J Ind Microbiol Biotechnol (2014) 41:315–331
DOI 10.1007/s10295-013-1361-8

REVIEW

SIMB
Society for Industrial Microbiology and Biotechnology

Lessons learned from the transformation of natural product discovery to a genome-driven endeavor

Caitlin D. Deane · Douglas A. Mitchell

Received: 10 August 2013 / Accepted: 30 September 2013 / Published online: 19 October 2013

Abstract Natural product discovery is currently undergoing a transformation from a phenotype-driven field to a genotype-driven one. The increasing availability of genome sequences, coupled with improved techniques for identifying biosynthetic gene clusters, has revealed that secondary metabolomes are strikingly vaster than previously thought. New approaches to correlate biosynthetic gene clusters with the compounds they produce have facilitated the production and isolation of a rapidly growing collection of what we refer to as "reverse-discovered" natural products, in analogy to reverse genetics. In this review, we present an extensive list of reverse-discovered natural products and discuss seven important lessons for natural product discovery by genome-guided methods: structure prediction, accurate annotation, continued study of model organisms, avoiding genome-size bias, genetic manipulation, heterologous expression, and potential engineering of natural product analogs.

Keywords Natural products · Genome mining · Reverse discovery · Biosynthetic gene clusters · Structure prediction · Genetic manipulation · Heterologous expression

C. D. Deane · D. A. Mitchell (✉)
Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: douglasm@illinois.edu

C. D. Deane · D. A. Mitchell
Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

D. A. Mitchell
Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Introduction

Natural products have played a significant role in medicine, and fittingly, the discovery of novel natural products continues to hold great promise for the development of new drugs [108]. Forward natural product discovery relies on the presence of an observable phenotype or chemical property, such as biological activity, color, or a known mass, which can be tracked through successive rounds of isolation. Successful "grind and find" isolation of a natural product can be followed by identification of the genes responsible for the biosynthesis of the compound of interest. This method was remarkably productive in its early years, but has been met with increasing frustration by researchers as the rediscovery of known compounds has become commonplace, as many of the easily accessible natural products have already been isolated [7].

Meanwhile, the increasing ease of DNA sequencing has led to an explosion of fully sequenced genomes and identified biosynthetic pathways over the last 20 years. In many cases, the biosynthetic genes responsible for well-known and long-studied natural products have been identified only recently (Fig. 1). Furthermore, the elucidation of biosynthetic pathways for natural products is by no means complete, as a number of important compounds (e.g., morphine, paclitaxel) still have major gaps in what is known of their biosynthesis (Fig. 1). Biosynthetic genes for plant natural products are often particularly difficult to identify because plants do not commonly cluster the responsible genes as bacteria and fungi often do, and plant genomes are much larger than those of bacteria and fungi [157]. That is not to say the task is easy in lower organisms, as history has shown us that characterization of bacterial and fungal pathways requires immense time and effort. A noteworthy example is the bottromycin family, whose biosynthetic

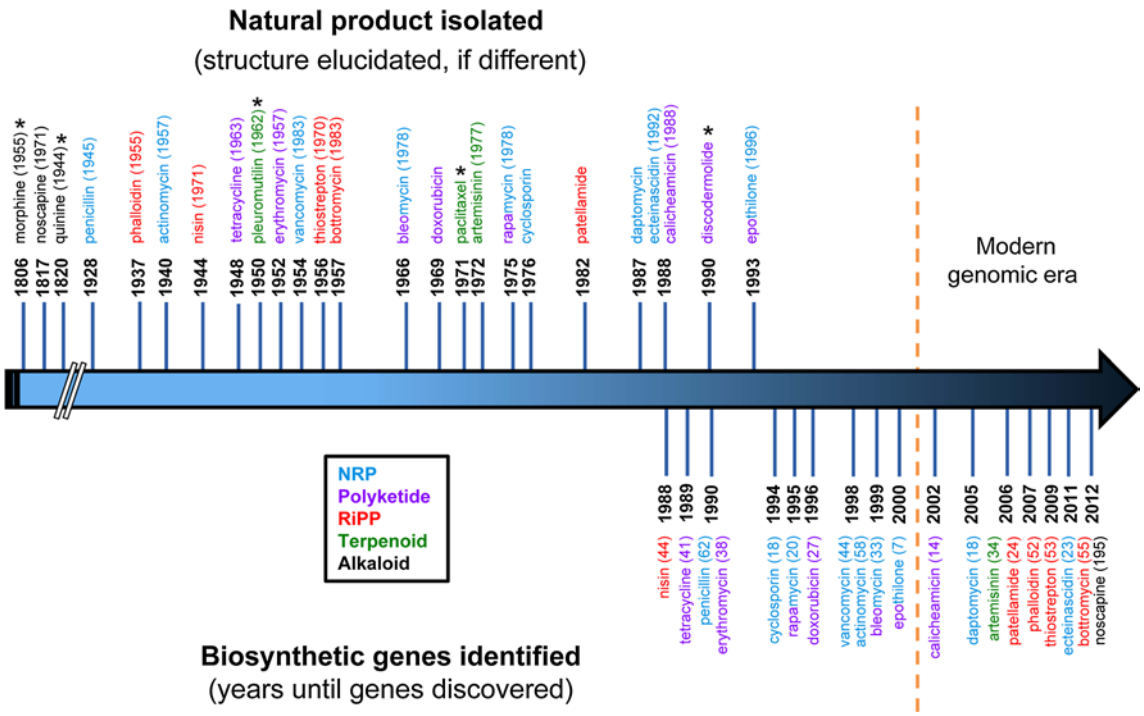

Fig. 1 Timeline of traditional "forward" natural product discovery, highlighting the frequently decades-long gap between isolation of a novel natural product and identification of its biosynthetic genes. A date above the *arrow* indicates when isolation of the indicated natural product was first reported, with the date of structure elucidation (if different) in parentheses. A date below the *arrow* indicates when the biosynthetic genes were first identified for the natural product, with the years since isolation in parentheses. This figure is not meant to be exhaustive but instead to provide a representative sample of natural products spanning multiple biosynthetic classes, biological targets, and producing organisms. *NRP* Non-ribosomal peptide, *RiPP* ribosomally synthesized and post-translationally modified peptide; * majority of biosynthetic pathway remains unknown

gene cluster was identified by four independent groups in 2012, which was 55 years after the first report of its isolation [35, 53, 66, 68]. Likewise, the genes responsible for the production of the thiopeptide antibiotics remained unknown until being reported upon nearly simultaneously by four independent research groups in 2009 [73, 89, 106, 156]. In the case of thiostrepton, a well-known thiopeptide, this report came 53 years after its initial isolation.

With the advent of inexpensive, massively parallel sequencing, a new route to discover natural products has emerged. The genomic revolution has fueled a shift in methods for identifying natural products from the traditional, phenotype-driven "forward" procedure to a genotype-driven, "reverse" discovery process, during which identification of the biosynthetic genes precedes and informs isolation of the natural product. With the availability of the of *Streptomyces coelicolor* and *Streptomyces avermitilis* genome sequences [11, 114], it became clear that even some highly studied strains whose biosynthetic capabilities had appeared to be exhausted harbored a surprising number of previously unknown biosynthetic gene clusters. For example, only three natural product gene clusters had been identified on the *S. coelicolor* chromosome prior to the completion of its genome sequence: those for actinorhodin [96, 97], prodiginine [46], and calcium-dependent antibiotic [29]. Genome sequencing of *S. coelicolor* revealed a number of "cryptic" gene clusters without known associated natural products (sometimes called orphan gene clusters). In total, *S. coelicolor* carries the potential for 29 structurally complex natural products (Fig. 2) [11, 34]. Many of these biosynthetic gene clusters remain cryptic to the present day, over a decade later, indicating that the biosynthetic capabilities of this and other organisms still have yet to be fully understood.

The number of complete bacterial genome sequences available in the National Center for Biotechnology Information (NCBI) database increased approximately 25-fold, to over 2,500, between 2003 and 2013. With the ready availability of sequenced genomes, particularly those from bacteria, it is increasingly possible to identify putative biosynthetic genes, use their sequence to predict the structure and properties of the potential product, and use those predictions to guide efficient isolation and characterization. One interesting example of this is the case of bacillaene, whose structure eluded scientists due to its chemical instability [115]. The genetic sequence of the bacillaene

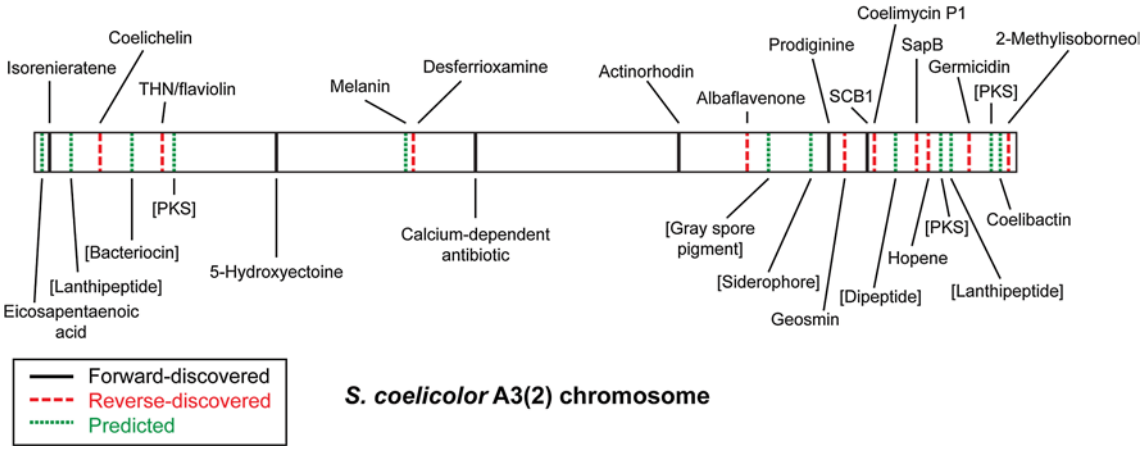

Fig. 2 Schematic representation of the linear 8.7 Mb *Streptomyces coelicolor* A3(2) chromosome, highlighting the locations of all known natural product biosynthetic gene clusters. *Names in brackets* indicate putative compounds for which no predicted structure appears in the literature [34]

producer led to identification of the protein sequences responsible for its biosynthesis, which in turn enabled a structural prediction for bacillaene that guided purification based on the perceived physical properties. Ultimately, this procedure was successful, and the structure was solved [18, 24]. In addition, the prototypical reverse-discovered natural product, coelichelin, had its structure predicted with a reasonably high degree of accuracy several years prior to its isolation [21, 83].

In Table 1, we present an extensive list of natural products that have been reverse discovered. For more detailed discussion of many of these natural products, we direct the reader to a number of reviews on this topic [20, 55, 150], in addition to those contained in the current issue of this journal. Many of the natural products listed in this table were discovered in 2008 or later, reflecting the exponential rise in available genome sequences. A number of these natural products are from well-studied model organisms (*S. coelicolor*, *Aspergillus nidulans*). Most are from bacteria, consistent with the larger number of sequenced bacterial genomes relative to fungi and plants.

In this review, we focus on seven lessons for the community to keep in mind during the reverse-discovery of natural products: (1) structure prediction, (2) accurate annotation, (3) continued study of model organisms, (4) avoiding genome-size bias, (5) genetic manipulation, (6) heterologous expression, and (7) potential engineering of analogs. To illustrate the utility of these lessons, we highlight five natural products from the recent literature (aureusimine, plantazolicin, astexin-1, the prochlorosins, and asperfuranone; Fig. 3) representing three prominent natural product classes: polyketides, non-ribosomal peptides (NRPs), and ribosomally synthesized and post-translationally modified peptides (RiPPs), as well as a variety of producing organisms (bacteria and fungi).

Lesson 1: structure prediction

Access to a predicted structure facilitated the isolation of the aureusimines, founding members of the pyrazinone class of natural products, by enabling mass spectrometry-guided isolation. These compounds are biosynthesized by a NRPS pathway in the human pathogen *Staphylococcus aureus* (Fig. 3a) [160, 166]. The aureusimine biosynthetic gene cluster was discovered by genome mining to identify possible NRPSs that are highly conserved among sequenced *S. aureus* strains. Homologous gene clusters have been found in over 50 *S. aureus* strains as well as other human pathogenic *Staphylococcus* species [160]. Prior to its isolation, the structure of aureusimine A was predicted based on the sequence of the NRPS gene and the previously established amino acid specificities for NRPS adenylation domains. The presence of a putative reductase domain at the C-terminus of the *S. aureus* NRPS was also factored into the structural prediction, as it indicated that the dipeptide was likely released from the synthetase as an aldehyde with the potential to spontaneously cyclize. This prediction was confirmed with the solved structures of aureusimines A and B (Fig. 3b) [160].

Structural prediction and mass spectrometry-guided isolation also proved useful in the isolation of plantazolicin, a member of the thiazole/oxazole-modified microcin (TOMM) subclass of RiPP natural products [3]. TOMM biosynthesis is characterized by the post-translational modification of ribosomally synthesized precursor peptides to generate thiazol(in)e and (methyl)oxazol(in)e heterocycles from the side chains of select cysteine, serine, and threonine residues [86]. The gene cluster for plantazolicin (Fig. 3a) was identified in 2008 during a search for genes with homology to the biosynthetic gene cluster responsible

Table 1 Reverse-discovered natural products, organized by year of reported isolation

Name	Class	Producer	Year isolated (year predicted, if different)	Genome sequenced	Reference(s)
Hopene	Polyketide	*Streptomyces coelicolor* A3(2)	2000	2002	[11, 121]
Collinone	Polyketide	*Streptomyces collinus* DSM 2012	2001	n/a	[98]
Geosmin	Terpenoid	*Streptomyces coelicolor* A3(2)	2003	2002	[11, 19]
Desferrioxamine	Other	*Streptomyces coelicolor* A3(2)	2004	2002	[9, 11]
Halstoctacosanolides	Polyketide	*Streptomyces halstedii* HC34	2004	n/a	[146, 147]
SapB	RiPP	*Streptomyces coelicolor* A3(2)	2004	2002	[11, 77]
Tetrahydroxy-naphthalene	Polyketide	*Streptomyces coelicolor* A3(2)	2004	2002	[4, 11]
Thalianol	Terpenoid	*Arabidopsis thaliana*	2004	2000	[2, 45]
Aurafurons	Polyketide	*Stigmatella aurantiaca* DW4/3-1	2005	2011	[67, 81]
Coelichelin	NRP	*Streptomyces coelicolor* A3(2)	2005 (2000)	2002	[11, 21, 83]
ECO-02301	Polyketide	*Streptomyces aizunensis* NRRL B-11277	2005	n/a	[100]
Myxochromides S	NRP	*Stigmatella aurantiaca* DW4/3-1	2005	2011	[67, 155]
Aspoquinolones	Alkaloid	*Aspergillus nidulans*	2006	2005	[49, 132]
Bacillaene	Polyketide/NRP	*Bacillus amyloliquefaciens* FZB42	2006	2007	[18, 22, 24]
DKxanthenes	Polyketide/NRP	*Myxococcus xanthus* DK1050	2006	2006	[51, 103]
ECO-501	Polyketide	*Amycolatopsis orientalis* ATCC 43491	2006	n/a	[8]
Germicidins	Polyketide	*Streptomyces coelicolor* A3(2)	2006	2002	[11, 138]
Haloduracin	RiPP	*Bacillus halodurans* C-125	2006	2000	[84, 102, 142]
Penocin A	RiPP	*Pediococcus pentosaceus* ATCC 25745	2006	2006	[39, 94]
Terrequinone A	NRP	*Aspergillus nidulans*	2006	2005	[16, 49]
Trichamide	RiPP	*Trichodesmium erythraeum* IMS101	2006	2006	[141]
Aeruginosides	NRP	*Planktothrix agardhii* CYA126/8	2007	Draft	[70]
Aspyridones	Polyketide/NRP	*Aspergillus nidulans*	2007	2005	[12, 49]
CBS-40	Polyketide	*Streptomyces* sp. CB2544	2007	n/a	[64]
Orfamide A	NRP	*Pseudomonas fluorescens* Pf-5	2007	2005	[56, 116]
Aerucyclamide C	RiPP	*Microcystis aeruginosa* PCC7806	2008	Draft	[122]
Albaflavenone	Terpenoid	*Streptomyces coelicolor* A3(2)	2008	2002	[11, 163]
Capistruin	RiPP	*Burkholderia thailandensis* E264	2008	2005	[75, 76]
Diazepinomicin (ECO-4601)	Other	*Micromonospora* sp.	2008	n/a	[101]

Table 1 continued

Name	Class	Producer	Year isolated (year predicted, if different)	Genome sequenced	Reference(s)
Emericellamide	Polyketide/NRP	*Aspergillus nidulans*	2008	2005	[28, 49]
2-Methylisoborneol	Terpenoid	*Streptomyces coelicolor* A3(2)	2008	2002	[11, 78]
Microcyclamide 7806A, B	RiPP	*Microcystis aeruginosa* PCC7806	2008	Draft	[164]
Thailandamides	Polyketide	*Burkholderia thailandensis* E264	2008	2005	[75, 109]
Vibi A-K	RiPP	*Viola biflora*	2008	n/a	[62]
Asperfuranone	Polyketide	*Aspergillus nidulans*	2009	2005	[27, 49]
Atrochrysone	Polyketide	*Aspergillus terreus*	2009	Draft	[5]
Emodin	Polyketide	*Aspergillus nidulans*	2009	2005	[15, 49]
F9775 A/B (blecanoric acid)	Polyketide	*Aspergillus nidulans*	2009	2005	[15, 49, 129]
Lecanoric acid	Polyketide	*Aspergillus nidulans*	2009	2005	[49, 135]
Lichenicidin	RiPP	*Bacillus licheniformis*	2009	2004	[10, 40, 126, 137]
Monodictyphenone	Polyketide	*Aspergillus nidulans*	2009	2005	[15, 49]
Mra4, 5	RiPP	*Melicytus ramiflorus*	2009	n/a	[148]
Nygerone A	Polyketide/NRP	*Aspergillus niger* CBS 513.88	2009	2007	[61, 119]
Orsellinic acid	Polyketide	*Aspergillus nidulans*	2009	2005	[49, 129, 135]
Anacylamides	RiPP	*Anabaena* sp. 90	2010	2012	[87, 153]
Aspernidine	Alkaloid	*Aspergillus nidulans*	2010	2005	[49, 133]
Aureusimine	NRP	*Staphylococcus aureus*	2010	2008	[6, 160, 166]
Bsa	RiPP	*Staphylococcus aureus*	2010	2004	[36, 63]
Coelimycin P1	Polyketide	*Streptomyces coelicolor* A3(2)	2010 (2007)	2002	[11, 117, 118]
Csypyrone B1	Polyketide	*Aspergillus oryzae*	2010	2005	[92, 136]
Diorcinol	Polyketide	*Aspergillus nidulans*	2010	2005	[49, 129]
Erythrochelin	NRP	*Saccharopolyspora erythraea* NRRL 2338	2010	2007	[85, 113, 127]
Gerfelin	Polyketide	*Aspergillus nidulans*	2010	2005	[49, 129]
Globa A, B	RiPP	*Gloeospermum blakeanum*	2010	n/a	[17]
Microcin H47	RiPP	*Escherichia coli* (various strains)	2010	n/a	[149]
Microcin M	RiPP	*Escherichia coli* (various strains)	2010	n/a	[149]
Microviridin L	RiPP	*Microcystis aeruginosa* NIES843	2010	2007	[72, 165]
Pneumococcin	RiPP	*Streptococcus pneumoniae* R6	2010	2001	[65, 93]
Prochlorosins	RiPP	*Prochlorococcus marinus* MIT9313	2010	2003	[88, 128]
TP-1161	RiPP	*Nocardiopsis* sp. TFS65-07	2010	n/a	[43]
Venezuelin	RiPP	*Streptomyces venezuelae* ATCC10712	2010	2011	[54, 124]
Austinol	Polyketide/terpenoid	*Aspergillus nidulans*	2011	2005	[49, 110]
Desmethylbassianin A	Polyketide	*Beauveria bassiana*	2011	Draft	[60]
Grisemycin	RiPP	*Streptomyces griseus* IFO 13350	2011 (2010)	2008	[31, 32, 112]

Table 1 continued

Name	Class	Producer	Year isolated (year predicted, if different)	Genome sequenced	Reference(s)
(Iso)flavipucine	Polyketide/NRP	*Aspergillus terreus*	2011	Draft	[125]
Koranimine	NRP	*Bacillus* spp.	2011	n/a	[44]
Plantazolicin	RiPP	*Bacillus amyloliquefaciens* FZB42	2011 (2008)	2007	[22, 86, 134]
Stambomycins	Polyketide	*Streptomyces ambofaciens* ATCC23877	2011	Draft	[82]
Thailandepsin	Polyketide/NRP	*Burkholderia thailandensis* E264	2011	2005	[75, 152]
Alternariol	Polyketide	*Aspergillus nidulans*	2012	2005	[1, 49]
Astexin-1	RiPP	*Asticcacaulis excentricus*	2012	2010	[95]
Azanigerones	Polyketide	*Aspergillus niger*	2012	Draft	[162]
Burkholderic acid	Polyketide/NRP	*Burkholderia thailandensis* E264	2012	2005	[48, 75]
Catenulipeptin	RiPP	*Catenulispora acidiphila* DSM 44928	2012	2009	[33, 154]
Cichorine	Polyketide	*Aspergillus nidulans*	2012	2005	[1, 49]
Curvopeptin	RiPP	*Thermomonospora curvata*	2012	2011	[26, 80]
Elgicin	RiPP	*Paenibacillus elgii* B69	2012	Draft	[145]
Fusarielins F-H	Polyketide	*Gibberella zeae*	2012	Draft	[139]
Geobacillin	RiPP	*Geobacillus thermodenitrificans* NG80-2	2012	2007	[47, 50]
Luminmycin	Polyketide/NRP	*Photorhabdus luminescens subsp. laumondii* TT01	2012	2003	[13, 41]
Malleilactone	Polyketide	*Burkholderia thailandensis* E264	2012	2005	[14, 75]
O-Methyldiaporthin	Polyketide	*Aspergillus oryzae*	2012	2005	[92, 107]
Pre-shamixanthone	Polyketide	*Aspergillus nidulans*	2012	2005	[49, 131]
Rhizopodin	Polyketide/NRP	*Stigmatella aurantiaca* Sg a15	2012	Draft	[120]
Caulosegnins	RiPP	*Caulobacter segnis*	2013	2010	[59]
Flavipeptin	RiPP	*Kribbella flavida*	2013	2010	[123, 151]
Flavopeptins	NRP	*Streptomyces flavogriseus* ATCC 33331	2013	2011	[25]
Fumicycline A/neosartoricin A	Polyketide	*Aspergillus fumigatus*	2013	2005	[30, 79, 111]
Neosartoricins B-D	Polyketide	*Trichophyton tonsurans*	2013	2012	[99, 161]
Thailanstatins	Polyketide/NRP	*Burkholderia thailandensis* MSMB43	2013	Draft	[91]

In cases where more than one species is known to produce a particular natural product, only the strain from which the compound was first isolated is listed

for the production of streptolysin S and microcin B17, which are TOMMs from *Streptococcus pyogenes* and *Escherichia coli*, respectively [86]. Like other RiPPs, the sequence of a predicted precursor peptide gene enabled mass spectrometric-guided separation to isolate plantazolicin from organic surface extracts of modified strains of *B. amyloliquefaciens* FZB42 [134] and aided significantly in the elucidation of its structure (Fig. 3b) [71, 105].

A concluding example of the utility of structure prediction during natural product discovery is provided by coelichelin, a prominent early example of a reverse-discovered natural product, which had its structure predicted several years prior to its isolation from *S. coelicolor* [21]. Although the initial structure was not entirely correct [83], the ability to predict structure based on genetic sequence is playing a growing role in reverse natural product

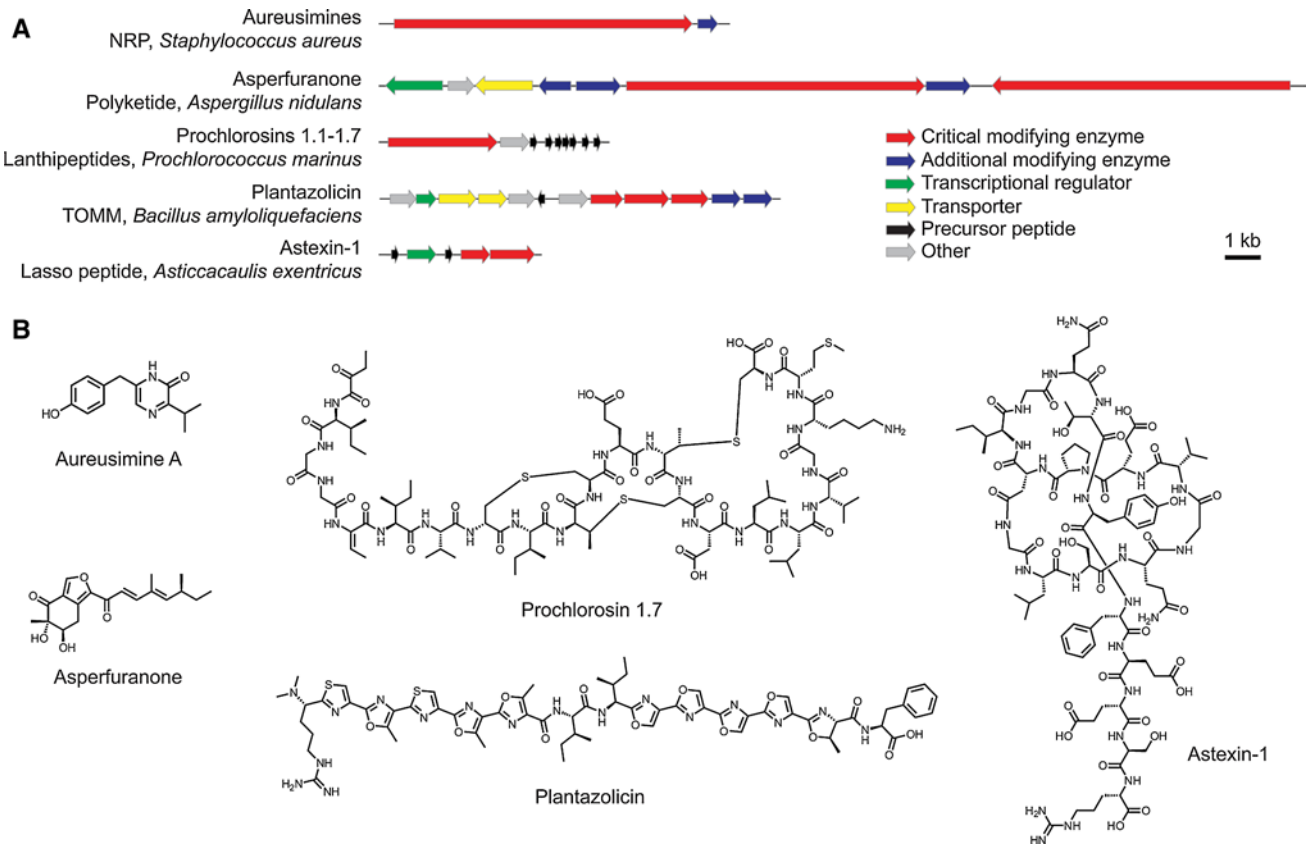

Fig. 3 Biosynthetic gene clusters (**a**) and structures (**b**) of select reverse-discovered natural products. Gene clusters in **a** are shown to scale and are shown with the name of the organism in which they were first identified

discovery, whether by providing insight into possible extraction conditions or by simplifying identification by mass spectrometry.

Lesson 2: accurate sequence annotation

Regardless of how many genome sequences are available, they are useless without interpretation. Complete and accurate sequence annotation also play crucial roles in the efficient reverse discovery of natural products. RiPP precursor peptide genes, in particular, are typically short and are often unannotated in published genomes [3, 104], complicating their identification during genome mining. For example, the open reading frame (ORF) encoding the plantazolicin precursor peptide was unannotated at the time of the initial report [86]. The same has also proven to be the case for the precursors of several lasso peptides, an emerging subclass of RiPP natural products [59, 76].

Despite the potential difficulty of identifying unannotated precursor genes, the discovery of the lasso peptide astexin-1 through a precursor-guided screen provides a promising avenue for future genome mining in the pursuit of novel RiPP natural products. Lasso peptides are characterized by the post-translational formation of a polypeptide loop structure with a covalent bond between the N-terminus of the precursor peptide and a specific aspartate or glutamate side chain elsewhere on the peptide. The C-terminal end of the precursor peptide is then threaded through the loop and often held in place by the side chains of bulky residues in the tail region [3]. The resulting threaded structure typically confers unique heat stability on the lasso peptides [167]. Genome mining for RiPP biosynthetic gene clusters commonly focuses on identifying homologs of a known *modifying enzyme*, as was the case for the lasso peptide capistruin [76], the prochlorosins, and plantazolicin. However, in the case of astexin-1, a genetic screen for *precursor peptides* was employed. This screen identified short ORFs, some unannotated, containing amino acid patterns consistent with the sequences of known lasso peptides. When a potential lasso peptide precursor was found, nearby regions of the genome were then searched for homology to the maturation enzymes [95]. This endeavor resulted in the identification of 79 putative lasso peptide gene clusters distributed across nine bacterial phyla and one archaeal phylum [95].

The identification of precursor peptide genes located outside of the primary gene cluster further underscores the advantage of searching widely for all possible biosynthetic genes during reverse natural product discovery [57]. Such was the case for the prochlorosins, a group of lanthipeptide RiPPs produced by some marine cyanobacteria species, most notably *Prochlorococcus marinus* [88]. The hallmark of lanthipeptide biosynthesis is the formation of (methyl)lanthionine rings between select cysteines and dehydroalanine or dehydrobutyrine residues, which arise from the dehydration of serines and threonines, respectively. Dehydration and lanthionine ring installation on the ribosomal precursor peptide may be catalyzed by one or more biosynthetic enzymes [3]. The gene clusters responsible for prochlorosin biosynthesis (Fig. 3a) were identified in *P. marinus* and related species through a bioinformatic search for bifunctional lanthipeptide synthetase homologs [88] and, independently, through the homology of prochlorosin precursor peptides to the Nif11 nitrogen-fixing proteins [58]. However, only seven of these 29 putative precursor peptide genes are clustered with the lanthipeptide synthetase gene [88], contrary to the norm for biosynthetic genes in bacteria and fungi.

Lesson 3: continued study of model organisms

The sequencing of the *S. coelicolor* genome, which brought to light the silent majority of biosynthetic gene clusters, also provides a reminder not to discount continued investigation into well-studied model organisms during the search for new natural products. These strains may contain the genetic potential to synthesize many more natural products than are superficially detectable. The fungus *A. nidulans* [27] was a highly studied model organism even prior to the sequencing of its genome, and since that time, it has been determined that *A. nidulans* FGSC A4 contains 24 PKS genes and 3 PKS-like genes [49, 130]. Several pairs of PKS genes are adjacent to each other (Fig. 3a), indicating a strong possibility that the pairs might work in concert, though *A. nidulans* was not previously known to produce any polyketides arising from the combined action of two PKS enzymes. The producer of plantazolicin, *B. amyloliquefaciens* FZB42, also produces a number of natural products [22], including the polyketide/NRP hybrid bacillaene, which likewise is a reverse-discovered natural product [18, 24].

Lesson 4: avoiding genome-size bias

One common assumption in the natural products field has been that organisms with small genomes are unlikely to produce structurally complex natural products. Indeed, bacteria with larger genomes, such as *S. coelicolor* (8.7 Mb), are recognized as prolific producers of secondary metabolites [11]. However, species with small genomes should not be ignored in the search for natural products, as exemplified by the isolation of structurally complex compounds from organisms whose genomes are a fraction the size of the traditionally recognized "natural product powerhouses". These species with smaller genomes, such as *S. aureus* (2.8 Mb) [6], are presumed to be less able to devote genome space to large biosynthetic gene clusters such as NRPSs, making the identification of a single NRPS gene that occupies 0.25 % of the *S. aureus* genome even more surprising [160].

Maintenance of multiple precursor genes and the ability to modify them with a single enzyme enables organisms with compact genomes, such as prochlorosin producer *P. marinus* (2.4 Mb), to produce a variety of structurally complex molecules without the need for large biosynthetic gene clusters. Indeed, the precursor hypervariability of the prochlorosin system endows *P. marinus* MIT9313 with the potential to produce as many natural products as *S. coelicolor*, which bears a genome nearly four times as large [11, 88].

Lesson 5: genetic manipulation

In many instances, natural products are difficult to detect under laboratory cultivation, due to low production levels or masking by the presence of other, more abundant, natural products. In such cases, genetic manipulation may be used to activate these silent gene clusters. The reverse discovery of the polyketide asperfuranone (Fig. 3b) illustrates how silent gene clusters can be activated via promoter engineering within the native host. In this case, the native promoter for a predicted transcriptional activator gene within the PKS cluster (Fig. 3a) was replaced with an inducible promoter [27]. As asperfuranone was not produced to any significant extent during laboratory cultivation, it would have been unlikely to identify this compound via any phenotype-driven effort. Other silent biosynthetic gene clusters that have been successfully activated by promoter engineering include those responsible for producing aspirydone [12], malleilactone [14], and stambomycin [82]. A limitation of this strategy is that it requires a genetically tractable host.

Another common genetic manipulation used for natural product discovery is biosynthetic gene disruption. Arguably, the most powerful method to establish gene function is to evaluate the phenotypes of the parent (wild-type) and the genetic deletion (mutant) strains. This reverse genetics method is often employed during the genome-guided

discovery of natural products. The phenotype compared between parent and mutant is typically either the metabolite profile (usually obtained by mass spectrometry) or the bioactivity of chromatographic fractions. The detection and isolation of plantazolicin demonstrated a secondary use of such a method to identify a novel compound whose production had been masked by other, more abundant compounds. When production of all known bioactive compounds was eliminated, the remaining bioactivity indicated that at least one additional biosynthetic gene cluster had yet to be identified [23, 134].

Lesson 6: heterologous expression

In cases where genetic manipulation in the native host is not feasible, the transfer of biosynthetic genes clusters to a genetically tractable host has the potential to facilitate compound production. The biosynthetic gene cluster for the lasso peptide astexin-1 (Fig. 3a) was identified in the freshwater α-proteobacterium *Asticcacaulis exentricus*, although compounds with the expected mass could not be detected after laboratory cultivation of this organism, necessitating an alternate expression method [95]. Astexin-1 was successfully expressed, purified, and structurally characterized using *E. coli* as a heterologous host (Fig. 3b) [95, 167]. Following extensive variation of culture conditions, a recent study succeeded in producing astexin-1 from *A. exentricus*, albeit at low levels compared to those from *E. coli* [167]. Likewise, successful high-level production of a number of prochlorosins required heterologous expression in *E. coli* and reconstitution in vitro, ultimately enabling structural confirmation of these natural products (Fig. 3b) [88, 144].

For further discussion and examples of heterologous expression, the interested reader is directed to other reviews specifically focused on this topic, also in this Special Issue [52, 69].

Lesson 7: potential engineering of natural product analogs

Beyond simply providing a novel natural product, identification of new biosynthetic gene clusters also has the potential to facilitate rational engineering of unnatural natural product variants. The production of prochlorosins from *P. marinus* MIT9313 is notable among lanthipeptides because of the exceptionally high number of precursor peptides (29) [88]. While it is not unusual for a lanthipeptide gene cluster to harbor more than one precursor peptide, these typically require multiple lanthionine synthetases to modify all the precursor peptides [10, 102]. Indeed, 17 prochlorosins from *P. marinus* MIT9313 have been successfully biosynthesized both in vitro and in vivo using this single lanthipeptide synthetase [88]. The potential for a single lanthipeptide synthetase enzyme to effect modification of up to 29 precursor peptides suggests that the prochlorosin synthetase is remarkably promiscuous, indicating that it may be a prime candidate for engineering novel lanthipeptides.

In addition to the natural combinatorial biosynthesis of prochlorosin and its potential use in engineering, several other examples presented in this review have shown promise for the future engineering of natural product variants. The NRPS responsible for aureusimine production has been heterologously expressed and used to explore the pyrazinone assembly line using alternate substrates [158, 159]. The enzymes responsible for asperfuranone biosynthesis have also provided a compliant system for the engineering of unnatural polyketides by domain swapping, an endeavor which has historically been challenging due to the complex nature of the requisite megasynthase enzymes [90]. The heterologous expression of plantazolicin in *E. coli* has also been achieved and exploited using precursor peptide replacement to generate unnatural variants [38]. In summary, the thorough understanding of biosynthetic gene clusters afforded by reverse natural product discovery has the potential to expedite the production of unnatural variants for applications in medicinal chemistry and structure–activity relationship studies.

Outlook

The sequencing and annotation of new genomes will continue to provide ample opportunity for the discovery of new natural products through genotype-based methods. It is possible, eventually, that natural product discovery will reach a point of diminishing returns from the investigation of new species, but the steadily increasing rate of reverse natural product discovery (Table 1) indicates that the field is still far from that point. While it is relatively straightforward to identify new putative PKS and NRPS clusters through sequence homology, other classes of natural products will require more sophisticated methods. For RiPP natural products, bioinformatics tools such as BAGEL [37] are useful in identifying genes for potential novel compounds, but less well-characterized and yet-to-be-discovered classes of natural products will require significant future investigation. The genes responsible for the biosynthesis of new or poorly understood natural products may currently be annotated as hypothetical proteins or domains of unknown function. The use of tools to group these unknown genes together by sequence homology [42], coupled with improved annotation in sequenced genomes, may allow for entirely new biosynthetic pathways to be discovered.

As the field of reverse natural product discovery moves forward, we anticipate an increasing dependence on structure prediction to facilitate isolation of interesting natural products and to provide starting points for structure elucidation. Structure prediction for new RiPP natural products is largely enabled by identification of the precursor peptide sequence, while programs such as ClustScan [140] utilize known domain specificities to predict the structure of novel PKS and NRPS gene clusters. While not strictly a method for reverse natural product discovery, peptidogenomics may also be helpful in this endeavor by identifying new peptide natural products and linking those to their respective biosynthetic genes [74]. Perhaps most widely useful, however, are comprehensive tools such as antibiotics and Secondary Metabolite Analysis Shell (antiSMASH), which incorporates information about many natural product classes beyond the comparatively well-studied PKS, NRPS, and RiPP systems. As the collective knowledge of the natural product discovery field increases, the predictive power of these tools will likewise increase and create a positive feedback loop to further enable the future isolation of novel compounds.

Natural product discovery will also be aided by the development of new genetic tools. To set a lofty goal, it would be of enormous utility to have a universal, user-friendly DNA manipulation platform for deleting genes of interest. Not only would this platform greatly enable deletion-guided approaches to natural product discovery, the whole of biology would benefit immensely from the development of a simple method for constructing genetic knockouts applicable to a broad range of organisms. If such a technique were developed and freely shared with the community, research groups focused on the chemical aspects of natural product research would now have the ability to construct useful strains without specialized knowledge in genetic manipulation.

Perhaps the most significant hurdle to isolating reverse-discovered natural products is that many gene clusters are silent (expression is undetectable) under laboratory cultivation. In such cases, no novel natural product is readily observable in extracts from cultures of the organism harboring the biosynthetic genes of interest. In order to isolate the compound in question, the genes responsible for its production must be activated. A number of methods have been developed to activate silent clusters, including variation of culture conditions, heterologous expression, promoter engineering, induction with gamma-butyrolactones, and co-culturing with other organisms to stimulate production, among others [20, 55, 130, 143]. One persistent issue with the use of heterologous hosts is the potential for inefficient export of the natural product, which may be addressed in part by the development of greater numbers of genetically tractable heterologous hosts. These tools will help to streamline the identification and production of natural products from gene clusters which are difficult, if not impossible, to express under laboratory culture conditions.

Conclusions

Rather than relying on the observation of a phenotype, reverse (genotype-driven) natural product discovery bases isolation of the compound(s) of interest on knowledge of the associated biosynthetic genes. The examples of reverse-discovered natural products presented in this review represent several biosynthetic classes, producing organisms, and biological targets. It is plausible that none of the discussed examples would have been discovered by the traditional phenotype-driven method due to unknown biological activity or activity that is extremely selective so as to preclude detection during bioassay-guided isolation. Perhaps more significantly, most of these examples, as well as others listed in Table 1, required some measure of genetic manipulation (promoter engineering, comparison of genetic deletion strains, heterologous expression, etc.) in order to produce useful levels of the natural product of interest. In some cases, a low level of production was masked by other, more prominent compounds, while in others, the biosynthetic gene cluster of interest was entirely silent despite extensive modification of culture conditions. The emergence of novel natural products even from well-studied model organisms indicates that the potential for production of structurally complex compounds remains largely untapped at this point, and the constantly increasing number of publicly available genomes will certainly provide a surplus of avenues for natural product discovery.

Acknowledgments We thank members of the Mitchell lab for the critical review of this manuscript. C.D.D. is supported by the Robert C. and Carolyn J. Springborn Endowment. D.A.M. is supported by the US National Institutes of Health (NIH, 1R01 GM097142) and the NIH Director's New Innovator Award Program (DP2 OD008463).

References

1. Ahuja M, Chiang YM, Chang SL, Praseuth MB, Entwistle R, Sanchez JF, Lo HC, Yeh HH, Oakley BR, Wang CC (2012) Illuminating the diversity of aromatic polyketide synthases in *Aspergillus nidulans*. J Am Chem Soc 134(19):8212–8221. doi:10.1021/ja3016395
2. Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814):796–815. doi:10.1038/35048692
3. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian KD, Fischbach MA, Garavelli JS, Goransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP,

Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Muller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJ, Rebuffat S, Ross RP, Sahl HG, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Sussmuth RD, Tagg JR, Tang GL, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 30(1):108–160. doi:10.1039/c2np20085f

4. Austin MB, Izumikawa M, Bowman ME, Udwary DW, Ferrer JL, Moore BS, Noel JP (2004) Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates. J Biol Chem 279(43):45162–45174. doi:10.1074/jbc.M406567200
5. Awakawa T, Yokota K, Funa N, Doi F, Mori N, Watanabe H, Horinouchi S (2009) Physically discrete beta-lactamase-type thioesterase catalyzes product release in atrochrysone synthesis by iterative type I polyketide synthase. Chem Biol 16 (6):613-623. doi:10.1016/j.chembiol.2009.04.004
6. Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K (2008) Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. J Bacteriol 190(1):300–310. doi:10.1128/JB.01000-07
7. Baltz RH (2006) Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? J Ind Microbiol Biot 33(7):507–513. doi:10.1007/s10295-005-0077-9
8. Banskota AH, McAlpine JB, Sorensen D, Ibrahim A, Aouidate M, Piraee M, Alarco AM, Farnet CM, Zazopoulos E (2006) Genomic analyses lead to novel secondary metabolites. Part 3. ECO-0501, a novel antibacterial of a new class. J Antibiot 59(9):533–542. doi:10.1038/ja.2006.74
9. Barona-Gomez F, Wong U, Giannakopulos AE, Derrick PJ, Challis GL (2004) Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. J Am Chem Soc 126(50):16282–16283. doi:10.1021/ja045774k
10. Begley M, Cotter PD, Hill C, Ross RP (2009) Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. Appl Environ Microbiol 75(17):5451–5460. doi:10.1128/AEM.00730-09
11. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885):141–147. doi:10.1038/417141a
12. Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C (2007) Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. Nat Chem Biol 3(4):213–217. doi:10.1038/nchembio869
13. Bian X, Plaza A, Zhang Y, Muller R (2012) Luminmycins A-C, cryptic natural products from *Photorhabdus luminescens* identified by heterologous expression in *Escherichia coli*. J Nat Prod 75(9):1652–1655. doi:10.1021/np300444e
14. Biggins JB, Ternei MA, Brady SF (2012) Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of *Burkholderia pseudomallei* group pathogens. J Am Chem Soc 134(32):13192–13195. doi:10.1021/ja3052156
15. Bok JW, Chiang YM, Szewczyk E, Reyes-Dominguez Y, Davidson AD, Sanchez JF, Lo HC, Watanabe K, Strauss J, Oakley BR, Wang CC, Keller NP (2009) Chromatin-level regulation of biosynthetic gene clusters. Nat Chem Biol 5(7):462–464. doi:10.1038/nchembio.177
16. Bok JW, Hoffmeister D, Maggio-Hall LA, Murillo R, Glasner JD, Keller NP (2006) Genomic mining for *Aspergillus* natural products. Chem Biol 13(1):31–37. doi:10.1016/j.chembiol.2005.10.008
17. Burman R, Gruber CW, Rizzardi K, Herrmann A, Craik DJ, Gupta MP, Goransson U (2010) Cyclotide proteins and precursors from the genus *Gloeospermum*: filling a blank spot in the cyclotide map of Violaceae. Phytochemistry 71(1):13–20. doi:10.1016/j.phytochem.2009.09.023
18. Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, Clardy J (2007) The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. Proc Natl Acad Sci USA 104(5):1506–1509. doi:10.1073/pnas.0610503104
19. Cane DE, Watt RM (2003) Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. Proc Natl Acad Sci USA 100(4):1547–1551. doi:10.1073/pnas.0337625100
20. Challis GL (2008) Genome mining for novel natural product discovery. J Med Chem 51(9):2618–2628. doi:10.1021/jm700948z
21. Challis GL, Ravel J (2000) Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. FEMS Microbiol Lett 187(2):111–114. doi:10.1111/j.1574-6968.2000.tb09145.x
22. Chen XH, Koumoutsi A, Scholz R, Eisenreich A, Schneider K, Heinemeyer I, Morgenstern B, Voss B, Hess WR, Reva O, Junge H, Voigt B, Jungblut PR, Vater J, Sussmuth R, Liesegang H, Strittmatter A, Gottschalk G, Borriss R (2007) Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. Nat Biotechnol 25(9):1007–1014. doi:10.1038/nbt1325
23. Chen XH, Scholz R, Borriss M, Junge H, Mogel G, Kunz S, Borriss R (2009) Difficidin and bacilysin produced by plant-associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. J Biotechnol 140(1–2):38–44. doi:10.1016/j.jbiotec.2008.10.015
24. Chen XH, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, Gottschalk G, Sussmuth RD, Borriss R (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB42. J Bacteriol 188(11):4024–4036. doi:10.1128/JB.00052-06
25. Chen Y, McClure RA, Zheng Y, Thomson RJ, Kelleher NL (2013) Proteomics guided discovery of flavopeptins: anti-proliferative aldehydes synthesized by a reductase domain-containing nonribosomal peptide synthetase. J Am Chem Soc. doi:10.1021/ja4031193
26. Chertkov O, Sikorski J, Nolan M, Lapidus A, Lucas S, Del Rio TG, Tice H, Cheng JF, Goodwin L, Pitluck S, Liolios K, Ivanova N, Mavromatis K, Mikhailova N, Ovchinnikova G, Pati A, Chen A, Palaniappan K, Djao OD, Land M, Hauser L, Chang YJ, Jeffries CD, Brettin T, Han C, Detter JC, Rohde M, Goker M, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Klenk HP, Kyrpides NC (2011) Complete genome sequence of *Thermomonospora curvata* type strain (B9). Stand Genomic Sci 4(1):13–22. doi:10.4056/sigs.1453580
27. Chiang YM, Szewczyk E, Davidson AD, Keller N, Oakley BR, Wang CC (2009) A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a

polyketide, asperfuranone, *Aspergillus nidulans*. J Am Chem Soc 131(8):2965–2970. doi:10.1021/ja8088185
28. Chiang YM, Szewczyk E, Nayak T, Davidson AD, Sanchez JF, Lo HC, Ho WY, Simityan H, Kuo E, Praseuth A, Watanabe K, Oakley BR, Wang CC (2008) Molecular genetic mining of the *Aspergillus* secondary metabolome: discovery of the emericellamide biosynthetic pathway. Chem Biol 15(6):527–532. doi:10.1016/j.chembiol.2008.05.010
29. Chong PP, Podmore SM, Kieser HM, Redenbach M, Turgay K, Marahiel M, Hopwood DA, Smith CP (1998) Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2). Microbiology 144(Pt 1):193–199. doi:10.1099/00221287-144-1-193
30. Chooi YH, Fang J, Liu H, Filler SG, Wang P, Tang Y (2013) Genome mining of a prenylated and immunosuppressive polyketide from pathogenic fungi. Org Lett 15(4):780–783. doi:10.1021/ol303435y
31. Claesen J, Bibb M (2010) Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. Proc Natl Acad Sci USA 107(37):16297–16302. doi:10.1073/pnas.1008608107
32. Claesen J, Bibb MJ (2011) Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by *Streptomyces griseus* IFO 13350. J Bacteriol 193(10):2510–2516. doi:10.1128/JB.00171-11
33. Copeland A, Lapidus A, Glavina Del Rio T, Nolan M, Lucas S, Chen F, Tice H, Cheng JF, Bruce D, Goodwin L, Pitluck S, Mikhailova N, Pati A, Ivanova N, Mavromatis K, Chen A, Palaniappan K, Chain P, Land M, Hauser L, Chang YJ, Jeffries CD, Chertkov O, Brettin T, Detter JC, Han C, Ali Z, Tindall BJ, Goker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP (2009) Complete genome sequence of *Catenulispora acidiphila* type strain (ID 139908). Stand Genomic Sci 1 (2):119–125. doi:10.4056/sigs.17259
34. Craney A, Ahmed S, Nodwell J (2013) Towards a new science of secondary metabolism. J Antibiot 66:387–400. doi:10.1038/ja.2013.25
35. Crone WJK, Leeper FJ, Truman AW (2012) Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. Chem Sci 3(12):3516–3521. doi:10.1039/C2sc21190d
36. Daly KM, Upton M, Sandiford SK, Draper LA, Wescombe PA, Jack RW, O'Connor PM, Rossney A, Gotz F, Hill C, Cotter PD, Ross RP, Tagg JR (2010) Production of the Bsa lantibiotic by community-acquired *Staphylococcus aureus* strains. J Bacteriol 192(4):1131–1142. doi:10.1128/JB.01375-09
37. de Jong A, van Hijum SA, Bijlsma JJ, Kok J, Kuipers OP (2006) BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Res 34 (Web Server issue):W273-279. doi:10.1093/nar/gkl237
38. Deane CD, Melby JO, Molohon KJ, Susarrey AR, Mitchell DA (2013) Engineering unnatural variants of plantazolicin through codon reprogramming. ACS Chem Biol 8:1998–2008. doi:10.1021/cb4003392
39. Diep DB, Godager L, Brede D, Nes IF (2006) Data mining and characterization of a novel pediocin-like bacteriocin system from the genome of *Pediococcus pentosaceus* ATCC 25745. Microbiology 152(Pt 6):1649–1659. doi:10.1099/mic.0.28794-0
40. Dischinger J, Josten M, Szekat C, Sahl HG, Bierbaum G (2009) Production of the novel two-peptide lantibiotic lichenicidin by *Bacillus licheniformis* DSM 13. PLoS ONE 4(8):e6788. doi:10.1371/journal.pone.0006788
41. Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S, Bocs S, Boursaux-Eude C, Chandler M, Charles JF, Dassa E, Derose R, Derzelle S, Freyssinet G, Gaudriault S, Medigue C, Lanois A, Powell K, Siguier P, Vincent R, Wingate V, Zouine M, Glaser P, Boemare N, Danchin A, Kunst F (2003) The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nat Biotechnol 21(11):1307–1313. doi:10.1038/nbt886
42. Dunbar KL, Melby JO, Mitchell DA (2012) YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. Nat Chem Biol 8(6):569–575. doi:10.1038/nchembio.944
43. Engelhardt K, Degnes KF, Kemmler M, Bredholt H, Fjaervik E, Klinkenberg G, Sletta H, Ellingsen TE, Zotchev SB (2010) Production of a new thiopeptide antibiotic, TP-1161, by a marine *Nocardiopsis* species. Appl Environ Microbiol 76(15):4969–4976. doi:10.1128/AEM.00741-10
44. Evans BS, Ntai I, Chen Y, Robinson SJ, Kelleher NL (2011) Proteomics-based discovery of koranimine, a cyclic imine natural product. J Am Chem Soc 133(19):7316–7319. doi:10.1021/ja2015795
45. Fazio GC, Xu R, Matsuda SP (2004) Genome mining to identify new plant triterpenoids. J Am Chem Soc 126(18):5678–5679. doi:10.1021/ja0318784
46. Feitelson JS, Malpartida F, Hopwood DA (1985) Genetic and biochemical characterization of the red gene cluster of *Streptomyces coelicolor* A3(2). J Gen Microbiol 131(9):2431–2441. doi:10.1099/00221287-131-9-2431
47. Feng L, Wang W, Cheng J, Ren Y, Zhao G, Gao C, Tang Y, Liu X, Han W, Peng X, Liu R, Wang L (2007) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. Proc Natl Acad Sci USA 104(13):5602–5607. doi:10.1073/pnas.0609650104
48. Franke J, Ishida K, Hertweck C (2012) Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic *Burkholderia* species. Angew Chem Int Ed 51(46):11611–11615. doi:10.1002/anie.201205566
49. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Penalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438(7071):1105–1115. doi:10.1038/nature04341
50. Garg N, Tang W, Goto Y, Nair SK, van der Donk WA (2012) Lantibiotics from *Geobacillus thermodenitrificans*. Proc Natl Acad Sci USA 109(14):5241–5246. doi:10.1073/pnas.1116815109
51. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB (2006) Evolution of sensory complexity recorded in a myxobacterial genome. Proc Natl Acad Sci USA 103(41):15200–15205. doi:10.1073/pnas.0607335103
52. Gomez-Escribano JP, Bibb MJ (2013) Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J Ind Microbiol Biot (in press)
53. Gomez-Escribano JP, Song LJ, Bibb MJ, Challis GL (2012) Posttranslational beta-methylation and macrolactamidination

in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. Chem Sci 3(12):3522–3525. doi:10.1039/C2sc21183a
54. Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA (2010) Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. PLoS Biol 8(3):e1000339. doi:10.1371/journal.pbio.1000339
55. Gross H (2009) Genomic mining–a concept for the discovery of new bioactive natural products. Curr Opin Drug Discov Devel 12(2):207–219
56. Gross H, Stockwell VO, Henkels MD, Nowak-Thompson B, Loper JE, Gerwick WH (2007) The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. Chem Biol 14(1):53–63. doi:10.1016/j.chembiol.2006.11.007
57. Haft DH (2009) A strain-variable bacteriocin in *Bacillus anthracis* and *Bacillus cereus* with repeated Cys-Xaa-Xaa motifs. Biol Direct 4:15. doi:10.1186/1745-6150-4-15
58. Haft DH, Basu MK, Mitchell DA (2010) Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. BMC Biol 8:70. doi:10.1186/1741-7007-8-70
59. Hegemann JD, Zimmermann M, Xie X, Marahiel MA (2013) Caulosegnins I-III: a highly diverse group of lasso peptides derived from a single biosynthetic gene cluster. J Am Chem Soc 135(1):210–222. doi:10.1021/ja308173b
60. Heneghan MN, Yakasai AA, Williams K, Kadir KA, Wasil Z, Bakeer W, Fisch KM, Bailey AM, Simpson TJ, Cox RJ, Lazarus CM (2011) The programming role of trans-acting enoyl reductases during the biosynthesis of highly reduced fungal polyketides. Chem Sci 2(5):972–979. doi:10.1039/C1sc00023c
61. Henrikson JC, Hoover AR, Joyner PM, Cichewicz RH (2009) A chemical epigenetics approach for engineering the in situ biosynthesis of a cryptic natural product from *Aspergillus niger*. Org Biomol Chem 7(3):435–438. doi:10.1039/b819208a
62. Herrmann A, Burman R, Mylne JS, Karlsson G, Gullbo J, Craik DJ, Clark RJ, Goransson U (2008) The alpine violet, *Viola biflora*, is a rich source of cyclotides with potent cytotoxicity. Phytochemistry 69(4):939–952. doi:10.1016/j.phytochem.2007.10.023
63. Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R, Barron A, Bason N, Bentley SD, Chillingworth C, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James KD, Lennard N, Line A, Mayes R, Moule S, Mungall K, Ormond D, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S, Barrell BG, Spratt BG, Parkhill J (2004) Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci USA 101(26):9786–9791. doi:10.1073/pnas.0402521101
64. Hornung A, Bertazzo M, Dziarnowski A, Schneider K, Welzel K, Wohlert SE, Holzenkampfer M, Nicholson GJ, Bechthold A, Sussmuth RD, Vente A, Pelzer S (2007) A genomic screening approach to the structure-guided identification of drug candidates from natural sources. ChemBioChem 8(7):757–766. doi:10.1002/cbic.200600375
65. Hoskins J, Alborn WE Jr, Arnold J, Blaszczak LC, Burgett S, DeHoff BS, Estrem ST, Fritz L, Fu DJ, Fuller W, Geringer C, Gilmour R, Glass JS, Khoja H, Kraft AR, Lagace RE, LeBlanc DJ, Lee LN, Lefkowitz EJ, Lu J, Matsushima P, McAhren SM, McHenney M, McLeaster K, Mundy CW, Nicas TI, Norris FH, O'Gara M, Peery RB, Robertson GT, Rockey P, Sun PM, Winkler ME, Yang Y, Young-Bellido M, Zhao G, Zook CA, Baltz RH, Jaskunas SR, Rosteck PR Jr, Skatrud PL, Glass JI (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. J Bacteriol 183(19):5709–5717. doi:10.1128/JB.183.19.5709- 5717.2001
66. Hou Y, Tianero MD, Kwan JC, Wyche TP, Michel CR, Ellis GA, Vazquez-Rivera E, Braun DR, Rose WE, Schmidt EW, Bugni TS (2012) Structure and biosynthesis of the antibiotic bottromycin D. Org Lett 14(19):5050–5053. doi:10.1021/ol3022758
67. Huntley S, Hamann N, Wegener-Feldbrugge S, Treuner-Lange A, Kube M, Reinhardt R, Klages S, Muller R, Ronning CM, Nierman WC, Sogaard-Andersen L (2011) Comparative genomic analysis of fruiting body formation in Myxococcales. Mol Biol Evol 28(2):1083–1097. doi:10.1093/molbev/msq292
68. Huo L, Rachid S, Stadler M, Wenzel SC, Muller R (2012) Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. Chem Biol 19(10):1278–1287. doi:10.1016/j.chembiol.2012.08.013
69. Ikeda H, Shin-ya K, Omura S (2013) Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. J Ind Microbiol Biot. doi:10.1007/s10295-013-1327-x
70. Ishida K, Christiansen G, Yoshida WY, Kurmayer R, Welker M, Valls N, Bonjoch J, Hertweck C, Borner T, Hemscheidt T, Dittmann E (2007) Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. Chem Biol 14(5):565–576. doi:10.1016/j.chembiol.2007.04.006
71. Kalyon B, Helaly SE, Scholz R, Nachtigall J, Vater J, Borriss R, Sussmuth RD (2011) Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. Org Lett 13(12):2996–2999. doi:10.1021/ol200809m
72. Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M, Nakamura Y, Kasai F, Watanabe A, Kawashima K, Kishida Y, Ono A, Shimizu Y, Takahashi C, Minami C, Fujishiro T, Kohara M, Katoh M, Nakazaki N, Nakayama S, Yamada M, Tabata S, Watanabe MM (2007) Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. DNA Res 14(6):247–256. doi:10.1093/dnares/dsm026
73. Kelly WL, Pan L, Li C (2009) Thiostrepton biosynthesis: prototype for a new family of bacteriocins. J Am Chem Soc 131(12):4327–4334. doi:10.1021/ja807890a
74. Kersten RD, Yang YL, Xu Y, Cimermancic P, Nam SJ, Fenical W, Fischbach MA, Moore BS, Dorrestein PC (2011) A mass spectrometry-guided genome mining approach for natural product peptidogenomics. Nat Chem Biol 7(11):794–802. doi:10.1038/nchembio.684
75. Kim HS, Schell MA, Yu Y, Ulrich RL, Sarria SH, Nierman WC, DeShazer D (2005) Bacterial genome adaptation to niches: divergence of the potential virulence genes in three *Burkholderia* species of different survival strategies. BMC Genomics 6:174. doi:10.1186/1471-2164-6-174
76. Knappe TA, Linne U, Zirah S, Rebuffat S, Xie X, Marahiel MA (2008) Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of *Burkholderia thailandensis* E264. J Am Chem Soc 130(34):11446–11454. doi:10.1021/ja802966g
77. Kodani S, Hudson ME, Durrant MC, Buttner MJ, Nodwell JR, Willey JM (2004) The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in *Streptomyces coelicolor*. Proc Natl Acad Sci USA 101(31):11448–11453. doi:10.1073/pnas.0404220101
78. Komatsu M, Tsuda M, Omura S, Oikawa H, Ikeda H (2008) Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. Proc Natl Acad Sci USA 105(21):7422–7427. doi:10.1073/pnas.0802312105

79. Konig CC, Scherlach K, Schroeckh V, Horn F, Nietzsche S, Brakhage AA, Hertweck C (2013) Bacterium induces cryptic meroterpenoid pathway in the pathogenic fungus *Aspergillus fumigatus*. ChemBioChem 14(8):938–942. doi:10.1002/cbic.201300070
80. Krawczyk B, Voller GH, Voller J, Ensle P, Sussmuth RD (2012) Curvopeptin: a new lanthionine-containing class III lantibiotic and its co-substrate promiscuous synthetase. ChemBioChem 13(14):2065–2071. doi:10.1002/cbic.201200417
81. Kunze B, Reichenbach H, Muller R, Hofle G (2005) Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (Myxobacteria). Fermentation, isolation, physico-chemical properties, structure and biological activity. J Antibiot 58(4):244–251. doi:10.1038/ja.2005.28
82. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, Aigle B (2011) Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. Proc Natl Acad Sci USA 108(15):6258–6263. doi:10.1073/pnas.1019077108
83. Lautru S, Deeth RJ, Bailey LM, Challis GL (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nat Chem Biol 1(5):265–269. doi:10.1038/nchembio731
84. Lawton EM, Cotter PD, Hill C, Ross RP (2007) Identification of a novel two-peptide lantibiotic, haloduracin, produced by the alkaliphile *Bacillus halodurans* C-125. FEMS Microbiol Lett 267(1):64–71. doi:10.1111/j.1574-6968.2006.00539.x
85. Lazos O, Tosin M, Slusarczyk AL, Boakes S, Cortes J, Sidebottom PJ, Leadlay PF (2010) Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. Chem Biol 17(2):160–173. doi:10.1016/j.chembiol.2010.01.011
86. Lee SW, Mitchell DA, Markley AL, Hensler ME, Gonzalez D, Wohlrab A, Dorrestein PC, Nizet V, Dixon JE (2008) Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci USA 105(15):5879–5884. doi:10.1073/pnas.0801338105
87. Leikoski N, Fewer DP, Jokela J, Wahlsten M, Rouhiainen L, Sivonen K (2010) Highly diverse cyanobactins in strains of the genus *Anabaena*. Appl Environ Microbiol 76(3):701–709. doi:10.1128/AEM.01061-09
88. Li B, Sher D, Kelly L, Shi Y, Huang K, Knerr PJ, Joewono I, Rusch D, Chisholm SW, van der Donk WA (2010) Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. Proc Natl Acad Sci USA 107(23):10430–10435. doi:10.1073/pnas.0913677107
89. Liao R, Duan L, Lei C, Pan H, Ding Y, Zhang Q, Chen D, Shen B, Yu Y, Liu W (2009) Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. Chem Biol 16(2):141–147. doi:10.1016/j.chembiol.2009.01.007
90. Liu T, Chiang YM, Somoza AD, Oakley BR, Wang CC (2011) Engineering of an "unnatural" natural product by swapping polyketide synthase domains in *Aspergillus nidulans*. J Am Chem Soc 133(34):13314–13316. doi:10.1021/ja205780g
91. Liu X, Biswas S, Berg MG, Antapli CM, Xie F, Wang Q, Tang MC, Tang GL, Zhang L, Dreyfuss G, Cheng YQ (2013) Genomics-guided discovery of thailanstatins A, B, and C as pre-mRNA splicing inhibitors and antiproliferative agents from *Burkholderia thailandensis* MSMB43. J Nat Prod. doi:10.1021/np300913h
92. Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto K, Arima T, Akita O, Kashiwagi Y, Abe K, Gomi K, Horiuchi H, Kitamoto K, Kobayashi T, Takeuchi M, Denning DW, Galagan JE, Nierman WC, Yu J, Archer DB, Bennett JW, Bhatnagar D, Cleveland TE, Fedorova ND, Gotoh O, Horikawa H, Hosoyama A, Ichinomiya M, Igarashi R, Iwashita K, Juvvadi PR, Kato M, Kato Y, Kin T, Kokubun A, Maeda H, Maeyama N, Maruyama J, Nagasaki H, Nakajima T, Oda K, Okada K, Paulsen I, Sakamoto K, Sawano T, Takahashi M, Takase K, Terabayashi Y, Wortman JR, Yamada O, Yamagata Y, Anazawa H, Hata Y, Koide Y, Komori T, Koyama Y, Minetoki T, Suharnan S, Tanaka A, Isono K, Kuhara S, Ogasawara N, Kikuchi H (2005) Genome sequencing and analysis of *Aspergillus oryzae*. Nature 438(7071):1157–1161. doi:10.1038/nature04300
93. Majchrzykiewicz JA, Lubelski J, Moll GN, Kuipers A, Bijlsma JJ, Kuipers OP, Rink R (2010) Production of a class II two-component lantibiotic of *Streptococcus pneumoniae* using the class I nisin synthetic machinery and leader sequence. Antimicrob Agents Chemother 54(4):1498–1505. doi:10.1128/AAC.00883-09
94. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee JH, Diaz-Muniz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D (2006) Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 103(42):15611–15616. doi:10.1073/pnas.0607117103
95. Maksimov MO, Pelczer I, Link AJ (2012) Precursor-centric genome-mining approach for lasso peptide discovery. Proc Natl Acad Sci USA 109(38):15223–15228. doi:10.1073/pnas.1208978109
96. Malpartida F, Hopwood DA (1984) Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. Nature 309(5967):462–464. doi:10.1038/309462a0
97. Malpartida F, Hopwood DA (1986) Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). Mol Gen Genet 205(1):66–73. doi:10.1007/BF02428033
98. Martin R, Sterner O, Alvarez MA, de Clercq E, Bailey JE, Minas W (2001) Collinone, a new recombinant angular polyketide antibiotic made by an engineered *Streptomyces* strain. J Antibiot 54(3):239–249. doi:10.7164/antibiotics.54.239
99. Martinez DA, Oliver BG, Graser Y, Goldberg JM, Li W, Martinez-Rossi NM, Monod M, Shelest E, Barton RC, Birch E, Brakhage AA, Chen Z, Gurr SJ, Heiman D, Heitman J, Kosti I, Rossi A, Saif S, Samalova M, Saunders CW, Shea T, Summerbell RC, Xu J, Young S, Zeng Q, Birren BW, Cuomo CA, White TC (2012) Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. mBio 3 (5):e00259–00212. doi:10.1128/mBio.00259-12
100. McAlpine JB, Bachmann BO, Piraee M, Tremblay S, Alarco AM, Zazopoulos E, Farnet CM (2005) Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. J Nat Prod 68(4):493–496. doi:10.1021/np0401664
101. McAlpine JB, Banskota AH, Charan RD, Schlingmann G, Zazopoulos E, Piraee M, Janso J, Bernan VS, Aouidate M, Farnet CM, Feng X, Zhao Z, Carter GT (2008) Biosynthesis of diazepinomicin/ECO-4601, a *Micromonospora* secondary metabolite with a novel ring system. J Nat Prod 71(9):1585–1590. doi:10.1021/np800376n
102. McClerren AL, Cooper LE, Quan C, Thomas PM, Kelleher NL, van der Donk WA (2006) Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. Proc Natl Acad Sci USA 103(46):17243–17248. doi:10.1073/pnas.0606088103

103. Meiser P, Bode HB, Muller R (2006) The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. Proc Natl Acad Sci USA 103(50):19128–19133. doi:10.1073/pnas.0606039103
104. Melby JO, Nard NJ, Mitchell DA (2011) Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. Curr Opin Chem Biol 15(3):369–378. doi:10.1016/j.cbpa.2011.02.027
105. Molohon KJ, Melby JO, Lee J, Evans BS, Dunbar KL, Bumpus SB, Kelleher NL, Mitchell DA (2011) Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. ACS Chem Biol 6(12):1307–1313. doi:10.1021/cb200339d
106. Morris RP, Leeds JA, Naegeli HU, Oberer L, Memmert K, Weber E, LaMarche MJ, Parker CN, Burrer N, Esterow S, Hein AE, Schmitt EK, Krastel P (2009) Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. J Am Chem Soc 131(16):5946–5955. doi:10.1021/ja900488a
107. Nakazawa T, Ishiuchi K, Praseuth A, Noguchi H, Hotta K, Watanabe K (2012) Overexpressing transcriptional regulator in *Aspergillus oryzae* activates a silent biosynthetic pathway to produce a novel polyketide. ChemBioChem 13(6):855–861. doi:10.1002/cbic.201200107
108. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75(3):311–335. doi:10.1021/np200906s
109. Nguyen T, Ishida K, Jenke-Kodama H, Dittmann E, Gurgui C, Hochmuth T, Taudien S, Platzer M, Hertweck C, Piel J (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat Biotechnol 26(2):225–233. doi:10.1038/nbt1379
110. Nielsen ML, Nielsen JB, Rank C, Klejnstrup ML, Holm DK, Brogaard KH, Hansen BG, Frisvad JC, Larsen TO, Mortensen UH (2011) A genome-wide polyketide synthase deletion library uncovers novel genetic links to polyketides and meroterpenoids in *Aspergillus nidulans*. FEMS Microbiol Lett 321(2):157–166. doi:10.1111/j.1574-6968.2011.02327.x
111. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, Bennett J, Bowyer P, Chen D, Collins M, Coulsen R, Davies R, Dyer PS, Farman M, Fedorova N, Fedorova N, Feldblyum TV, Fischer R, Fosker N, Fraser A, Garcia JL, Garcia MJ, Goble A, Goldman GH, Gomi K, Griffith-Jones S, Gwilliam R, Haas B, Haas H, Harris D, Horiuchi H, Huang J, Humphray S, Jimenez J, Keller N, Khouri H, Kitamoto K, Kobayashi T, Konzack S, Kulkarni R, Kumagai T, Lafon A, Latge JP, Li W, Lord A, Lu C, Majoros WH, May GS, Miller BL, Mohamoud Y, Molina M, Monod M, Mouyna I, Mulligan S, Murphy L, O'Neil S, Paulsen I, Penalva MA, Pertea M, Price C, Pritchard BL, Quail MA, Rabbinowitsch E, Rawlins N, Rajandream MA, Reichard U, Renauld H, Robson GD, Rodriguez de Cordoba S, Rodriguez-Pena JM, Ronning CM, Rutter S, Salzberg SL, Sanchez M, Sanchez-Ferrero JC, Saunders D, Seeger K, Squares R, Squares S, Takeuchi M, Tekaia F, Turner G, Vazquez de Aldana CR, Weidman J, White O, Woodward J, Yu JH, Fraser C, Galagan JE, Asai K, Machida M, Hall N, Barrell B, Denning DW (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature 438(7071):1151–1156. doi:10.1038/nature04332
112. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. J Bacteriol 190(11):4050–4060. doi:10.1128/JB.00204-08
113. Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. Nat Biotechnol 25(4):447–453. doi:10.1038/nbt1297
114. Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc Natl Acad Sci USA 98(21):12215–12220. doi:10.1073/pnas.211433198
115. Patel PS, Huang S, Fisher S, Pirnik D, Aklonis C, Dean L, Meyers E, Fernandes P, Mayerl F (1995) Bacillaene, a novel inhibitor of procaryotic protein synthesis produced by *Bacillus subtilis*: production, taxonomy, isolation, physico-chemical characterization and biological activity. J Antibiot 48(9):997–1003. doi:10.7164/antibiotics.48.997
116. Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GS, Mavrodi DV, DeBoy RT, Seshadri R, Ren Q, Madupu R, Dodson RJ, Durkin AS, Brinkac LM, Daugherty SC, Sullivan SA, Rosovitz MJ, Gwinn ML, Zhou L, Schneider DJ, Cartinhour SW, Nelson WC, Weidman J, Watkins K, Tran K, Khouri H, Pierson EA, Pierson LS 3rd, Thomashow LS, Loper JE (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. Nat Biotechnol 23(7):873–878. doi:10.1038/nbt1110
117. Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E (2007) A cryptic type I polyketide synthase (cpk) gene cluster in *Streptomyces coelicolor* A3(2). Arch Microbiol 187(2):87–99. doi:10.1007/s00203-006-0176-7
118. Pawlik K, Kotowska M, Kolesinski P (2010) *Streptomyces coelicolor* A3(2) produces a new yellow pigment associated with the polyketide synthase Cpk. J Mol Microb Biotech 19(3):147–151. doi:10.1159/000321501
119. Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albang R, Albermann K, Andersen MR, Bendtsen JD, Benen JA, van den Berg M, Breestraat S, Caddick MX, Contreras R, Cornell M, Coutinho PM, Danchin EG, Debets AJ, Dekker P, van Dijck PW, van Dijk A, Dijkhuizen L, Driessen AJ, d'Enfert C, Geysens S, Goosen C, Groot GS, de Groot PW, Guillemette T, Henrissat B, Herweijer M, van den Hombergh JP, van den Hondel CA, van der Heijden RT, van der Kaaij RM, Klis FM, Kools HJ, Kubicek CP, van Kuyk PA, Lauber J, Lu X, van der Maarel MJ, Meulenberg R, Menke H, Mortimer MA, Nielsen J, Oliver SG, Olsthoorn M, Pal K, van Peij NN, Ram AF, Rinas U, Roubos JA, Sagt CM, Schmoll M, Sun J, Ussery D, Varga J, Vervecken W, van de Vondervoort PJ, Wedler H, Wosten HA, Zeng AP, van Ooyen AJ, Visser J, Stam H (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. Nat Biotechnol 25(2):221–231. doi:10.1038/nbt1282
120. Pistorius D, Muller R (2012) Discovery of the rhizopodin biosynthetic gene cluster in *Stigmatella aurantiaca* Sg a15 by genome mining. ChemBioChem 13(3):416–426. doi:10.1002/cbic.201100575
121. Poralla K, Muth G, Hartner T (2000) Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). FEMS Microbiol Lett 189(1):93–95. doi:10.1111/j.1574-6968.2000.tb09212.x
122. Portmann C, Blom JF, Kaiser M, Brun R, Juttner F, Gademann K (2008) Isolation of aerucyclamides C and D and structure revision of microcyclamide 7806A: heterocyclic ribosomal peptides from *Microcystis aeruginosa* PCC 7806 and their antiparasite evaluation. J Nat Prod 71(11):1891–1896. doi:10.1021/np800409z

123. Pukall R, Lapidus A, Glavina Del Rio T, Copeland A, Tice H, Cheng JF, Lucas S, Chen F, Nolan M, Labutti K, Pati A, Ivanova N, Mavromatis K, Mikhailova N, Pitluck S, Bruce D, Goodwin L, Land M, Hauser L, Chang YJ, Jeffries CD, Chen A, Palaniappan K, Chain P, Rohde M, Goker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Brettin T (2010) Complete genome sequence of *Kribbella flavida* type strain (IFO 14399). Stand Genomic Sci 2(2):186–193. doi:10.4056/sigs.731321
124. Pullan ST, Chandra G, Bibb MJ, Merrick M (2011) Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. BMC Genomics 12:175. doi:10.1186/1471-2164-12-175
125. Qiao K, Zhou H, Xu W, Zhang W, Garg N, Tang Y (2011) A fungal nonribosomal peptide synthetase module that can synthesize thiopyrazines. Org Lett 13(7):1758–1761. doi:10.1021/ol200288w
126. Rey MW, Ramaiya P, Nelson BA, Brody-Karpin SD, Zaretsky EJ, Tang M, Lopez de Leon A, Xiang H, Gusti V, Clausen IG, Olsen PB, Rasmussen MD, Andersen JT, Jorgensen PL, Larsen TS, Sorokin A, Bolotin A, Lapidus A, Galleron N, Ehrlich SD, Berka RM (2004) Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. Genome Biol 5(10):R77. doi:10.1186/gb-2004-5-10-r77
127. Robbel L, Knappe TA, Linne U, Xie X, Marahiel MA (2010) Erythrochelin–a hydroxamate-type siderophore predicted from the genome of *Saccharopolyspora erythraea*. FEBS J 277(3):663–676. doi:10.1111/j.1742-4658.2009.07512.x
128. Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, Chisholm SW (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424(6952):1042–1047. doi:10.1038/nature01947
129. Sanchez JF, Chiang YM, Szewczyk E, Davidson AD, Ahuja M, Elizabeth Oakley C, Woo Bok J, Keller N, Oakley BR, Wang CC (2010) Molecular genetic analysis of the orsellinic acid/F9775 gene cluster of *Aspergillus nidulans*. Mol Biosyst 6(3):587–593. doi:10.1039/b904541d
130. Sanchez JF, Somoza AD, Keller NP, Wang CC (2012) Advances in *Aspergillus* secondary metabolite research in the postgenomic era. Nat Prod Rep 29(3):351–371. doi:10.1039/c2np00084a
131. Sarkar A, Funk AN, Scherlach K, Horn F, Schroeckh V, Chankhamjon P, Westermann M, Roth M, Brakhage AA, Hertweck C, Horn U (2012) Differential expression of silent polyketide biosynthesis gene clusters in chemostat cultures of *Aspergillus nidulans*. J Biotechnol 160(1–2):64–71. doi:10.1016/j.jbiotec.2012.01.015
132. Scherlach K, Hertweck C (2006) Discovery of aspoquinolones A-D, prenylated quinoline-2-one alkaloids from *Aspergillus nidulans*, motivated by genome mining. Org Biomol Chem 4(18):3517–3520. doi:10.1039/b607011f
133. Scherlach K, Schuemann J, Dahse HM, Hertweck C (2010) Aspernidine A and B, prenylated isoindolinone alkaloids from the model fungus *Aspergillus nidulans*. J Antibiot 63(7):375–377. doi:10.1038/ja.2010.46
134. Scholz R, Molohon KJ, Nachtigall J, Vater J, Markley AL, Sussmuth RD, Mitchell DA, Borriss R (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. J Bacteriol 193(1):215–224. doi:10.1128/JB.00784-10
135. Schroeckh V, Scherlach K, Nutzmann HW, Shelest E, Schmidt-Heck W, Schuemann J, Martin K, Hertweck C, Brakhage AA (2009) Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. Proc Natl Acad Sci USA 106(34):14558–14563. doi:10.1073/pnas.0901870106
136. Seshime Y, Juvvadi PR, Kitamoto K, Ebizuka Y, Fujii I (2010) Identification of csypyrone B1 as the novel product of *Aspergillus oryzae* type III polyketide synthase CsyB. Bioorg Med Chem 18(12):4542–4546. doi:10.1016/j.bmc.2010.04.058
137. Shenkarev ZO, Finkina EI, Nurmukhamedova EK, Balandin SV, Mineev KS, Nadezhdin KD, Yakimenko ZA, Tagaev AA, Temirov YV, Arseniev AS, Ovchinnikova TV (2010) Isolation, structure elucidation, and synergistic antibacterial activity of a novel two-component lantibiotic lichenicidin from *Bacillus licheniformis* VK21. Biochemistry 49(30):6462–6472. doi:10.1021/bi100871b
138. Song L, Barona-Gomez F, Corre C, Xiang L, Udwary DW, Austin MB, Noel JP, Moore BS, Challis GL (2006) Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by *Streptomyces coelicolor* genome mining. J Am Chem Soc 128(46):14754–14755. doi:10.1021/ja065247w
139. Sorensen JL, Hansen FT, Sondergaard TE, Staerk D, Lee TV, Wimmer R, Klitgaard LG, Purup S, Giese H, Frandsen RJ (2012) Production of novel fusarielins by ectopic activation of the polyketide synthase 9 cluster in *Fusarium graminearum*. Environ Microbiol 14(5):1159–1170. doi:10.1111/j.1462-2920.2011.02696.x
140. Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res 36(21):6882–6892. doi:10.1093/nar/gkn685
141. Sudek S, Haygood MG, Youssef DT, Schmidt EW (2006) Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. Appl Environ Microbiol 72(6):4382–4387. doi:10.1128/AEM.00380-06
142. Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with Bacillus subtilis. Nucleic Acids Res 28(21):4317–4331. doi:10.1093/nar/28.21.4317
143. Takano E (2006) Gamma-butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. Curr Opin Microbiol 9(3):287–294. doi:10.1016/j.mib.2006.04.003
144. Tang W, van der Donk WA (2012) Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria. Biochemistry 51(21):4271–4279. doi:10.1021/bi300255s
145. Teng Y, Zhao WP, Qian CD, Li O, Zhu L, Wu XC (2012) Gene cluster analysis for the biosynthesis of elgicins, novel lantibiotics produced by *Paenibacillus elgii* B69. BMC Microbiol 12. doi:10.1186/1471-2180-12-45
146. Tohyama S, Eguchi T, Dhakal RP, Akashi T, Otsuka M, Kakinuma K (2004) Genome-inspired search for new antibiotics. Isolation and structure determination of new 28-membered polyketide macrolactones, halstoctacosanolides A and B, from *Streptomyces halstedii* HC34. Tetrahedron 60(18):3999–4005. doi:10.1016/j.tet.2004.03.027
147. Tohyama S, Kakinuma K, Eguchi T (2006) The complete biosynthetic gene cluster of the 28-membered polyketide

macrolactones, halstoctacosanolides, from *Streptomyces halstedii* HC34. J Antibiot 59(1):44–52. doi:10.1038/ja.2006.7
148. Trabi M, Mylne JS, Sando L, Craik DJ (2009) Circular proteins from *Melicytus* (Violaceae) refine the conserved protein and gene architecture of cyclotides. Org Biomol Chem 7(11):2378–2388. doi:10.1039/b823020j
149. Vassiliadis G, Destoumieux-Garzon D, Lombard C, Rebuffat S, Peduzzi J (2010) Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. Antimicrob Agents Chemother 54(1):288–297. doi:10.1128/AAC.00744-09
150. Velasquez JE, van der Donk WA (2011) Genome mining for ribosomally synthesized natural products. Curr Opin Chem Biol 15(1):11–21. doi:10.1016/j.cbpa.2010.10.027
151. Voller GH, Krawczyk B, Ensle P, Sussmuth RD (2013) Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation. J Am Chem Soc. doi:10.1021/ja402296m
152. Wang C, Henkes LM, Doughty LB, He M, Wang D, Meyer-Almes FJ, Cheng YQ (2011) Thailandepsins: bacterial products with potent histone deacetylase inhibitory activities and broad-spectrum antiproliferative activities. J Nat Prod 74(10):2031–2038. doi:10.1021/np200324x
153. Wang H, Sivonen K, Rouhiainen L, Fewer DP, Lyra C, Rantala-Ylinen A, Vestola J, Jokela J, Rantasarkka K, Li Z, Liu B (2012) Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium *Anabaena* sp. strain 90. BMC Genomics 13:613. doi:10.1186/1471-2164-13-613
154. Wang H, van der Donk WA (2012) Biosynthesis of the class III lantipeptide catenulipeptin. ACS Chem Biol 7(9):1529–1535. doi:10.1021/cb3002446
155. Wenzel SC, Kunze B, Hofle G, Silakowski B, Scharfe M, Blocker H, Muller R (2005) Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. ChemBioChem 6(2):375–385. doi:10.1002/cbic.200400282
156. Wieland Brown LC, Acker MG, Clardy J, Walsh CT, Fischbach MA (2009) Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. Proc Natl Acad Sci USA 106(8):2549–2553. doi:10.1073/pnas.0900008106
157. Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE, Walker C, Bowser TA, Graham IA (2012) A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science 336(6089):1704–1708. doi:10.1126/science.1220757
158. Wyatt MA, Magarvey NA (2013) Optimizing dimodular nonribosomal peptide synthetases and natural dipeptides in an *Escherichia coli* heterologous host. Biochem Cell Biol 91(4):203–208. doi:10.1139/bcb-2012-0097
159. Wyatt MA, Mok MC, Junop M, Magarvey NA (2012) Heterologous expression and structural characterisation of a pyrazinone natural product assembly line. ChemBioChem 13(16):2408–2415. doi:10.1002/cbic.201200340
160. Wyatt MA, Wang W, Roux CM, Beasley FC, Heinrichs DE, Dunman PM, Magarvey NA (2010) *Staphylococcus aureus* nonribosomal peptide secondary metabolites regulate virulence. Science 329(5989):294–296. doi:10.1126/science.1188888
161. Yin WB, Chooi YH, Smith AR, Cacho RA, Hu Y, White TC, Tang Y (2013) Discovery of cryptic polyketide metabolites from dermatophytes using heterologous expression in *Aspergillus nidulans*. ACS Synth Biol. doi:10.1021/sb400048b
162. Zabala AO, Xu W, Chooi YH, Tang Y (2012) Characterization of a silent azaphilone gene cluster from *Aspergillus niger* ATCC 1015 reveals a hydroxylation-mediated pyran-ring formation. Chem Biol 19(8):1049–1059. doi:10.1016/j.chembiol.2012.07.004
163. Zhao B, Lin X, Lei L, Lamb DC, Kelly SL, Waterman MR, Cane DE (2008) Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). J Biol Chem 283(13):8183–8189. doi:10.1074/jbc.M710421200
164. Ziemert N, Ishida K, Quillardet P, Bouchier C, Hertweck C, de Marsac NT, Dittmann E (2008) Microcyclamide biosynthesis in two strains of *Microcystis aeruginosa*: from structure to genes and vice versa. Appl Environ Microbiol 74(6):1791–1797. doi:10.1128/AEM.02392-07
165. Ziemert N, Ishida K, Weiz A, Hertweck C, Dittmann E (2010) Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides. Appl Environ Microbiol 76(11):3568–3574. doi:10.1128/AEM.02858-09
166. Zimmermann M, Fischbach MA (2010) A family of pyrazinone natural products from a conserved nonribosomal peptide synthetase in *Staphylococcus aureus*. Chem Biol 17(9):925–930. doi:10.1016/j.chembiol.2010.08.006
167. Zimmermann M, Hegemann JD, Xie X, Marahiel MA (2013) The astexin-1 lasso peptides: biosynthesis, stability, and structural studies. Chem Biol 20(4):558–569. doi:10.1016/j.chembiol.2013.03.013