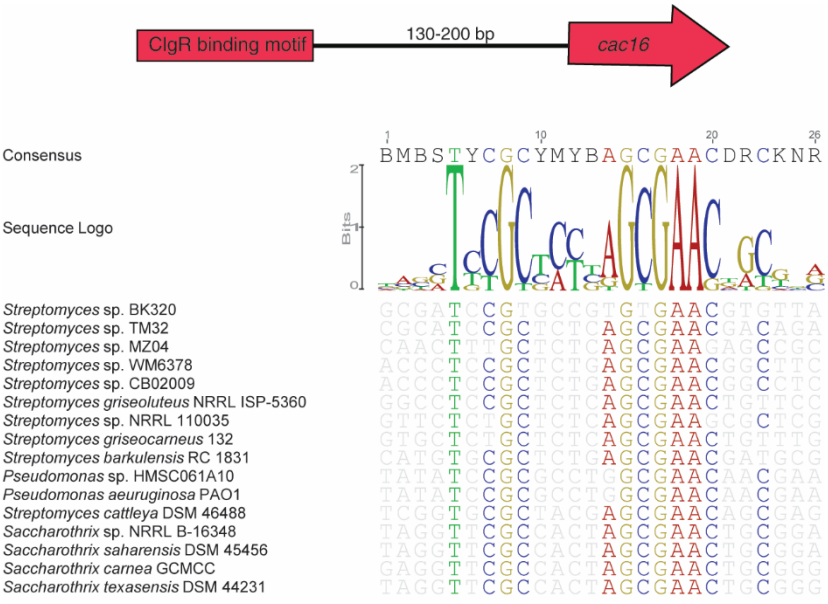
Supplementary Figure 20. The ^{1}H-^{13}C HSQC NMR spectrum of compound 1 (d^{4}-MeOD)



Supplementary Figure 21. The ^{1}H-^{13}C HMBC NMR spectrum of compound 1 (d^{4}-MeOD)



Supplementary Figure 22. LC-MS/MS of azabicyclenes.
Exact mass of predicted ions are shown. MS spectra were acquired in positive mode with 10eV collision energy.



Supplementary Figure 23. ClgR binding motif.
The ClgR regulon binding motif, as predicted in promoter regions of *cac16* homologs listed, was used to construct a sequence logo similar to the previously characterized binding motif controlling *clpP1* expression in *S. lividans,* CGC-5N-GCGAAC[32].

Supplementary Table 1. Final 10 hits from ClpP target-directed genome mining.

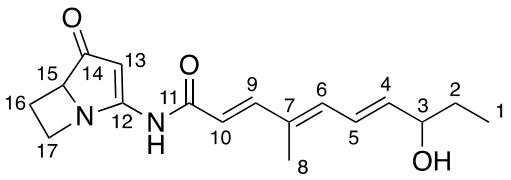
Strain	Genbank genome accession	Associated with Bimodular NRPS?
Streptomyces cattleya	NC_017586	Y
Streptomyces sp. NBRC 110035, TP-A0873	NZ_BBNN01000058	Y
Senegalimassilia anaerobia JC110	NZ_HE611016	N
Streptomyces griseoluteus ISP-5360	NZ_JOBE01000059	Y
Streptomyces scopuliridis RB72	NZ_JOEI01000002	N
Streptosporangium roseum strain NRRL B-2638	NZ_JOEP01000017	N
Amycolatopsis balhimycina FH 1894	NZ_KB913037	N
Streptomyces sp. WM6378 P402	NZ_LGDD01000350	Y
Saccharothrix sp. NRRL B-16348	NZ_LGED01000125	Y
Streptomyces sp. CB02009	NZ_LIVV01000003	Y

Supplementary Table 2. Nucleotide sequences used for CORASON analysis in Figure 1b.

Strain	Genbank accession	% amino acid identity to Cac9	ClpP associated?
Streptomyces cattleya	NC_017586	100%	Y
Streptomyces pini strain PL19 NRRL B-24728	NZ_FOSG01000037.1	75.23	Y
Kitasatospora phosalacinea NRRL B-16228	NZ_JNYE01000189	74.5	Y
Streptomyces barkulensis RC 1831	NZ_PGSG01000092.1	74.83	Y
Streptomyces griseocarneus strain 132	PENC01000002.1	58.85	Y
Streptomyces sp. NBRC 110035, TP-A0873	NZ_BBNN01000058.1	55.88	Y
Streptomyces sp. TM32	NZ_SDIG01000007.1	55.96	Y
Streptomyces sp. CB02009	LIVV01000003.1	55.43	Y
Streptomyces sp. BK340	NZ_VISR01000015.1	56.01	Y
Streptomyces sp. MZ04	NZ_SRIC01000011.1	55.51	Y
Streptomyces sp. WM6378 P402	NZ_LGDD01000350.1	54.58	Y
Saccharothrix sp. NRRL B-16348	NZ_LGED01000125.1	54.42	Y
Saccharothrix saharensis SA152	NZ_VFPP01000001.1	54.84	Y
Saccharothrix carnea strain CGMCC 4.7097	PYAX01000004.1	54.64	Y
Saccharothrix texasensis DSM 44231	NZ_RJKM01000001.1	54.04	Y
Amycolatopsis sp. CA-126428	NZ_PPHF01000129.1	56.9	Y
Streptomyces griseoluteus ISP-5360	JOBE00000000	58.71	Y
Streptomyces exfoliatus NRRL B-2924	JNZP00000000.1	58.88	N
Streptomyces rimosus NRRL B-8076	NZ_JNYK01000026.1	55.17	N
Streptomyces chattanoogensis NRRL ISP-5002	NZ_LGKG01000112.1	55.66	N
Kitasatospora aureofaciens NRRL B-2658	LGUY01000185.1	55.17	N
Streptomyces sp. NEAU-D10	NZ_QUAC01000057.1	55.02	N
Streptomyces tubercidicus NBRC 13090	BLIR01000001.1	55.48	N
Streptomyces lydicus A02	NZ_CP007699.2	55.73	N
Streptomyces subrutilus 10-1-1	NZ_MEHK01000001.1	55.59	N
Murinocardiopsis flavida DSM 45312	PYGA01000001.1	57.56	N
Pseudomonas sp. SHC52	CBLV010000330.1	36.14	N
Pseudomonas corrugata RM1-1-4	NZ_CP014262.1	35.40	N
Pseudomonas mediterranea CFBP 5447	LT629790.1	35.21	N
Pseudomonas aeruginosa PAO1	AE004091.2	52.70	N
Pseudomonas sp. HMSC061A10	NZ_KV830129.1	52.65	N

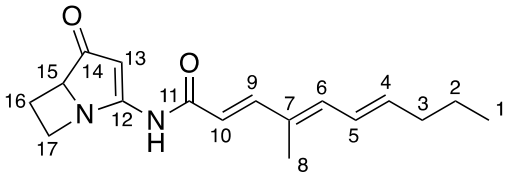
Supplementary Table 3. Detailed *cac* annotations.

S. cattleya DMS8057			*S. sp. CB02009 bhm* (identity)	*P. aeruginosa PAO1 aze* (identity)	Top hit tblastx	
Locus tag	**ORF**	**putative function**			**species**	**identity (coverage)**
SCATT_RS15315	Cac1	tRNA or nucleotide binding protein	-	-	*Streptomyces pini*	76(100)
SCATT_RS15320	Cac2	MtaP - 5'-methylthioadenosine phosphorylase	-	-	*Kitasatospora phosalacinea*	84(98)
SCATT_RS15325	Cac3	MtnC - acireductone synthase	-	-	*Streptomyces sp.* CB03234	65(93)
SCATT_RS15330	Cac4	MtnD - cupin acireductone dioxygenase	-	-	*Kitasatospora phosalacinea*	83(98)
SCATT_RS15335	Cac5	MtnAB - bifunctional isomerase/dehydratase	-	-	*Kitasatospora phosalacinea*	74(98)
SCATT_RS15340	Cac6	SAM dependent methyltransferase: L-AZC synthase	-	AzeJ (42%)	*Kitasatospora phosalacinea*	74(95)
SCATT_RS15345	Cac7	crotonyl-CoA carboxylase/reductase	-	-	*Streptomyces melanosporofaciens*	85(99)
SCATT_RS15350	Cac8	cytochrome p450	BhmA (21%)	AzeF (19.6%)	*Streptomyces pini*	80(98)
SCATT_RS15355	Cac9	NRPS	BhmJ (55.5%)	AzeB (52.3%)	*Streptomyces pini*	75(100)
SCATT_RS15360	Cac10	FAD-dependent monooxygenase	BhmK (56.3%)	AzeC (57.5%)	*Streptomyces pini*	83(98)
SCATT_RS15365	Cac11	Condensation domain	-	AzeD (45.2%)	*Streptomyces pini*	71(98)
SCATT_RS15370	Cac12	Oxioreductase/ketoreductase: ACP oxioreductase	-	AzeE (62.2%)	*Streptomyces pini*	79(100)
SCATT_RS15375	Cac13	PhzA/B homolog	-	-	*Streptomyces pini*	77(98)
SCATT_RS15380	Cac14	Type I PKS - 4 modules	-	-	*Actinomadura macra*	57(99)
SCATT_RS15385	Cac15	XRE transcriptional regulator	-	-	*Streptomyces pini*	62(95)
SCATT_RS15390	Cac16	ClpP	BhmB (78.2%)	AzeA (34.8%)	*Streptomyces pini*	88(94)
SCATT_RS15395	Cac17	ClpP	BhmC (67.4%)	AzeA (33.9%)	*Kitasatospora phosalacinea*	79(94)
SCATT_RS15400	Cac18	MFS transporter	-	-	*Streptomyces pini*	76(99)



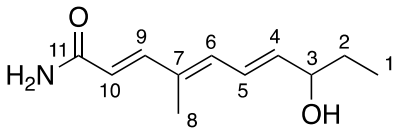
Supplementary Table 4. ^{1}H and ^{13}C NMR data for azabicyclene C (d^4-MeOD)

No.	δ_H (ppm)	δ_C (ppm)
1	0.95 (3H, t, J 7.0 Hz)	10.1
2	1.57 (2H, quin., J 7.0 Hz)	31.1
3	4.12 (1H, q, J 6.0 Hz)	74.3
4	6.03 (1H, dd, J 15.0 and 6.0 Hz)	143.8
5	6.69 (1H, ddd, J 15.0, 11.5 and 1.0 Hz)	126.7
6	6.56 (1H, d, J 11.5 Hz)	140.8
7	-	134.3
8	1.96 (3H, s)	12.6
9	7.46 (1H, d, J 15.0 Hz)	150.5
10	6.20 (1H, d, J 15.0 Hz)	118.7
11	-	166.7
12	-	177.7
13	6.20 (1H, s)	100.7
14	-	209.1
15	4.38 (1H, dd, J 10.0 and 6.5 Hz)	68.6
16	2.76 (1H, dddd, J 11.5, 9.5, 9.0 and 5.5) 2.27 (1H, dddd J 11.5, 10.0, 8.5 and 6.5)	21.8
17	4.21 (1H, q, J 9.0) 3.50 (1H, td, J 9.5, 5.5)	54.9



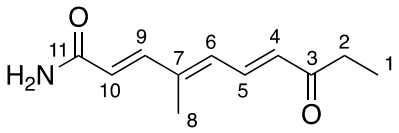
Supplementary Table 5. ^{1}H and ^{13}C NMR data for azabicyclene D (d^{4}-MeOD)

No.	δ_H (ppm)	δ_C (ppm)
1	0.95 (3H, t, J 7.5 Hz)	14.0
2	1.48 (2H, sext., J 7.5 Hz)	23.4
3	2.20 (2H, q, J 7.5 Hz)	36.4
4	6.09-6.03 (1H, m)	142.7
5	6.53-6.48 (1H, m)	128.0
6	6.53-6.48 (1H, m)	141.8
7	-	132.5
8	1.92 (3H, s)	12.5
9	7.44 (1H, d, J 15.0 Hz)	150.9
10	6.15 (1H, d, J 15.0 Hz)	118.0
11	-	166.8
12	-	177.7
13	6.19 (1H, s)	100.6
14	-	209.0
15	4.38 (1H, dd, J 10.0 and 6.5 Hz)	68.6
16	2.76 (1H, dddd, J 11.5, 9.5, 8.5 and 5.5) 2.26 (1H, dddd J 11.5, 10.0, 8.5 and 6.5)	21.8
17	4.21 (1H, q, J 8.5) 3.50 (1H, td, J 9.5, 5.5)	54.9



Supplementary Table 6. ^{1}H and ^{13}C NMR data for compound 2 (d^4-MeOD)

No.	δ_H (ppm)	δ_C (ppm)
1	0.94 (3H, t, J 7.5 Hz)	10.1
2	1.60-1.53 (2H, m)	31.1
3	4.10 (1H, q, J 6.5 Hz)	74.4
4	5.93 (1H, dd, J 15.0 and 6.5 Hz)	142.1
5	6.66 (1H, ddd, J 15.0, 11.5 and 1.0 Hz)	126.8
6	6.41 (1H, d, J 11.5 Hz)	138.4
7	-	134.3
8	1.92 (3H, s)	12.6
9	7.23 (1H, d, J 15.5 Hz)	147.1
10	6.07 (1H, d, J 15.5 Hz)	120.1
11	-	171.5



Supplementary Table 7. ^{1}H and ^{13}C NMR data for compound 1 (d^{4}-MeOD)

No.	δ_H (ppm)	δ_C (ppm)
1	1.10 (3H, t, J 7.5 Hz)	8.5
2	2.69 (2H, q, J 7.5 Hz)	35.0
3	-	203.7
4	6.40 (1H, d, J 15.0 Hz)	132.5
5	7.65 (1H, dd, J 15.0 and 11.5 Hz)	145.5
6	6.56 (1H, d, J 11.5 Hz)	135.7
7	-	143.7
8	2.07 (3H, s)	13.1
9	7.27 (1H, d, J 15.5 Hz)	138.4
10	6.29 (1H, d, J 15.5 Hz)	123.8
11	-	170.6

Supplementary Table 8. Optimization of *in vitro* ClpP1P2 and Cac16/17 activity.
The activity of a variety of substrates and the effects of activator peptides were measured. % activity was calculated based on rate of the reaction, normalized to that with Suc-LLVY-AMC substrate and no activator peptide.

		% activity	
Substrate	**EcClpP**	**ClpP1/P2**	**Cac16/17**
Suc-LLVY-AMC	100	100	100
Suc-LY-AMC	4251.7	14	0.7
Z-GGL-AMC	178	0	27.2
Ac-DEHD-AMC	0	0	0
Suc-AAPF-AMC	0	0	0
Suc-AFK-AMC	0	0	0
Activator			
Z-LLNva-CHO (MG115)	ND	88.2	7.2
Z-LLF-CHO	ND	109.3	6.1
Z-LL-CHO	ND	127.2	18.1

Supplementary Table 9. *In vitro* processing of ClpPs.
Observed masses were determined by whole protein MS after incubation alone or with cognate ClpPs, as listed. Incubations were carried out in the presence or absence of ADEP to determine its effect.

	Incubated with	**Predicted unmodified mass (Da)**	**Observed mass (Da)**	**Predicted processing**	**Processed without ADEP?**
ClpP1	Alone	23765.83	23634.82	Loss of met	N/A
	ClpP2		23634.78	Loss of met	N/A
	Cac17		23634.78	Loss of met	N/A
ClpP2	Alone	26058.03	26058.57	None	N/A
	ClpP1		23593.90	Pos 24-234	Y
	Cac16		23593.91	Pos 24-234	Y
Cac16	Alone	24307.13	24176.70	loss of met	N/A
	ClpP2		22461.96	pos 19-226	Y
	Cac17		22461.95	pos 19-226	Y
Cac17	Alone	27022.58	26820.50	loss of met-ala	N/A
	ClpP1		23851.92	Pos 33-249	Y
	Cac16		23851.92	Pos 33-249	Y

Supplementary Table 10. Primers used in this study.

Oligo name	Sequence	Purpose
cac-scrn1-F	CTCCTCCATCTCACCGGTT	PCR screening TAR colonies
cac-scrn1-R	CGAATGTACGCAGGAAGTCT	
cac-scrn2-F	CATCGACCTGAAGCTGGAG	PCR screening TAR colonies
cac-scrn2-R	GTGCCCCACCAGTAGATTTG	
Cac16del-block-F	CCGAGTTGCGGCCGCGGCGGCTGACGAGGTGCGGGCC GACGTTTAAACTGCAGCTCACGG	amplification of apra recombineering casette for cac16-17 deletion
Cac16del-block-R	ACCTCCGAGGACGCCATGGCCCGGGGCGAGATCACCT GGTGTTAACGCCAATCGACTGGC	
Leu2-BSG	ACTCGCAACCCCTATATACAAAGCTTATGTCTGCCCCT AAGAA	Leu2 cloning into pSASS5
Leu2-BSG-R	CTGACATCAGCAGAGGACATCCGCGGTTAAGCAAGGA TTTTCTTAACTTC	
His3-BSG-F	ACTCGCAACCCCTATATACAAAGCTTATGACAGAGCAG AAAGC	His3 cloning into pSASS4
His3-BSG-R	CTGACATCAGCAGAGGACATCCGCGGTACATAAGAAC ACCTTTGGT	
Cac8-A26-F	CCGTCGCCTCATCGTGTAC	amplification of A26 promoter with homology arm to cac8
Cac8-A26-R	CCCTGATGACTTGACGCTCGAGAAGGTTCACGATTAGC ATGCCGGTTACGGACCTTTGAGGAGGGCTCAATGAGCA AGCAGGCGCCGTT	
Cac9-A35-F	GCGTCCGCCCTTGCCCCATCCCCGAAATACACGCTACA GTGTCGACAACCTACGAACGAGGAGGCCCCATTTGAAT TCATCCGAGTGGTC	amplification of A35 promoter with homology arm to cac9
Cac9-A35-R	GAAGAAGAGGTTGAGCGCC	
Cac9-His-F	CGACGTGCACCAACTTGC	amplification of His casette from pSASS4
Cac9-His-R	TGGGGCAAGGGCGGACGCCGCGCTGCGGATATTTCTA AG	
cac14-xnr-F	ATTCATCTCTTGTATGGTTGACAACCGCGCCGCCGGCC CGAC	amplification of xnr1700 promoter from *S. albus*
xnr-R	GGGTTCCTCCTCATGAGTCC	
cac14-F	GGACTCATGAGGAGGAACCCATGACCGACCACAGCCG TGACGGC	amplification of cac14 right homology arm
cac14-xnr-F	GTGAGGCGGTAGTGGTCGAGGG	
Cac14-leftarm-F	AACCCTCCACGCACAGAAAGCGAG	amplification of cac14 left homology arm overlapping with xnr1700 promoter
Cac14-leftarm-R	CAGCGCCTGCTTCCAGGCG	
Cac14-Leu-F2	GCCTGGAAGCAGGCGCTGGCCGCGCTCGCCCAGGGCT GAATTTACCACCACAAAAGTTGA	amplification of Leu casette from pSASS4
Cac14-Leu-R	GTCGGGCCGGCGGCGCGGTTGTCAACCATACAAGAGA TGAAT	
pSETerm-16-F	AGGATCTCTAGAGGATCCCCAGTGGCTTACCAGGACGG	pSET152-ermEp*:cac16n17 cloning
pSET-17-R	GATTACGAATTCGATATCTCAGACATGGCGCCCGGC	
pIJ-Cac15Trunc-F	GGTAGGATCGTCTAGAACAGGAGGCCCCACATATGCTC GCCGGGGTCAGT	pIJ10257-cac15 cloning

Oligo name	Sequence	Purpose
pIJ-cac15-5'-R	CTCCGCCAGGGAGTCCGACGTCGACCGGCATG	pIJ10257-cac15 cloning
Sc-popR-F2	GCTGAAGGAGGTCGCCGAG	RTPCR
Sc-popR-R2	CCCAGGCCCAGTCCGAGG	
ScoClpP1-F1	CGACATCGCCAACAAGATCAC	RTPCR
ClpP1-R1	ACGTCGTTCTTGATGTACTGCA	
Sc-ClpP3-F1	CCGACGCTCAGCAGGAAC	RTPCR
Sc-ClpP3-R1	CGACATCAGCCTGTACGTCAA	
cac-8-F1	CATCGCGAAGTACTGTCCGA	RTPCR
cac-8-R1	TATCCGTCGGTGGCCATCAT	
cac9-F8	TCTACCACAGCACCGAGGAC	RTPCR
cac9-R8	TCTGGAGCGGTTCGGAGTAG	
cac14-F1	GGGTTGTACGATCCGGATCC	RTPCR
cac14-R1	ATACCGAAGAACTCCGCGTC	
cac15-F1	TGCCGTCAGATCCGTCAAAA	RTPCR
cac15-R1	ACGCCCTATGAGCGCAAATA	
cac16-F2	AGAACGACGTGGTGACCATC	RTPCR
cac16-R2	TCAGCTGTTCCGCGTAGATC	
hrdB-coel-F3	GAGGACAAGCTGGCGAACT	RTPCR both *S. coelicolor* and *S. cattleya*
hrdB-coel-R3	AGCCCTTGGTGTAGTCGAAC	

Supplementary Table 11. gBlocks used in this study

gblocks	Sequence	Use
gblock-cac	CCTCCCATGGTATAAATAGTGGCAAGCTTGCCGCCGGTCTGC CTCTCTGTTCAACGAGTGTTGGTGTTCCGCCCGCGGGTTTAA ACCAAGGCATTCGCCCCGGAGTACCAGGGCGCCCTCAGCTC GCTGTCGGTCAAGCTTTATGTCGAAAGCTACATATAAGGAAC	TAR cloning homology arms
gblock-apra-cac5	GTCCCAGGTGAGCGACGGGGTGGTGGCACGCCGGGGCATAC GTCACCTCTTCAACTCAGATACTAGTAACTCCCCCAGTCCTG CACGCTGTCGTATTCTCCTGGCCACGACTTTACAACACCGCA CAGCATGTTGTCAAAGCAGAGACCGTTCGAATGTGAACAGG ATCCCGCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCA AGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTGCT CAGCGGTGGCAGCAGCCAACTCAGCTCTAGAGTCGACCTGC AGCCCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGAC TGGGAAAACCCTGGCGTTACCCAACTTAATCGTTAACGCCAA TCGACTGGCGAGCGGCATCGCATTCTTCGCATCCCGCCTCTG GCGGATGCAGGAAGATCAACGGATCTCGGCCCAGTTGACCC AGGGCTGTCGCCACAATGTCGCGGGAGCGGATCAACCGAGC AAAGGCATGACCGACTGGACCTTCCTTCTGAAGGCTCTTCTC CTTGAGCCACCTGTCCGCCAAGGCAAAGCGCTCACAGCAGT GGTCATTCTCGAGATAATCGACGCGTACCAACTTGCCATCCT GAAGAATGGTGCAGTGTCTCGGCACCCCATAGGGAACCTTT GCCATCAACTCGGCAAGATGCAGCGTCGTGTTGGCATCGTGT CCCACGCCGAGGAGAAGTACCTGCCCATCGAGTTCATGGAC ACGGGCGACCGGGCTTGCAGGCGAGTGAGGTGGCAGGGGCA ATGGATCAGAGATGATCTGCTCTGCCTGTGGCCCCGCTGCCG CAAAGGCAAATGGATGGGCGCTGCGCTTTACATTTGGCAGG CGCCAGAATGTGTCAGAGACAACTCCAAGGTCCGGTGTAAC GGGCGACGTGGCAGGATCGAACGGCTCGTCGTCCAGACCTG ACCACGAGGGCATGACGAGCGTCCCTCCCGGACCCAGCGCA GCACGCAGGGCCTCGATCAGTCCAAGTGGCCCATCTTCGAG GGGCCGGACGCTACGGAAGGAGCTGTGGACCAGCAGCACAC CGCCGGGGGTAACCCCAAGGTTGAGAAGCTGACCGATGAGC TCGGCTTTTCGCCATTCGTATTGCACGACATTGCACTCCACC GCTGATGACATCAGTCGATCATAGCACGATCAACGGCACTG TTGCAAATAGTCGGTGGTGATAAACTTATCATCCCCTTTTGC TGATGGAGCTGCACATGAACCCATTCAAAGGCCGGCATTTTC AGCGTGACATCATTCTGTGGGCCGTACGCTGGTACTGCAAAT ACGGCATCAGTTACCGTGAGCTGCAGTTTAAACGGGGCCTCC GTGGGTCGGCGAAGGTCGCGGCGAGCGGGT	cac5 recombineering apra casette

gblocks	Sequence	Use
cac16-gblock	ATGGCGTATCAGGACGGACGCGTTGCTGCAGCAGCAGCCGA TGAGGTCCGCGCTGATGGTGGCAGCCCCGCATGGGCAGTTTC CGACCAGGTGTTTAACCGCCTTCTGAAGGACCGTATTATCTT TCTGGGTCAGGCTGTGGACGATGACATTGCTAACAAGTTAAC CGCCCAGTTGCTGATGCTGGCTGCCGATTCACAGGAAGACAT CTATTTGTACATTAATTCTCCGGGTGGATCTGTCACTGCAGG AATGGCCGTGTACGACACTATGCAGTATATCAAAAATGATG TCGTAACTATTGCTATGGGATTCTGTGCTAGCATGGGTCAGT TCCTTCTTACTGCAGGAACCGCCGGCAAACGTTTTGCTCTTC CGCATACCAAGATCTTAATGCACCAACCGTCTGCGGGGTTGG CGGGTTCCGCGTCAGACATCAAAATTTATGCAGAACAGTTA ATCCGCACTAAGAAAGAGATGGCCGAGCTGATTGCGGTCCA TAGCGGACAAGATGTCGAAAAAATCACGAAGGACAGCGATC GTGATCGCTGGTTCACCGCTCAAGAGGCCAAAGAGTATGGG CTTATCGACGAGGTGATGACTAGCACTGTTGGCGTCCCCGGT GGCGGGGGCACAAAAACTGAT	pET28-cac16 protein expression
cac17-gblock	ATGGCACCGCATAGTCATCCCGCTGCTCCGCATCAGGGCCCA GGTGCTCCATTTCAAGCATTAGCTCCTGCCGGCCACCTTATG CCGCAAGCACGTTATGTAGTACCACACTTCGTAGAACGCAC GTCCCAAGGGGTGCGCGAGTACGACCCTTATGCTAAGTTGTT CGAGGAACGCGTTGTGTTCCTTGGAACGCAGGTCGATGACA CTAGCGCGAACGACATTATGGCTCAATTACTGTGCTTAGAGT CGACGGACCCGGATCGCGATATCTCTATTTATATTAACTCAC CTGGCGGAAGTATGACTTCGCTTATGGCAATTTACGATACTA TGCAATTTATTCGTCCGCAAATTCAAACGGTGTGCATCGGAC AGGCTAGCAGTGCTGCAGCAGTCATCTTGGCGGCAGGAACG CCGGGGAAGCGTTGTGCCCTGCCGAATGCCAAAATTCTTATC CATCAGCCTGCGTCCCAGGGATCAGAGGGACAAGTAAGCGA TTTGGAGATCCAGGCCGCCGAAATCTTGCGCGTGCGCAGTCT GGTAGAGACGATTCTTGCAAAACACACCGGCCGTGAAGTAG AGCAGGTGCGTGAAGATATTGAGCGCGACAACATCTTAACT GCACAACAGGCTATGGAGTATGGAATTGTTGACCAAGTTAT CTCCCCACGTGCCATGGCATCTTCAGAAGTCGCTGGTCGCCA CGTA	pET28-cac17 protein expression
clpP1-gblock	ATGACTACTCCGCAGATTGAAATGCCAGGGCTGCTTATGCCA ACTGCAAACGGAGCTCCCAGTGGCCCTGGGTTGGGCGACGC TGTTTACAATCGTTTGCTGGATGAGCGCATCATTTTCTTGGG ACAAGCTGTAGATGATGATATCGCTAATCGTATTTGTGCTCA AATGTTGTTACTTGCCGCCGACCCGACCAAAGATATTTACCT GTATATTAACTCCCCTGGGGGTTCAGTGACGGCTGGATTCGC AATCTATGACACCATGCAGTATATCAAGAACGACGTGGTCA CAATCGCAATGGGCTTGGCTGCCAGTATGGGACAGTTCTTAT TGACGGCTGGGACTCCTGGTAAACGTTTCGCCCTTCCGAATG CAGACATCCTGATGCATCAAGGGAGTGCAGGGCTGGGCGGC AGCGCGAGCGACATTAAAATCCAGGCAGAACAACTTTTACG CACTAAGAAACGCATGGAACAACTGACGGCCCAACACTCCG GGCAATCGGTGGAGACGATTAACCGTGACGCCGATCGCGAC CGCTGGTTTACGGCCGAGGAAGCGAAGGAATACGGACTTAT TGATGAGGTAATGCACACTGCTGCAGATGTCCCAGGAGGGG GTGGTACCGGAGCG	pET28-clpP1 protein expression

gblocks	Sequence	Use
clpP2-gblock	ATGAACACACGCAATCCCCTGGCTTTGCAGCAAGGGGTAGA CGCGTCCCGCAGTCATGGCATGGAGGCCCGCTATATTGTGCC GCGCTTTGTCGAGCGTACCAGTCAGGGTATCCGCGAATACG ACCCCTATGCAAAATTGTTTGAAGAGCGTGTAATTTTTCTTG GCGTACAAATTGATGACGCTTCAGCCAACGATGTGATGGCC CAATTGTTATGCCTGGAATCAATGGACCCTGATCGTGACATC TCCATTTACATTAACAGCCCAGGAGGTAGTTTTACCGCTATG ACGGCTATTTATGACACAATGCAATTCGTCAAACCGGACATT CAGACTGTGTGCCTTGGTCAGGCGGCTTCTGCAGCAGCAGTG TTACTTGCGGCTGGCACCCCAGGGAAGCGTATGGCGCTTCCG AATGCGCGCGTTTTGATTCATCAACCATACACGGAGACAGG CCGCGCACAGATTAGCGACTTGGAGATTCAGGCTAACGAGA TTACTCGCATGCGTGTTCAGATGGAGGAGATGCTGGCCAAG CACTCGCATCGTACTGAAGAACAAATTCGCGAGGACATTGA ACGCGACAAGATTCTGACTGCCGAAGATGCCTTATCGTTTGG ATTCGTGGATCAGATCACAAGCTCACGCAAGGCCGCGTTATC GGCA	pET28-clpP2 protein expression

Supplementary Table 12. Strains used in this study.

Plasmid name	Description	Source
Plasmids		
pCGW	Engineered copy number controlled capturing vector for TAR, *aac(3)IV ura3 cen/ARS trp1 sopABC repE ori2 oriV cat neo traJ-oriT attP-int$_{\Phi C31}$*	Xu et al., submitted
pCGW-*cac*	Full *cac* cluster	this work
pCGW-*cac*-L	*leu2 xnr1700*p:*cac14*	this work
pCGW-*cac*-LH	*leu2 xnr1700*p:*cac14 his3* A26:*cac8* A35:*cac9*	this work
pCGW-*cac*-LHK	*leu2 xnr1700*p*:cac14 his3 A26:cac8 A35:cac9 aac*(3)IV *kasO*p**:cac5*	this work
pCGW-*cac*-LHK-hpa	*leu2 xnr1700*p*:cac14 his3 A26:cac8 A35:cac9 kasO*p**:cac5*	this work
pCGW-*cac*-LHK-Δcac16-17	*leu2 xnr1700*p*:cac14 his3 A26:cac8 A35:cac9 kasO*p**:cac5 Δcac16-17*	this work
pSASS5	*pMB1** amp^r ENO2prom VPS13term	Gifted by Dr. Mike Tyers
pSASS4	*pMB1** amp^r CRH1prom GRE3term	Gifted by Dr. Mike Tyers
pIJ10257-*cac15*	pIJ10257 backbone, *ermE*p* driving expression of *cac15*	this work
pSET152-*ermE*p**:cac16n17*	pSET152 backbone, *ermE*p* driving expression of *cac16-17*	this work
pUZ8002 neo::bla	pUZ8002 derived *E. coli-Streptomyces* conjugation helper plasmid through replacing the *neo* with *bla*, ampicillin resistant	
pR9406	pUB307 derived *E. coli-Streptomyces* conjugation helper plasmid through replacing the *neo* with *bla*, ampicillin resistant	[54]
pET-28a(+)	protein expression	Novagen
pKD46	*gam-bet-exo*, λ-red recombination	[55]
Strains		
S. coelicolor M1154	*Δact Δred Δcpk Δcda rpoB[C1298T] rpsL[A262G]*	[56]
S. cattleya DSM 46488		ATCC
S. venezuelae ATCC 10712		ATCC
S. aureus ATCC 29213		ATCC
M. smegmatis MC^{2}155		ATCC
B. subtilis 168	*trpC2*	laboratory strain
B. subtilis 168 Δ*clpP*	*trpC2 ΔclpP::*spc^r	Gifted by Dr. Heike Brotz Oesterhelt
S. cerevisiae VL6-48N	MAT α, *his3-D200*, *trp1-Δ1*, *ura3-Δ1*, *lys2*, *ade2−101*, *met14*, *psi+cir^O*, TAR host strain	Gifted by Vladmir Larinov
S. cerevisiae SASy31	*MATalpha his3Δ leu2Δ lys2Δ trp1Δ::pACT1-Z3EV-ENO2term ura3Δ*	Gifted by Dr. Mike Tyers
S. cerevisiae SASy35	*MATa his3Δ leu2Δ lys2Δ trp1Δ::pACT1-Z3EV-ENO2term ura3Δ met15Δ*	Gifted by Dr. Mike Tyers
E. coli Top10	General cloning and plasmid maintenance	Invitrogen
E. coli EPI300	Inducible trfA controlling oriV	Gifted by Dr. Ming Jiang
E. coli ET12567	*dam-13::Tn9 dcm-6 hsdMRS cat*	[57]
E. coli BW25113	recombineering targeting system host strain	laboratory strain
E. coli BL21(DE3) 1146D	BL21(DE3) Δ*clpP::*cm^r	Gifted by Dr. Walid Houry

Chapter V – Conclusions

SUMMARY

The work presented in this thesis describes three unique, hypothesis-driven approaches to antibiotic discovery from Actinomycetes. Each of these strategies uses the information encoded in BGCs in a different way, whether it be to guide dereplication efforts (Chapter 2), avoid known resistance mechanisms (Chapter 3), or target a desired mode of action (Chapter 4). By doing so, these three chapters address the barriers to novel antibiotic discovery: disrupting streptothricin and streptomycin's biosynthesis (Chapter 2) addresses the dereplication challenge, while the evolution-guided discovery of corbomycin (Chapter 3) and target-directed discovery of ClpP targeting azabicyclenes (Chapter 4) address BGC prioritization. Along the way, strategies for cryptic BGC activation, such as manipulating media composition, overexpressing transcriptional activators, and employing heterologous expression, were employed. Direct application of the tools developed here and the molecules we describe will no doubt be useful, but there are also general themes and lessons to be taken from these three stories. In this concluding chapter, I will explore the successes and failures of each technique and suggest future prospects for continued discovery efforts.

EVALUATING THE SUCCESS OF BGC-GUIDED ANTIBIOTIC DISCOVERY

The premise of this thesis, as described in Chapter 1, was to overcome the major challenges in novel antibiotic discovery from actinomycetes: dereplication of known compounds and access to cryptic clusters. Were the strategies employed successful and where is there room for improvement?

CRISPR-Cas9 knockout of common antibiotics

To start, the CRISPR-Cas9 platform for removal of nuisance antibiotics was largely successful in generating knockouts and uncovering rare antibiotics from these strains. The strength of the platform is in its numbers: since such as large proportion of strains produce a handful of known compounds, it is applicable to thousands of strains in historic collections. Therefore, it is not necessary that knockout is successful in every strain in order to salvage the biosynthetic potential of a large numbers of isolates. However, to have strength in numbers requires that the platform be applied to more than just streptothricin and streptomycin, and more than just 12 *Streptomyces* spp, as described in Chapter 2. As it stands, the low throughput of the platform limits discovery of novel compounds to known, nonetheless rare, ones.

The major bottlenecks that limit throughput of CRISPR in actinomycetes are relatively low editing efficiency and conjugation frequencies, both of which contribute to limited multiplex editing capability. Variation in conjugation frequency between actinomycete strains is at least partially attributable to differences in their restriction-modification (RM) systems, where exogenous DNA is degraded by restriction enzymes that cleave highly specific DNA sequences based on methylation profile. In 1997, this observation led to adoption of the conjugal donor strain *E. coli* ET12567 which lacks three common methyltransferases (*dam, dcm-6, hsdM*) and can improve conjugation efficiency $>10^4$ fold[106]. However, *E. coli* ET12567 still does not protect against RM systems that cleave un-methylated motifs and few advancements in this area have been made. Recently, Johnston *et al.* demonstrated that by systematically removing all possible

restriction target motifs from a nucleotide sequence, DNA molecules that are invisible to RM systems can be introduced into previously intractable *Staphylococcus* strains[107]. A similar approach could be applied to the common actinomycete RM recognition motifs to likewise improve conjugation efficiency.

Relatively low editing efficiency and large genomic deletions in some of our engineered strains could also be due to off-target effects of sgRNAs leading to aberrant double strand breaks and associated Cas9 toxicity. Since our original study, improved CRISPR platforms boasting multiplexed editing and reduced off-target effects have been developed in *Streptomyces*[108,109]. One of these, called CRISPR-BEST, avoids the introduction of double strand breaks entirely by fusing Cas9 nickase to uracil glycosylase inhibitor and cytidine deaminase[109]. After localized target binding of the sgRNA/Cas9n complex, cytidine deaminates to uracil, uracil glycosylase inhibitor blocks nucleotide excision repair, and conserved cellular mismatch repair converts U:G to U:A. In this way, point mutations and stop codons can be introduced without introducing double strand breaks and avoiding large genomic rearrangements.

Even with these suggestions, it is likely that further innovation would be required to fully optimize a CRISPR based knockout strategy for common antibiotics. If realized, however, it could be applied in multiplex to knockout the top nuisance antibiotics in high throughput to facilitate the discovery of novel compounds.

Evolution-guided BGC prioritization

Evolution-guided discovery of a new functional class of GPAs was arguably the most successful of the three approaches described in the thesis in terms of novelty of the

antibiotics discovered and their mechanism of action. The incremental improvement on previous phylogeny-guided approaches by searching for the lack of known resistance genes was key to steering our efforts towards unknown modes of action. Inherent in the unknown is the risk that this strategy will yield molecules that are not antibiotics at all, and so undoubtedly some 'luck' is involved. However, we posit that our approach is not completely agnostic to mechanism, as an antibiotic's scaffold is selected through evolution to be biologically active and so phylogenetically related BGCs may also yield molecules whose biological function is related. This was indeed the case for complestatin and corbomycin, but it would be interesting to see if the same holds true for other antibiotic families. The only requirement for applying evolution-guided genome mining is sequencing data and that the resistance mechanism is specific to a mechanism of action (e.g. target modification is specific, efflux is not). There are dozens of examples of antibiotic classes using specific resistance mechanisms that could therefore be applied (e.g. tetracyclines - ribosomal protection, ansamycins - *rpoB* duplication, macrolides - 23S rRNA modification, coumermycins - duplicated *gyrB*).

Corbomycin and complestatin have clinical potential but come with several liabilities that must be addressed before their use. Namely, protein binding, poor solubility, and gram-negative impermeability restrict their use. These problems can be addressed through semi-synthesis and is encouraged by the fact that derivatization of traditional GPAs has resulted in FDA approval of three next-generation drugs, telavancin, dalbavancin and oritavancin, with superior pharmacokinetics and activity profiles[110]. However, in our unpublished work, we have had limited success so far improving

complestatin and corbomycin, owing partly to the lack of chemical handles for derivatization. Continuing to explore the chemical space of this family of GPAs through genome mining may provide an improved or more amenable scaffold to base these efforts upon.

Target-directed genome mining for ClpP directed NPs

In sharp contrast to evolution-guided discovery of complestatin and corbomycin, target-directed genome mining relies on the presence, rather than absence, of a predicted resistance gene. This allows researchers to exploit specific targets of interest, and in addition to target duplication might also be feasibly applied to target modifying resistance proteins if they are known (e.g. rRNA modifying enzymes). However, even when applied to these two forms of resistance, target-directed genome mining limits researcher's discovery potential and projects preconceived notion about what resistance should look like. Perhaps as a perfect illustration of how resistance mechanisms can be more intricate than predicted *a priori*, we identified a family of ClpP associated BGCs that do not appear to encode antibiotics. While further work is required to fully elucidate the azabicyclenes' biological function, their association with ClpP ensures that they are biologically interesting molecules that will perhaps have other useful properties.

FUTURE PROSPECTS FOR EXPLORING THE UNKNOWN

Every antibiotic discovery campaign must strike a balance between using prior knowledge to guide effort towards interesting compounds, while remaining unbiased to what is possible. This balance between known and unknown is positioned differently depending on the purpose of the study. For example, it is often the goal to expand the

chemical diversity within a known antibiotic class, and so by using known BGCs or resistance mechanisms, unknown molecules can be discovered (e.g. [50,111]). The three chapters described in this thesis also fall to different extremes along this spectrum: CRISPR-inactivation of nuisance BGCs works with the known-knowns (known mechanism, known molecule), target-directed genome mining towards known-unknowns (known mechanism, unknown molecule), and evolution-guided genome mining pushes the limit of unknown-unknowns (unknown mechanism, unknown molecule). It is this far, unknown side of the spectrum that is the most difficult to access, but also where the most interesting compounds are found. As in the discovery of corbomycin, hypotheses about the evolution of BGCs based on their ecological role could be used to guide discovery of novel, biologically interesting compounds with the fewest preconceived notions about mechanism. Based on this premise and lessons learned in preparing this thesis, here I will pose relevant questions and ideas about how this unknown territory could be explored in the future.

BGC abundance and diversity

It is well established that there exists a vast array of BGCs in the environment, that BGC diversity and richness varies from location to location[112,113], and that a large proportion of BGCs do not resemble anything previously known[10]. Tapping into unknown BGC families is no doubt an excellent way to discover novel chemical matter, but it requires refinement if one is to discover biologically interesting compounds. Evolution-guided discovery of corbomycin and complestatin is a contrasting example of how known NP families can encode molecules that are biologically active in new ways. Our

evolution-guided strategy relied on resistance genes and phylogenetic structure, but other indicators of evolutionary forces might also be useful markers of BGC function.

Firstly, it was interesting to note neighbouring branches surrounding complestatin and corbomycin BGCs with potential to share a common mechanism of action. These bunches of BGC diversity built around a biologically active scaffold may be the result of evolutionary selection on a BGC to probe the chemical space surrounding a useful molecule. Thus, BGC diversity within a family could be used as a marker to prioritize BGCs more likely to have interesting bioactivity.

Secondly, it is interesting to think about the relative abundance of different BGCs/BGC families. What is more interesting: common BGCs that remain uncharacterized, or BGCs that are rarely found? On the one hand, rare BGCs might encode novel scaffolds, but on the other hand, what might it mean about the compound's bioactivity? Limited spread of a BGC could simply be the result of lack of time or opportunity to disseminate, but it could also indicate lack the selective advantage to promote its spread and maintenance. Thus, the abundance of a BGC could be an indicator of physiological importance, making common, uncharacterized BGCs more likely to be interesting. Perhaps these molecules only remain uncharacterized because the BGC is usually cryptic or does not elicit a phenotype typically screened for. The azabicyclenes are a good example of the later. Still, care should be taken to avoid ubiquitous BGCs that are relatively uninteresting: molecules among the company of hopenes, siderophores, and ectoines. These housekeeping BGCs tend to be encoded in central regions of actinomycete genomes which are maintained more stably than peripheral regions of the

chromosome, and so chromosomal position can be yet another marker for BGC function[114].

A relevant case study that could answer some of these questions is that of streptothricin. Why is streptothricin so common? Is it actually so common in the environment, or is its apparent success an artefact of amenability to laboratory culturing? Is it common because of its evolutionary age or selective advantage? What is the ecological role of streptothricin: an antibiotic or signalling molecule? If it is acting as an antibiotic, how widespread is resistance in the environment? Another relevant case study is corbomycin. How does it measure up in terms of rarity and family diversity? Is the combination of relative rarity but diversity within a family a unique marker of interesting families like corbomycin's? What is the ecological function of corbomycin, anyway? Is it truly an antibiotic, or perhaps meant to protect from lytic phages or regulate spore germination?

Broadening the scope of these questions, does biological function, both in an ecological and clinical sense, correlate with BGC abundance? To answer this last question, systematic genome mining that compares the abundance of different NP families with their biological function might be useful. While surveys of BGC diversity have been made in the past[10,112,113], they lack either the fine grained categorization beyond broad NP families that would be required, or complete measurements of gene cluster similarity beyond short sequence tags.

BGC regulation

A BGC family's abundance and diversity are just two readouts of how evolution acts on the encoded metabolites' function. Regulation of the BGC is also impacted by its function, and so could likewise be used to predict a BGC's biological role. In some cases, the link between regulation and function is obvious, for example iron limitation inducing siderophore biosynthesis. It is also sometimes easy to present compelling hypotheses, for example the production of protective antibiotics by plant endophytes in response to plant hormones[114]. In most cases, however, regulatory networks are vague, with onset coinciding with morphological development, and so shed little light on the metabolite's specific role. A broader question therefore might be to ask why some clusters are transcriptionally silent while others are not. Are some families of BGCs always turned "on" and others always turned "off"? Similar to the BGC rarity line of questioning, are BGCs that are more often "on" more biologically useful, or are those usually turned "off" interesting because they are more likely novel?

A couple of studies have started to answer these questions. One transcriptomics study comparing the expression of shared BGCs in different *Salinispora* genomes found that BGC conservation was correlated with higher expression[115]. My anecdotal experience supports this: streptothricin, a common BGC, is usually robustly expressed. In some cases of silent BGCs they saw evidence of BGC silencing (e.g. loss of a transcriptional activator), but in other cases, nearly identical BGCs in different strains were differentially expressed. This suggests that the same compounds might be expressed in response to different environmental cues and so perhaps play different roles in different

strains. Another metatranscriptomic study directly extracted RNA from the soil, sampling 120 time points in a microcosm manipulation experiment, allowing for a truer representation of expression networks in an ecological setting[116]. More large-scale studies such as these will help to decipher regulation and how it relates to BGC function, likewise informing which BGCs should be prioritized.

CONCLUDING REMARKS

As the world faces unprecedented crisis in the year 2020, it is more important that ever not to lose sight of the growing antibiotic resistance problem. Fighting antibiotic resistance and spurring a renaissance of the "Golden Age" of discovery will require new techniques inspired by modern technology. Adding to this toolbox of creative techniques, this thesis presents three approaches to antibiotic discovery guided by BGCs in actinomycetes. On top of these advances in BGC dereplication and prioritization, accelerating progress in this field requires improved understanding of BGC regulation and ecological function as well as genetic techniques to more easily access cryptic clusters. The good news is that there is no shortage of chemical diversity in nature, and by unlocking this biosynthetic potential, clinically useful drugs will undoubtedly follow.

REFERENCES

1. Council of Canadian Academies. *When Antibiotics Fail*. (2019).
2. Fleming, A. On the antibacterial action of cultures of a penicillium with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* **10**, 226 (1929).
3. Wright, G. D. Something old, something new: revisiting natural products in antibiotic drug discovery. *Can. J. Microbiol.* **60**, 147–54 (2014).
4. The Pew Charitable Trusts. Antibiotics Currently in Clinical Development. (2019).
5. Bérdy, J. Thoughts and facts about antibiotics: Where we are now and where we are heading. *J. Antibiot. (Tokyo).* **65**, 441–441 (2012).
6. Gomez-Escribano, J. P. & Bibb, M. J. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: From genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**, 425–431 (2014).
7. Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. **11**, (2015).
8. Blin, K., Kim, H. U., Medema, M. H. & Weber, T. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief. Bioinform.* **20**, 1103 (2019).
9. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).
10. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).
11. Cruz-Morales, P. *et al.* Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes. *Genome Biol. Evol.* **8**, 1906–1916 (2016).
12. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
13. Kautsar, S. A. *et al.* MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **48**, D454–D458 (2019).
14. Wang, G., Tang, W. & Bidigare, R. R. Terpenoids as therapeutic drugs and pharmaceutical agents. in *Natural Products: Drug Discovery and Therapeutic Medicine* 197–227 (Humana Press, 2005). doi:10.1007/978-1-59259-976-9_9
15. Fischbach, M. A. & Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics : Logic , Machinery , and Mechanisms. **5**, 3468–

3496 (2006).

16. Röttig, M. *et al.* NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, 362–367 (2011).

17. Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: Expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **45**, W49–W54 (2017).

18. Baltz, R. H. Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J. Ind. Microbiol. Biotechnol.* **33**, 507–513 (2006).

19. Bobzin, S. C., Yang, S. & Kasten, T. P. *LC-NMR: a new tool to expedite the dereplication and identification of natural products*.

20. Gaudêncio, S. P. & Pereira, F. Dereplication: Racing to speed up the natural products discovery process. *Natural Product Reports* **32**, 779–810 (2015).

21. Williams, R. B. *et al.* Dereplication of natural products using minimal NMR data inputs. *Org. Biomol. Chem.* **13**, 9957–9962 (2015).

22. Zani, C. L. & Carroll, A. R. Database for rapid dereplication of known natural products using data from MS and fast NMR experiments. *J. Nat. Prod.* **80**, 1758–1766 (2017).

23. Dictionary of Natural Products. (2020). Available at: http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml. (Accessed: 28th August 2020)

24. MarinLit - A database of the marine natural products literature. Available at: http://pubs.rsc.org/marinlit/. (Accessed: 28th August 2020)

25. AntiBase: The Natural Compound Identifier. Available at: https://www.wiley.com/en-us/AntiBase%3A+The+Natural+Compound+Identifier-p-9783527343591. (Accessed: 28th August 2020)

26. ChemSpider. Available at: http://www.chemspider.com/. (Accessed: 28th August 2020)

27. PubChem. Available at: https://pubchem.ncbi.nlm.nih.gov/. (Accessed: 28th August 2020)

28. Guijas, C. *et al.* METLIN: A technology platform for identifying knowns and unknowns. *Anal. Chem.* **90**, 3156–3164 (2018).

29. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).

30. Gurevich, A. *et al.* Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. *Nat. Microbiol.* **3**, 319–327 (2018).

31. Mohimani, H. *et al.* Dereplication of microbial metabolites through database search of mass spectra. *Nat. Commun.* **9**, 1–12 (2018).

32. Mohimani, H. *et al.* Automated genome mining of ribosomal peptide natural products. *ACS Chem. Biol.* **9**, 1545–1551 (2014).

33. Wong, W. R., Oliver, A. G. & Linington, R. G. Development of antibiotic activity profile screening for the classification and discovery of natural product antibiotics. *Chem. Biol.* **19**, 1483–1495 (2012).

34. Zubyk, H. L., Cox, G. & Wright, G. D. Antibiotic dereplication using the antibiotic resistance platform. *J. Vis. Exp.* **2019**, (2019).

35. Cox, G. *et al.* A common platform for antibiotic dereplication and adjuvant discovery. *Cell Chem. Biol.* **24**, 98–109 (2017).

36. Liu, J.-T. *et al.* Bioactive natural products from the antarctic and arctic organisms. *Mini-Reviews Med. Chem.* **13**, 617–626 (2013).

37. Jiang, Z. K. *et al.* Xiakemycin A, a novel pyranonaphthoquinone antibiotic, produced by the Streptomyces sp. CC8-201 from the soil of a karst cave. *J. Antibiot. (Tokyo).* **68**, 771–774 (2015).

38. Fondi, M. *et al.* "Every gene is everywhere but the environment selects": Global geolocalization of gene sharing in environmental samples through network analysis. *Genome Biol. Evol.* **8**, 1388–1400 (2016).

39. Schorn, M. A. *et al.* Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology (United Kingdom)* **162**, 2075–2086 (2016).

40. Feling, R. H. *et al.* Salinosporamide A: A highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chemie - Int. Ed.* **42**, 355–357 (2003).

41. Gulder, T. A. & Moore, B. S. Chasing the treasures of the sea - bacterial marine natural products. *Current Opinion in Microbiology* **12**, 252–260 (2009).

42. Milshteyn, A., Colosimo, D. A. & Brady, S. F. Accessing bioactive natural products from the human microbiome. *Cell Host and Microbe* **23**, 725–736 (2018).

43. Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511–516 (2016).

44. Chu, J. *et al.* Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat. Chem. Biol.* **12**, 1004–1006 (2016).

45. Chiumento, S. *et al.* Ruminococcin C, a promising antibiotic produced by a human gut symbiont. *Sci. Adv.* **5**, 9969–9994 (2019).

46. AMANN & J. Die direkte Zahlung der Wasserbakterien mittels des Ultramikroskops. *Cent. f. Bakteriol.* **29**, 381–384 (1911).

47. Nichols, D. *et al.* Use of Ichip for high-throughput in situ cultivation of "uncultivable" microbial species. *Appl. Environ. Microbiol.* **76**, 2445–2450 (2010).

48. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).

49. Owen, J. G. *et al.* 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 (2015).

50. Hover, B. M. *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol.* **3**, 415–422 (2018).

51. Peek, J. *et al.* Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat. Commun.* **9**, 4147 (2018).

52. Kallifidas, D. & Brady, S. F. Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. in *Methods in Enzymology* **517**, 225–239 (Academic Press Inc., 2012).

53. Li, W. *et al.* Promysalin, a salicylate-containing pseudomonas putida antibiotic, promotes surface colonization and selectively targets other pseudomonas. *Chem. Biol.* **18**, 1320–1330 (2011).

54. Robbins, N. *et al.* Discovery of ibomycin, a complex macrolactone that exerts antifungal activity by impeding endocytic trafficking and membrane function. *Cell Chem. Biol.* **23**, 1383–1394 (2016).

55. Schmitt, E. K. *et al.* The natural product cyclomarin kills *Mycobacterium tuberculosis* by targeting the ClpC1 subunit of the caseinolytic protease. *Angew. Chemie* **123**, 6011–6013 (2011).

56. Gao, W. *et al.* The Cyclic Peptide Ecumicin Targeting ClpC1 Is Active against Mycobacterium tuberculosis In Vivo. *Antimicrob. Agents Chemother.* **59**, 880–889 (2015).

57. Gavrish, E. *et al.* Lassomycin, a ribosomally synthesized cyclic peptide, kills *Mycobacterium tuberculosis* by targeting the ATP-dependent protease ClpC1P1P2. *Chem. Biol.* **21**, 509–518 (2014).

58. Wu, S.-C., Liu, F., Zhu, K. & Shen, J.-Z. Natural products that target virulence factors in antibiotic-resistant *Staphylococcus aureus*. (2019).

doi:10.1021/acs.jafc.9b05595

59. Muñoz-Cazares, N., García-Contreras, R., Soto-Hernández, M., Martínez-Vázquez, M. & Castillo-Juárez, I. Natural products with quorum quenching-independent antivirulence properties. in *Studies in Natural Products Chemistry* **57**, 327–351 (Elsevier B.V., 2018).

60. Gehrke, S. S. *et al.* Exploiting the sensitivity of nutrient transporter deletion strains in discovery of natural product antimetabolites. *ACS Infect. Dis.* **3**, 955–965 (2017).

61. King, A. M. *et al.* Aspergillomarasmine A overcomes metallo-β-lactamase antibiotic resistance. *Nature* **510**, 503–6 (2014).

62. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147 (2002).

63. Challis, G. L. Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**, 219–32 (2014).

64. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big effects from small changes: Possible ways to explore nature's chemical diversity. *ChemBioChem* **3**, 619–627 (2002).

65. Moussa, M. *et al.* Co-culture of the fungus *Fusarium tricinctum* with *Streptomyces lividans* induces production of cryptic naphthoquinone dimers. *RSC Adv.* **9**, 1491–1500 (2019).

66. Shin, D. *et al.* Coculture of marine *Streptomyces* sp. with *Bacillus* sp. produces a new piperazic acid-bearing cyclic peptide. *Front. Chem.* **6**, 498 (2018).

67. Lee, N. *et al.* Iron competition triggers antibiotic biosynthesis in *Streptomyces coelicolor* during coculture with *Myxococcus xanthus*. *ISME J.* **14**, 1111–1124 (2020).

68. Onaka, H., Mori, Y., Igarashi, Y. & Furumai, T. Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl. Environ. Microbiol.* **77**, 400–406 (2011).

69. Pishchany, G. *et al.* Amycomicin is a potent and specific antibiotic discovered with a targeted interaction screen. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 10124–10129 (2018).

70. Kong, D., Wang, X., Nie, J. & Niu, G. Regulation of antibiotic production by signaling molecules in *Streptomyces*. *Frontiers in Microbiology* **10**, (2019).

71. Zhang, Y. *et al.* Activation of paulomycin production by exogenous γ-butyrolactone signaling molecules in *Streptomyces albidoflavus* J1074. *Appl.*

Microbiol. Biotechnol. **104**, 1695–1705 (2020).

72. Kitani, S. *et al.* Avenolide, a Streptomyces hormone controlling antibiotic production in *Streptomyces avermitilis*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16410–16415 (2011).

73. Craney, A., Ozimok, C., Pimentel-Elardo, S. M., Capretta, A. & Nodwell, J. R. Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem. Biol.* **19**, 1020–1027 (2012).

74. Pimentel-Elardo, S. M. *et al.* Activity-independent discovery of secondary metabolites using chemical elicitation and cheminformatic inference. *ACS Chem. Biol.* **10**, 2616–2623 (2015).

75. Okamoto, S. *et al.* Ribosome engineering and secondary metabolite production. (2004). doi:10.1016/S0065-2164(04)56005-7

76. Martín, J. F., Santos-Beneit, F., Sola-Landa, A. & Liras, P. Cross-talk of global regulators in Streptomyces. in *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria* **1**, 257–267 (Wiley Blackwell, 2016).

77. Sun, D., Liu, C., Zhu, J. & Liu, W. Connecting metabolic pathways: Sigma factors in *Streptomyces* spp. *Frontiers in Microbiology* **8**, (2017).

78. Rodríguez, H., Rico, S., Díaz, M. & Santamaría, R. I. Two-component systems in *Streptomyces*: Key regulators of antibiotic complex pathways. *Microbial Cell Factories* **12**, 127 (2013).

79. Xu, F., Nazari, B., Moon, K., Bushin, L. B. & Seyedsayamdost, M. R. Discovery of a cryptic antifungal compound from *Streptomyces albus* J1074 using high-throughput elicitor screens. *J. Am. Chem. Soc.* **139**, 9203–9212 (2017).

80. Covington, B. C., McLean, J. A. & Bachmann, B. O. Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Nat. Prod. Rep.* **34**, 6–24 (2017).

81. Xu, M. & Wright, G. D. Heterologous expression-facilitated natural products' discovery in actinomycetes. *Journal of Industrial Microbiology and Biotechnology* **46**, 415–431 (2019).

82. Tang, X. *et al.* Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10**, 2841–2849 (2015).

83. Jiang, W. *et al.* Cas9-Assisted Targeting of CHromosome cloning of large gene clusters. *Nat. Commun.* **6**, 1–8 (2015).

84. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440–446 (2012).

85. Greunke, C. *et al.* Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **47**, 334–345 (2018).

86. Li, L. *et al.* A stepwise increase in pristinamycin II biosynthesis by *Streptomyces pristinaespiralis* through combinatorial metabolic engineering. *Metab. Eng.* **29**, 12–25 (2015).

87. 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**, (2017).

88. Smanski, M. J. *et al.* Expression of the platencin biosynthetic gene cluster in heterologous hosts yielding new platencin congeners. *J. Nat. Prod.* **75**, 2158–2167 (2012).

89. 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 (2018).

90. Myronovskyi, M. *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **49**, 316–324 (2018).

91. Gomez-Escribano, J. P. & Bibb, M. J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207–215 (2011).

92. Komatsu, M. *et al.* Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* **2**, 384–396 (2013).

93. 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 (2010).

94. Zhao, Z. *et al.* Hybrubins: Bipyrrole tetramic acids obtained by crosstalk between a truncated undecylprodigiosin pathway and heterologous tetramic acid biosynthetic genes. *Org. Lett.* **18**, 572–575 (2016).

95. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509–523 (2015).

96. Teijaro, C. N., Adhikari, A. & Shen, B. Challenges and opportunities for natural product discovery, production, and engineering in native producers versus heterologous hosts. *J. Ind. Microbiol. Biotechnol.* **46**, 433–444 (2019).

97. Zhang, M. M., Wang, Y., Ang, E. L. & Zhao, H. Engineering microbial hosts for production of bacterial natural products. *Natural Product Reports* **33**, 963–987

(2016).

98. Ji, C.-H., Kim, J.-P. & Kang, H.-S. Library of synthetic *Streptomyces* regulatory sequences for use in promoter engineering of natural product biosynthetic gene clusters. *ACS Synth. Biol.* **7**, 1946–1955 (2018).

99. Myronovskyi, M. & Luzhetskyy, A. Native and engineered promoters in natural product discovery. *Natural Product Reports* **33**, 1006–1019 (2016).

100. Shao, Z. & Zhao, H. Manipulating natural product biosynthetic pathways via DNA assembler. *Curr. Protoc. Chem. Biol.* **6**, 65–100 (2014).

101. 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.* **112**, 8953–8958 (2015).

102. Kang, H.-S., Charlop-Powers, Z. & Brady, S. F. Multiplexed CRISPR/Cas9- and TAR-mediated promoter engineering of natural product biosynthetic gene clusters in yeast. *ACS Synth. Biol.* **5**, 1002–10 (2016).

103. Muyrers, J. P. P., Zhang, Y., Testa, G. & Stewart, A. F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* **27**, 1555–1557 (1999).

104. Doroghazi, J. R. *et al.* Aroadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **10**, 963–968 (2014).

105. Libis, V. *et al.* Uncovering the biosynthetic potential of rare metagenomic DNA using co-occurrence network analysis of targeted sequences. *Nat. Commun.* **10**, 1–9 (2019).

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

107. Johnston, C. D. *et al.* Systematic evasion of the restriction-modification barrier in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **166**, 11454–11459 (2019).

108. Li, L. *et al.* CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in *Streptomyces*. *Appl. Environ. Microbiol.* **84**, e00827-18 (2018).

109. Tong, Y. *et al.* Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20366–20375 (2019).

110. Blaskovich, M. A. T. *et al.* Developments in glycopeptide antibiotics. *ACS Infectious Diseases* **4**, 715–735 (2018).

111. Thaker, M. N. *et al.* Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat. Biotechnol.* **31**, 922–927 (2013).

112. Lemetre, C. *et al.* Bacterial natural product biosynthetic domain composition in soil correlates with changes in latitude on a continent-wide scale. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11615–11620 (2017).

113. Sharrar, A. M. *et al.* Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *MBio* **11**, 1–17 (2020).

114. van Bergeijk, D. A., Terlouw, B. R., Medema, M. H. & van Wezel, G. P. Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nature Reviews Microbiology* **18**, 546–558 (2020).

115. Amos, G. C. A. *et al.* Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci.* **114**, E11121–E11130 (2017).

116. Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C. & Banfield, J. F. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**, 440–444 (2018).