BMC Bioinformatics

RESEARCH ARTICLE

Open Access

Automatic reconstruction of metabolic pathways from identified biosynthetic gene clusters

Snorre Sulheim[1,2*], Fredrik A. Fossheim[1], Alexander Wentzel[2] and Eivind Almaas[1,3]

*Correspondence:
snorre.sulheim@sintef.no
[1] Department of Biotechnology and Food Science, NTNU - Norwegian University of Science and Technology, Sem Sælands vei 8, 7034 Trondheim, Norway
Full list of author information is available at the end of the article

Abstract

Background: A wide range of bioactive compounds is produced by enzymes and enzymatic complexes encoded in biosynthetic gene clusters (BGCs). These BGCs can be identified and functionally annotated based on their DNA sequence. Candidates for further research and development may be prioritized based on properties such as their functional annotation, (dis)similarity to known BGCs, and bioactivity assays. Production of the target compound in the native strain is often not achievable, rendering heterologous expression in an optimized host strain as a promising alternative. Genome-scale metabolic models are frequently used to guide strain development, but large-scale incorporation and testing of heterologous production of complex natural products in this framework is hampered by the amount of manual work required to translate annotated BGCs to metabolic pathways. To this end, we have developed a pipeline for an automated reconstruction of BGC associated metabolic pathways responsible for the synthesis of non-ribosomal peptides and polyketides, two of the dominant classes of bioactive compounds.

Results: The developed pipeline correctly predicts 72.8% of the metabolic reactions in a detailed evaluation of 8 different BGCs comprising 228 functional domains. By introducing the reconstructed pathways into a genome-scale metabolic model we demonstrate that this level of accuracy is sufficient to make reliable in silico predictions with respect to production rate and gene knockout targets. Furthermore, we apply the pipeline to a large BGC database and reconstruct 943 metabolic pathways. We identify 17 enzymatic reactions using high-throughput assessment of potential knockout targets for increasing the production of any of the associated compounds. However, the targets only provide a relative increase of up to 6% compared to wild-type production rates.

Conclusion: With this pipeline we pave the way for an extended use of genome-scale metabolic models in strain design of heterologous expression hosts. In this context, we identified generic knockout targets for the increased production of heterologous compounds. However, as the predicted increase is minor for any of the single-reaction knockout targets, these results indicate that more sophisticated strain-engineering strategies are necessary for the development of efficient BGC expression hosts.

Keywords: Biosynthetic gene clusters, Genome-scale metabolic model, AntiSMASH, Polyketide synthases, Natural products, Heterologous expression, Non-ribosomal peptide synthetases

Background

Natural products provide an immense source of bioactive small molecules of medical and agricultural importance [1–3]. The biosynthesis of these small-molecule bioactive compounds is usually governed by genes that are clustered in physical close proximity on the genome in fungal [4] or bacterial species [5], commonly known as biosynthetic gene clusters (BGCs). The revolution in sequencing technology has enabled access to complete genome sequences for an increasing number of bacteria and fungi. Mining of these genomes has revealed a vast abundance of BGCs, many more than the number of bioactive compounds observed in vitro [6, 7], suggesting that many BGCs are not expressed or that their respective compounds are not produced at detectable amounts in laboratory conditions. The activation of these silent BGCs may lead to the discovery of many novel bio-pharmaceuticals [8].

One promising avenue towards exploration of the bioactive potential of these silent BGCs is heterologous expression in host strains that are engineered to achieve maximal production of the encoded natural products [9, 10]. With current software [11] it is possible to quickly mine a genome for BGCs and retrieve information about the class, location, and functional domains of every gene in each cluster [12]. One may further prioritize BGC candidates for heterologous expression based on this information, (dis)similarity to known BGCs, bioactivity assays and mass spectrometry profiles of produced compounds, and subsequently transfer the selected BGCs to a chosen host strain using available genetic tools [13, 14]. However, the cloning and transfer of BGCs can be time-consuming and difficult depending on the genetic tools available for the native and the heterologous host strains, as well as the size of the BGC in question [15]. Additionally, it is not clear which host strain or which genetic modifications will maximize the yield of the secondary metabolite synthesized through the metabolic pathway catalyzed by the enzymes, or enzyme complexes, encoded by the heterologously expressed BGC [16, 17].

Genome-scale metabolic models (GEMs) can predict the consequence of genetic modifications [18] and are routinely used to guide strain design for a wide range of purposes [19]. However, this approach has still not gained traction in guiding strain-engineering efforts to increase the heterologous production of complex natural compounds, despite a number of available GEMs for *Actinobacteria* [20], a phylum known for an extremely diverse secondary metabolism responsible for about two-thirds of all known antibiotics in use today [21]. Previous efforts are limited to maximization of native secondary metabolites [22–24] or precursor pools [25]. One reason for the lack of computational efforts leveraging GEMs to assess heterologous production from BGCs is the significant amount of work required to map out the associated metabolic pathway, although most of the required information is contained in the output from software used to identify and annotate BGCs, such as antiSMASH [12]. In this work, we address this hurdle by developing a pipeline that parses the output obtained from antiSMASH and constructs

the corresponding metabolic-synthesis pathway, thereby making BGCs available for constraint-based analysis and strain engineering guided by GEMs.

We have chosen to focus on non-ribosomal peptide synthetases (NRPSs) and two types of polyketide synthases (PKSs), namely type 1 PKSs and trans-AT PKSs. These BGC classes are of particular interest because of their vast abundance [26, 27] and great prospect to become novel biopharmaceuticals [28, 29]. For an exhaustive description of NRPS and PKS biosynthesis, we refer the reader to a range of excellent reviews [27, 30–33], but we provide the brief summary required as a context for the later description of the pipeline and results. Both NRPS, and type 1 and trans-AT PKS biosynthesis are performed by multidomain enzyme complexes that create a polymer from amino acid or acyl-CoA building blocks, respectively. The chain elongation is performed by well-defined modules that makes it tractable to predict the biosynthetic pathways producing the associated compounds from the annotated sequence data, but the presence of iterative modules can complicate predictions [34–36]. An active chain elongating module in an NRPS cluster requires at least three functional domains: a condensation (C) domain, an adenylation (A) domain and a peptidyl carrier (PCP) domain. The A domain activates a specific amino acid (or in some cases a carboxylic acid) and facilitates the attachment of the amino acid to the PCP domain, while the C domain catalyzes the formation of peptide bonds required to elongate the peptide. In addition to these three domains, NRPS modules can replace the C domain by a Cy domain performing condensation and heterocyclisation or additionally contain a methyltransferase (MT) and/or an epimerase (E) domain. The load module initiating biosynthesis usually lacks the C domain, while the terminating module contains either a thioesterase (TE) or a thioester reductase (TR) domain.

Similar to NRPSs, chain elongating modules of PKSs rely on three functional domains: an acyltransferase (AT) domain that recognizes a specific extender unit and attaches it to the acyl carrier (ACP) domain. The third domain, ketosynthase (KS) catalyzes the Claisen condensation required to extend the polyketide chain. A standard PKS load module contains only the AT and ACP domain, and a TE or TR domain is required for the release of the polyketide chain by the final PKS module. PKS modules can also feature the reducing domains ketoreductase (KR), dehydratase (DH) and enoylreductase (ER), and different combinations of functional domains yield a large variety of molecular transformations, in particular for the trans-AT PKSs [32]. These trans-AT PKSs not only differ from normal (*cis*) modular PKSs by having a larger module diversity and deviations from canonical rules, but they are also recognized by freestanding AT domains that perform the chain elongation [32]. The diversity of PKS and NRPS natural products is further extended by hybrid variants containing both NRPS and PKS domains and modules.

We acknowledge that experimental analyses of the final and intermediate products, as well as enzyme activity assays, are required to fully unravel the details of the metabolic pathways associated with a BGC. However, for the chosen classes of BGCs (NRPS, type 1 PKS, and trans-AT PKS), we hypothesize that the information acquired from genome mining is sufficient to make *in silico* predictions that are biologically relevant. After assembling and evaluating the accuracy of the new pipeline presented in this work, we demonstrate its value towards high-throughput assessment of BGCs by reconstructing

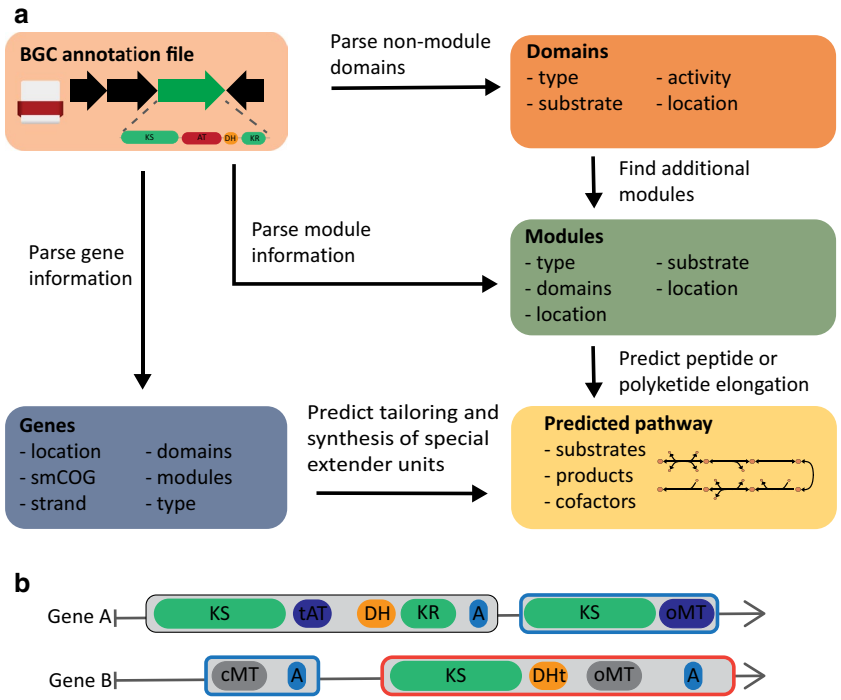

Fig. 1 Overview of the BiGMeC pipeline. **a** Schematic description of how the BGC annotation file produced by antiSMASH is parsed and used to construct the associated metabolic pathway. **b** BiGMeC extends the rule-based identification of modules from antiSMASH with bridging modules and analysis of module activity, as exemplified here with this toy BGC: The last module in gene A and the first module in gene B (marked by green edges) constitute an active bridging module that is not identified by antiSMASH [12]. The last module on gene C (red edge color) are most often found to be inactive, a feature currently incorporated into BiGMeC

the metabolic pathways for 943 of the BGCs currently in MIBiG [37]. Furthermore, we predict the optimal single reaction inactivation (by gene knockout) strain-engineering strategy for natural product synthesis based on each BGC when introduced into a genome-scale metabolic model of *Streptomyces coelicolor*, a model organism among the *Actinobacteria* and a popular heterologous BGC expression host [15, 38].

Results

We have developed the Biosynthetic Gene cluster Metabolic pathway Construction (BiGMeC) pipeline that leverages antiSMASH results to create the metabolic pathway corresponding to a PKS or NRPS biosynthetic gene cluster (Fig. 1a). The pipeline details each enzymatic reaction of the metabolic pathway, including redox cofactors and energy demand. The results are stored in a format that is easily introduced into a GEM using popular tools for constraint-based reconstruction and analysis, such as cobrapy [39] or COBRA Toolbox [40].

The hallmarks of PKS- and NRPS-genes are adjacent functional domains that in total make up one or several modules that initiate, extend or cleave off the polyketide or peptide product, respectively [30, 32, 33]. The output from antiSMASH comprises information about these modules and their functional domains, and occasionally also the specific extender unit or chemical transformation associated with each functional domain [12]. The BiGMeC pipeline not only parses this information, but uses well-reasoned heuristics

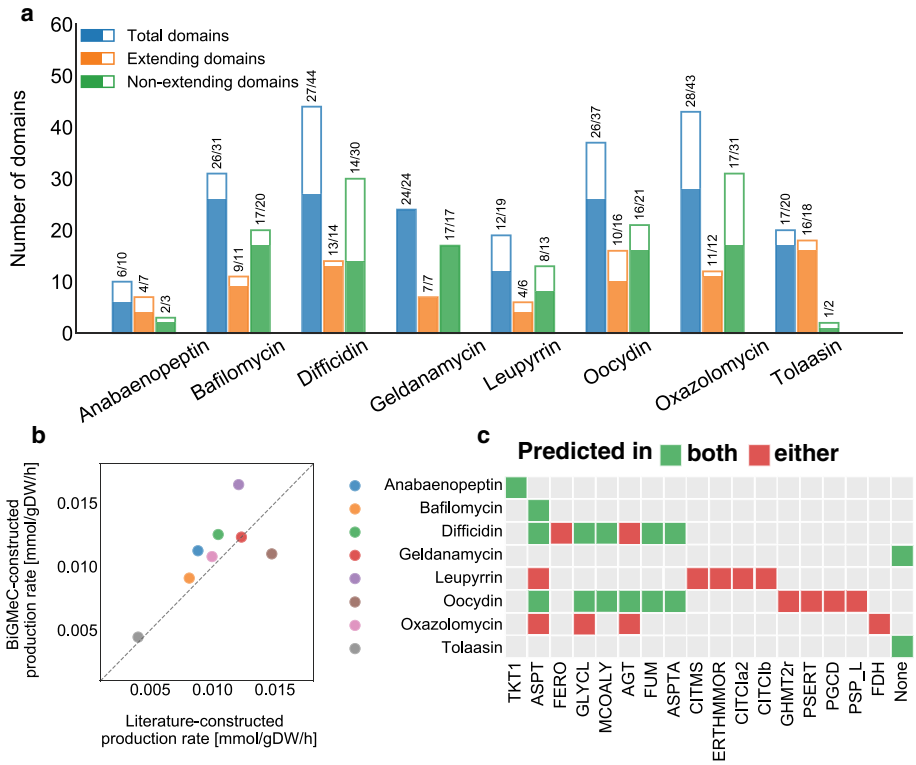

Fig. 2 Analysis of BiGMeC prediction accuracy for eight selected BGCs. **a** Barplot showing the number of correct domains when comparing BiGMeC-constructed pathways with pathways as they are detailed in the literature (Additional file 2). The filled part of each bar, as well as the ratio printed above, shows the number of correct domains for each BGC. Extending domains comprise the domains that append an extender unit to the polyketide or peptide backbone, while the non-extending domains cover all other domains. **b** Predicted maximum production rate when introduced into a *S. coelicolor* GEM. The x- and y-axis represent the maximal production rate using the metabolic pathway created based on literature or reconstructed with BiGMeC, respectively. **c** This panel shows a comparison of the predicted reaction-knockout targets (x-axis) when using a metabolic pathway created based on literature or with BiGMeC. Similar predictions are shown as green tiles, while incorrect predictions (predicted in either but not both of the two cases) are shown as red tiles. The names of the model reaction IDs are: TKT1: transketolase; ASPT: aspartate ammonia-lyase; FERO: ferroxidase; GLYCL: glycine cleavage system; MCOALY: malyl-CoA lyase; AGT: alanine-glyoxylate aminotransferase; FUM: fumarase; ASPTA: aspartate transaminase; CITMS: (R)-citramalate synthase; ERTHMMOR: 3-isopropylmalate dehydrogenase; CITCIa2: (R)-2-Methylmalate hydro-lyase; CITCIb: 2-methylmaleate hydratase; GHMT2r: glycine hydroxymethyltransferase; PSERT: phosphoserine transaminase; PGCD: phosphoglycerate dehydrogenase; PSP_L: phosphoserine phosphatase; FDH: formate dehydrogenase

to handle deviations from canonical rules and cases where information is missing (see Materials and Methods). Improvements in determining module function includes identification of bridging modules in trans-AT PKSs and non-extending modules due to the presence of oMT domains [32] (Fig. 1b).

We first assessed the accuracy of the BiGMeC pipeline by comparing its predictions with experimentally characterized and manually curated metabolic pathways. To this end, we compared the substrates, cofactors, and reaction products of each step of the metabolic pathway associated with eight well-characterized BGCs (Fig. 2a, Additional file 2). These BGCs cover a range of BGC classes, including type 1 PKS, trans-AT PKS, NRPS and hybrids, and we believe they provide a test set that is sufficiently diverse to probe the pipeline for its strengths and weaknesses. Overall, BiGMeC appends the

correct metabolic reaction for 72.8% (166/228) of the functional domains in all eight BGCs. Of these functional domains, BiGMeC chooses the correct extender unit for 81.3% (74/91) of the domains extending the peptide or polyketide. For all other domains, including chain initiation, reductive domains, methyltransferases and final tailoring reactions, the accuracy is 67.2% (92/137).

A large number of the incorrect predictions derive from wrong assignments of inactive KR domains by antiSMASH [12]. Across the eight closely inspected BGCs, KR domains are almost always active, but on several occasions antiSMASH predicts that these domains are inactive. The incorrect predictions of KR domain activity are to a large extent associated with adjacent MT domains. Furthermore, this leads to incorrect assignment of the activity of succeeding DH and ER domains because they act on the functional moiety produced by the preceding domain. For the prediction of extender units, most incorrect assignments derive from missing recognition of non-elongating modules caused by inactive KS domains devoid of a conserved histidine residue required for carboxylative condensation [32]. More specifically, only 10 of 16 KS domains are active in the oocydin BGC [32, 41]. Another significant source of incorrect domains is the anabaenopeptin cluster that has two consecutive genes, each having two modules that initiate biosynthesis and perform first chain elongation, respectively, yielding two slightly different variants of the final compound. The BiGMeC pipeline treats these two genes as consecutive steps of the same pathway, and therefore, predicts too many chain elongations in the biosynthesis.

To investigate how much the errors in the constructed metabolic pathways affect model predictions, we introduced both the literature-based and the BiGMeC pathway reconstructions into the consensus GEM of *S. coelicolor* (Sco-GEM) [16] and compared the maximal production rate of the final compound (Fig. 2b). In general, we observe quite similar rates for the eight BGCs (Pearson $\rho = 0.75$, $P = 0.03$), suggesting that the incorrect domains only have a minor impact on the predicted production rates. The offset in the production of leupyrrin likely comes from an incorrect starter unit while the offset in oocydin production is caused by a fairly large error in the predicted number of malonyl-CoA extender units (10 vs. 16).

The anticipated use of the developed pipeline towards strain engineering of expression hosts underscores the need to elucidate if model-based strain designs using BiGMeC-constructed pathways deviate from results using pathways reconstructed according to literature. To this end, we predicted optimal single-reaction knockout mutants that should increase the production rate of the associated product (Fig. 2c). Note that, a reaction knockout is the practical implication of disrupting one or more of the genes encoding the enzyme catalyzing the corresponding reaction. For 6 out of 8 BGCs there is a good overlap between pairwise pathway reconstructions. This includes the cases of tolaasin and geldanamycin, where no knockout target is identified with either of the two pathway reconstructions.

To demonstrate the power of BiGMeC in high-throughput assessment of BGCs, we employed the pipeline on 1883 of the 1923 BGCs in the MIBiG database (version 2.0) [37]. For 40 of the 1923 BGCs, we could not obtain the antiSMASH output file because the link from MIBiG was broken. The 943 (50.1%) metabolic pathways that were successfully reconstructed with BiGMeC cover both fungi and a range of different bacteria

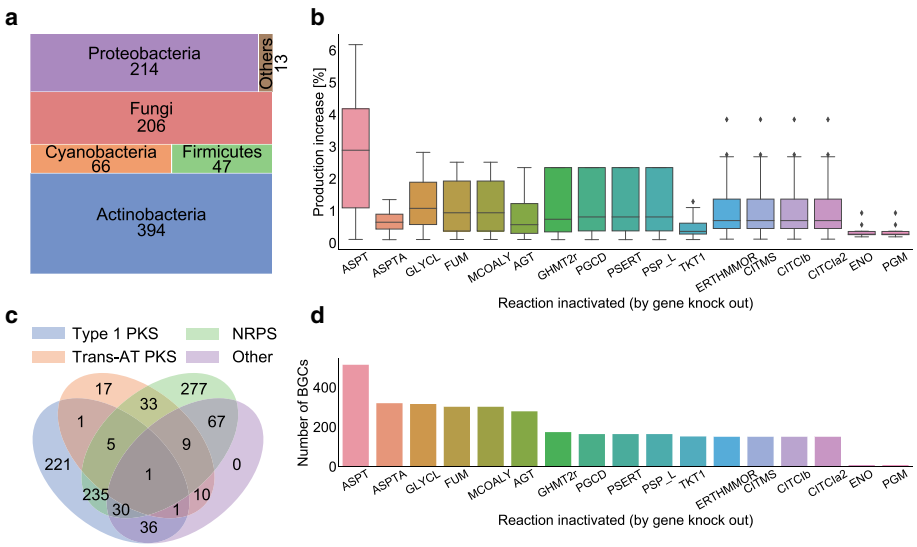

Fig. 3 Automatic reconstruction and analysis of 943 BGCs from MIBiG. These BGCs cover **a** a range of different organisms and **b** a wide variety of hybrid BGCs. **c** Box plot showing the increase in production for the 17 different reaction knockouts that increase the production of one or more of the analysed BGCs. **d** Bar chart showing the number of BGCs where the knockout of each reaction is predicted to increase production of the target secondary metabolite. The names of the model reactions used in panel C and D: ASPT: aspartate ammonia-lyase; ASPTA: aspartate transaminase; GLYCL: glycine cleavage system; FUM: fumarase; MCOALY: malyl-CoA lyase; AGT: alanine-glyoxylate aminotransferase; GHMT2r: glycine hydroxymethyltransferase; PGCD: phosphoglycerate dehydrogenase; PSERT: phosphoserine transaminase; PSP_L: phosphoserine phosphatase; TKT1: transketolase; ERTHMMOR: 3-isopropylmalate dehydrogenase; CITMS: (R)-citramalate synthase; CITCIb: 2-methylmaleate hydratase; CITCIa2: (R)-2-Methylmalate hydro-lyase; ENO: enolase: PGM: phosphoglycerate mutase

(Fig. 3a). Most clusters are either type 1 PKS, NRPS, or hybrids of these two, and only 77 of the BGCs share similarity with trans-AT PKS (Fig. 3b). The 940 remaining BGCs were not analyzed either because the BGC class was not covered by BiGMeC (such as RiPPs, terpenes, Type 2 and Type 3 PKSs) or because functional modules and domains were lacking in the results from antiSMASH.

We introduced each of the 943 reconstructed pathways into Sco-GEM [16], and predicted single-reaction knockout strategies improving the production of the final pathway product. Surprisingly, only 17 different reactions were suggested as a knockout target in one or more of the 943 *in silico* heterologous expression experiments (Fig. 3c, d). Of these 17 reactions, aspartate transaminase is predicted to provide on average the largest increase in production (Fig. 3c) and is also the most frequently suggested candidate (Fig. 3d). However, the predicted production increase is minor for all of the 17 suggested reactions, including aspartate transaminase, with a maximum increase of 6% relative to the wild-type production rate.

Discussion

To make novel natural product pathways encoded by BGCs accessible to the constraint-based reconstruction and analysis framework, we have developed a pipeline that creates a draft reconstruction of the metabolic pathway encoded by a BGC. This pipeline

outlines the correct metabolic reaction for 72.8% of the functional domains in our test set comprised of 8 experimentally characterized BGC-encoded biosynthetic pathways. One may question whether this accuracy extends to uncharacterized BGCs. In principle, as the pathway reconstruction is solely based on genome mining results from antiSMASH, there should not be a significant difference in accuracy between well-characterized and uncharacterized BGCs. However, as antiSMASH relies on annotation rules learnt from well-characterized BGCs [42, 43], one may anticipate that uncharacterized BGCs that deviate from known canonical rules are less accurately annotated by antiSMASH, and therefore less accurately reconstructed by BiGMeC.

By applying the BiGMeC pipeline to 943 BGCs covering NRPSs, PKSs and NRPS-PKSs hybrids from a wide range of organisms we have demonstrated how the pipeline enables high-throughput assessment of potential candidates for heterologous expression. In an assessment of 943 BGCs, we explored general single-gene knockout strategies for increased heterologous production, and although we identify a set of 17 general targets, none provides a drastic increase in production. This result suggests that multiple knockouts, over-expression of genes, or strategies that perturb regulatory mechanisms are necessary to reroute a large amount of precursors from growth towards secondary metabolism, at least in the organism *S. coelicolor*.

Although the accuracy of the BiGMeC pipeline is sufficient to make biologically relevant pathway reconstructions, this work has also revealed aspects where there is room for further improvement. Incorrect assignment of KS and KR domains as active or inactive is a large source of error in PKS metabolic pathways, and incorporation of the recently developed transATor algorithm would provide an improvement in this context [44]. Synthesis of rare precursors and tailoring of the polyketide or peptide succeeding the release from the multidomain enzyme complex are two other features with opportunity for improvement. Although the genes encoding enzymes responsible for the synthesis of rare precursors or for the post-release tailoring steps usually are contained in the BGC, neither their exact function nor their functional order can be accurately predicted. Therefore, the current pipeline relies in certain aspects on assumptions and heuristics that apply in general, but with several exceptions. However, with a continuous improvement in algorithms for annotation and identification of BGCs [12, 44, 45] and increased experimental characterization [37], current generalisations can develop into more accurate pathway reconstructions that encompass a larger range of deviations from canonical rules. Furthermore, as the knowledgebase and algorithms for annotation of iterative PKSs and ribosomally synthesised and post-translationally modified peptides improves [46, 47], these types of BGCs represent obvious targets for further development. Other possible targets include terpenes, alkaloids and glycosides, frequently encoded in plant and fungal genomes [48–50], or polysaccharides which are of large value in dairy industry [51] and medical applications [52], and the most abundant class of prokaryotic BGCs [5]. Nevertheless, accurate pathway reconstruction for these classes of BGCs will require accurate descriptions of the biosynthetic rules encoded in the gene clusters. In this context, tailoring reactions and post-translational modifications represent particular challenges. Further improvement should also aim to accept the output from other annotation software, such as PRISM [53].

Conclusion

The BiGMeC pipeline is, to our knowledge, the first tool for automatic metabolic pathway reconstruction specifically targeting PKS and NRPS BGCs. Although the reconstructed pathways are not able to capture the entire diversity seen in the biosynthesis of NRPSs and PKSs [30, 32], the predicted production rates and reaction knockout targets are comparable to predictions provided using manually reconstructed pathways. Furthermore, the pipeline can aid model reconstruction efforts, both as a decent starting point for further manual curation and as a complement to standard model-reconstruction pipelines [54]. This is in particular relevant for organisms with a rich secondary metabolism, such as the *Actinobacteria* which are of utmost interest in drug discovery. We anticipate that the pipeline presented here can increase the use of GEMs in this context, e.g. to screen different combinations of BGCs and expression hosts or, as shown in this work, to explore strain-engineering opportunities. The pipeline is developed in an open source environment on GitHub and we encourage interested readers to engage in future development through pull request or by raising issues. We also encourage developers of genome mining tools and databases to converge towards standardized and consistent file formats, such as the Minimum Information about a Biosynthetic Gene Cluster (MIBiG) initiative [37]. This will ease the development and maintenance of downstream pipelines such as BiGMeC, and promote integration of data from different genome mining tools. This is intended as a reminder rather than a criticism of existing software.

Materials and methods

Software implementation

We developed BiGMeC to translate information about PKS and NRPS BGCs to detailed outlines of the metabolic reactions governing the production of the associated secondary metabolites. The BiGMeC software and all other associated scripts are implemented in Python 3 and publicly available at https://github.com/AlmaasLab/BiGMeC. BiGMeC runs from a command-line interface and takes an annotated NRPS or PKS BGC in the format of a region-specific GenBank file as produced by antiSMASH 5.1 [12]. It leverages the included gene, domain, and module information to make a description of the enzymatic reactions encoded by the BGC, including substrate and co-factor usage (Fig. 1a). BiGMeC uses a reference model as a library of metabolites and reactions, and in the current work, we have used Sco-GEM version 1.2.1, the consensus *S. coelicolor* GEM [16]. This model was obtained from https://github.com/SysBioChalmers/Sco-GEM.

The BiGMeC pipeline first parses information about the location and annotation of the genes and modules as annotated by antiSMASH from the GenBank file (Fig. 1). If available, the gene information includes strand, secondary metabolism Clusters of Orthologous Groups (smCOG) annotation [55], type of gene, extender unit, annotated functional domains and if the gene is a core gene or not. The core genes synthesize the core structure of the PKS or NRPS molecule. The module information contains details about the type of module and its functional domains. Then, the pipeline assesses the presence and order of domains not included in a module, e.g. special load or bridging modules (in trans-AT PKS, Fig. 1b) [32], and combines these domains into functional modules when possible. The peptide or polyketide backbone is subsequently constructed

Table 1 List of domains and associated reactions as implemented in BiGMeC

Abbrv.	Name	Note	Reaction
A	Adenylation	Activates and attaches AA to PCP	ATP + AA + PCP→AA-PCP + AMP + P_{ii}
ACP	Acyl carrier protein	Facilitates transport in PKSs	
AT	Acyltransferase	Loads extender unit onto ACP	Acyl-CoA + ACP → Acyl-ACP + CoA
C	Condensation	Elongates the peptide by condensation	AA-PCP + X_n → X_{n+1} + H_2O + ACP
CAL	Coenzyme A ligase	Catalyzes the incorporation of different starter units, e.g. fatty acids, AHBA, and shikimic acid [32, 56, 57]	Y + PCP → Y-PCP + H_2O
cMT	Carbon methyltransferase	Methylates peptide/polyketide	SAM + X_n → X_n + SAH
Cy	Heterocylization	Elongates the peptide by condensation and cyclization	AA-PCP + X_n → X_{n+1} + H_2O + ACP
DH	Dehydratase	Forms double bound by removal of H_2O	X_n → X_n + H_2O
E	Epimerase	Stereochemical inversion	X_n → X_n
ECH	Enoyl-CoA hydratase/isomerase	Not able to discriminate, so BiGMeC assumes isomerase	X_n → X_n
ER	Enoyl reductase	Reduces double bound formed by the DH domain to a methylene group	NADPH + H^+ + X_n → X_n + $NADP^+$
FkbH	FkbH-like domain	Domain in an alternative loading module. Dephosphorylates 1,3-bpg [58]	1,3-bpg + ACP → D-lactate-ACP + 2 P_i
GNAT	GCN5-related N-acetyl transferase	Alt. load module that decarboxylates malonyl-CoA and adds acetyl group to ACP [59]	Malonyl-CoA + ACP → Acetyl-ACP + CoA + CO_2
KR	Keto reductase	Reduces carbonyl group to hydroxyl group	NADPH + H^+ + X_n → X_n + $NADP^+$
KS	Keto synthase	Appends extender unit to polyketide	Acyl-ACP + X_n → X_{n+1} + CO_2 + ACP
nMT	Nitrogen methyltransferase	Methylates peptide/polyketide	SAM + X_n → X_n + SAH
oMT	Oxygen methyltransferase	Methylates peptide/polyketide	SAM + X_n → X_n + SAH
PCP	Peptidyl carrier domain	Facilitates transport in NRPSs	
TD	Thioester reductase	Releases product from ACP/PCP	NADPH + H^+ + X_n → detached product + $NADP^+$
TE	Thioesterase	Releases product from ACP/PCP	H_2O + X_n → detached product

The peptide or polyketide backbone is referred to as X_n, and in reactions that extend the backbone we refer to the elongated backbone as X_{n+1}

AA, generic amino acid; 1,3-bpg, 1,3 biphosphoglycerate; CoA, Coenzyme A; P_{ii}, diphosphate; SAH, S-Adenosyl-L-homocysteine; SAM, S-Adenosyl methionine; Y, generic starter unit

based on the order of the identified domains and the function of each domain within each module. Although NRPS and type 1 PKS modules can be iterative, we here assume that the selected BGCs are modular such that each module only performs one chain elongation. The reactions associated with the functional domains are listed in Table 1. Domains in the BGC that are not contained in a module are assumed to not affect the backbone structure. If a terminating domain (thioesterase or thioester reductase) domain is encountered, no further chain elongations are carried out. The activity of reducing domains (DH, ER, KR) are based on the annotation of the KR domain from antiSMASH. Tailoring reactions post PKS synthesis are predicted from the smCOG annotations of each gene. The currently implemented tailoring reactions relate to the

smCOGs 1256, 1084, 1002, 1109 and 1062 and includes glycosylation, glycosyltransferase and incorporation of 2-Amino-3-hydroxycyclopent-2-enone (Additional file 1).

Rare extender units appear in both PKS and NRPS biosynthesis. The synthesis of rare extender units is usually carried out by genes in the BGC [60], and we therefore include the synthesis of the most common rare extender units (not in the reference library) when necessary. This includes hydroxyphenylglycine, beta-hydroxytyrosine, 2-aminobutyric acid, pipecolic acid, dihydroxyphenylglycine and 3-amino-5-hydroxybenzoate [56]. Synthesis of the rare extender unit methoxymalonyl-ACP [60] is based on the presence of genes with specific smCOG annotations (Additional file 1). For the remaining rare extender units, or in the case of missing information or nonspecific antiSMASH annotation, we use a conservative approach where a generic amino acid is used as the extender unit in NRPS modules and malonyl-CoA is used in PKS modules. In the case of using a generic amino acid as the extender unit, we add a set of pseudo-reactions that can convert every proteogenic amino acid into this generic molecule to ensure that the biosynthetic pathway is functional.

The pipeline also handles a number of deviations from the canonical rules, for example the deactivation of the KS domain often seen in modules containing O-methyltransferases [32]. Furthermore, it is found that the presence of a C domain in the initiating NRPS module acylates the initial amino acid [31, 61]. Both in tolaasin [62] and surfactin, currently the best studied example of this type of NRPS initiation, the acylating agent is a CoA-activated β-hydroxy fatty acid [61, 63]. It is likely that the C-domain has a strong selectivity for a specific acylating agent, but since this specificity is not identified by antiSMASH we use a generic fatty acid molecule. A third example of exceptions that are handled by BiGMeC is bridging modules in trans-AT PKSs where the KS domain is encoded in the first gene and the DH and ACP domains follow immediately on the second gene. These modules are called dehydratase docking domains (DHD) and are usually not active [32].

Evaluation of the BiGMeC pipeline

To evaluate how well biosynthetic pathways can be constructed solely based on antiSMASH data we compared BiGMeC-constructed pathways with literature-based reconstructions for 8 different BGCs, covering different species and classes of BGCs (Additional file 2). The 8 BGCs were (MIBiG ID in parenthesis): bafilomycin from *Streptomyces lohi* [64–66] (BGC0000028), geldanamycin from *Streptomyces hygroscopicus* [67–69] (BGC0000066), difficidin from *Bacillus velezensis* FZB42 [70, 71] (BGC0000176), oocydin from *Serratia plymuthica* [32, 41] (BGC0001032), oxazolomycin from *Streptomyces albus* [71, 72] (BGC0001106), leupyrrin from *Sorangium cellulosum* [73] (BGC0000380), anabaenopeptin from *Anabaena* sp. 90 [74] (BGC0000302) and tolaasin from *Pseudomonas costantinii* [62] (BGC0000447). For each domain in each of the 8 different BGCs we compared the BiGMeC-constructed reaction with the *real* reaction, i.e. the associated reaction as described in the literature. When clearly defined in the literature, tailoring reactions were included, but we focused on the synthesis of the core peptide/polyketide. The very complex tailoring of leupyrrin [73] was not included.

An initial evaluation was performed by counting the number of correct domains (Fig. 2a). The total number of domains include all domains either predicted by BiGMeC or described in the literature, and the correct predictions include both true positives and true negatives. Next, we incorporated the BiGMeC and literature-based

pathway reconstructions into Sco-GEM and predicted the maximum production rate of the secondary metabolite produced by each pathway (Fig. 2b). To do so, we performed Flux Balance Analysis (FBA) [75, 76] in cobrapy [39] with the final reaction of the BGC encoded pathway as objective and with growth limited to minimum 90% of the maximum value. The growth and production were simulated in a growth medium with glucose and ammonium as the sole carbon and nitrogen sources, respectively, and with a maximum glucose uptake rate of 0.8 mmol gDW^{-1} h^{-1}. We did not constrain the uptake of ammonium, sulphate, phosphate, oxygen and metal ions. Finally, using both the BiGMeC and literature-based pathway reconstructions, we predicted reaction inactivation targets (by gene knockout) that would increase the production of the associated compound, with a maximum growth rate reduction of 50% (Fig. 2c). We limited the set of possible reaction targets to non-essential gene-annotated reactions. The search for optimal knockouts was carried out in a brute-force manner: we conducted an iterative knockout of each reaction (within the predefined set of possible reactions) and, first used FBA to predict the maximum growth of the mutant phenotype, and secondly predict the maximum production rate at 99.9% of the knockout-mutant's maximum growth rate. All knockouts that resulted in more than 0.1% increase in production rate compared to the wild-type were considered knockout candidates.

Large-scale reconstruction of BGC pathways

To demonstrate the value and efficiency enabled by BiGMeC we applied this pipeline to all relevant BGCs from the MIBiG database [37]. To get the antiSMASH-generated output for all BGCs in MIBiG we automatically downloaded all GenBank-files with a url on the form: https://mibig.secondarymetabolites.org/repository/BGC0000001/generated/BGC0000001.1.region001.gbk, with the MIBiG ID ranging from BGC0000001 to BGC0002057. The MIBiG database currently reports on a total of 1923 BGCs but due to different reasons (e.g. missing entries) we could only obtain the antiSMASH result for 1883 of the entries. For all BGCs at least annotated to either type 1 PKS, trans-AT PKS or NRPS we used the BiGMeC pipeline to reconstruct the corresponding metabolic pathway. We predicted optimal knockout strategies for each of successfully constructed pathway using the same procedure as described for the 8 BGCs used to evaluate the BiGMeC pipeline.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-021-03985-0.

Additional file 1. Details on tailoring reactions and synthesis of the rare extender unit methoxymalonyl-ACP, as well as a description of the analysis used to develop the heuristics that indicate the presence of these reactions from smCOG annotations.

Additional file 2. Detailed comparison of 8 BGCs for evaluation the accuracy of the BiGMeC pipeline.

Acknowledgements
Not applicable.

Authors' contributions
Conceptualization, SS, EA; Methodology and Software, SS, FF; Validation and Formal Analysis, SS, FF; Data curation SS, FF; Writing: Original Draft, SS, FF; Reviewing and editing, SS, EA, FF, AW; Visualization SS; Supervision SS, EA; Project Administration, AW, EA; Funding Acquisition, AW, EA. All authors read and approved the final manuscript.

Funding
This research was conducted within the project INBioPharm of the Center for Digital Life Norway (Research Council of Norway grant #248885), with additional support of SINTEF internal funding.

Availability of data and materials
The BiGMeC pipeline and the data analysed/generated during the current study is available at https://github.com/AlmaasLab/BiGMeC. We have also deposited the latest version of the repository to Zenodo (https://doi.org/10.5281/zenodo.4434667) to ensure persistent access.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Biotechnology and Food Science, NTNU - Norwegian University of Science and Technology, Sem Sælands vei 8, 7034 Trondheim, Norway. [2] Department of Biotechnology and Nanomedicine, SINTEF Industry, Richard Birkelands vei 3, 7034 Trondheim, Norway. [3] K.G. Jebsen Center for Genetic Epidemiology, NTNU - Norwegian University of Science and Technology, Håkon Jarls gate 11, 7030 Trondheim, Norway.

Received: 26 November 2020 Accepted: 18 January 2021
Published online: 23 February 2021

References

1. Clardy J, Fischbach MA, Walsh CT. New antibiotics from bacterial natural products. Nat Biotechnol. 2006;24(12):1541–50.
2. Demain AL, Sanchez S. Microbial drug discovery: 80 years of progress. J Antibiot. 2009;62(1):5–16.
3. Cantrell CL, Dayan FE, Duke SO. Natural products as sources for new pesticides. J Nat Prod. 2012;75(6):1231–42.
4. Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH. Biosynthetic gene clusters and the evolution of fungal chemodiversity. Nat Prod Rep. 2020;37:868–78.
5. Cimermancic P, Medema MH, Claesen J, Kurita K, Brown LCW, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014;158(2):412–21.
6. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson N, James KD, Harris DE, Quail MA, Kieser H, Harper D, et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor a3 (2). Nature. 2002;417(6885):141–7.
7. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Ōmura S. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nat Biotechnol. 2003;21(5):526–31.
8. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov. 2015;14(2):111–29.
9. Xu M, Wright GD. Heterologous expression-facilitated natural products' discovery in actinomycetes. J Ind Microbiol Biotechnol. 2019;46(3–4):415–31.
10. Myronovskyi M, Luzhetskyy A. Heterologous production of small molecules in the optimized Streptomyces hosts. Nat Prod Rep. 2019;36(9):1281–94.
11. Kim HU, Blin K, Lee SY, Weber T. Recent development of computational resources for new antibiotics discovery. Curr Opin Microbiol. 2017;39:113–20.
12. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. Antismash 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47(W1):81–7.
13. Huo L, Hug JJ, Fu C, Bian X, Zhang Y, Müller R. Heterologous expression of bacterial natural product biosynthetic pathways. Nat Prod Rep. 2019;36(10):1412–36.
14. Sekurova ON, Schneider O, Zotchev SB. Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. Microb biotechnol. 2019;12(5):828–44.
15. Nah H-J, Pyeon H-R, Kang S-H, Choi S-S, Kim E-S. Cloning and heterologous expression of a large-sized natural product biosynthetic gene cluster in Streptomyces species. Front Microbiol. 2017;8:394.
16. Sulheim S, Kumelj T, van Dissel D, Salehzadeh-Yazdi A, Du C, van Wezel GP, Nieselt K, Almaas E, Wentzel A, Kerkhoven EJ. Enzyme-constrained models and omics analysis of *Streptomyces coelicolor* reveal metabolic changes that enhance heterologous production. iScience. 2020;23(9):101525.
17. Ke J, Yoshikuni Y. Multi-chassis engineering for heterologous production of microbial natural products. Curr Opin Biotechnol. 2020;62:88–97.

18. Famili I, Förster J, Nielsen J, Palsson BO. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci. 2003;100(23):13134–9.
19. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. Genome Biol. 2019;20(1):121.
20. Mohite OS, Weber T, Kim HU, Lee SY. Genome-scale metabolic reconstruction of actinomycetes for antibiotics production. Biotechnol J. 2019;14(1):1800377.
21. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk H-P, Clément C, Ouhdouch Y, van Wezel GP. Taxonomy, physiology, and natural products of actinobacteria. Microbiol Mol Biol Rev. 2016;80(1):1–43.
22. Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, Nielsen J, Kerkhoven EJ. Raven 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. PLoS Comput Biol. 2018;14(10):1006541.
23. Borodina I, Siebring J, Zhang J, Smith CP, van Keulen G, Dijkhuizen L, Nielsen J. Antibiotic overproduction in Streptomyces coelicolor a3 (2) mediated by phosphofructokinase deletion. J Biol Chem. 2008;283(37):25186–99.
24. Huang D, Li S, Xia M, Wen J, Jia X. Genome-scale metabolic network guided engineering of Streptomyces tsukubaensis for fk506 production improvement. Microb Cell Factories. 2013;12(1):1–18.
25. Kumelj T, Sulheim S, Wentzel A, Almaas E. Predicting strain engineering strategies using iks1317: a genome-scale metabolic model of Streptomyces coelicolor. Biotechnol J. 2019;14(4):1800180.
26. Doroghazi JR, Metcalf WW. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC Genom. 2013;14(1):611.
27. Masschelein J, Jenner M, Challis GL. Antibiotics from gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights. Nat Prod Rep. 2017;34(7):712–83.
28. Bozhüyük KA, Micklefield J, Wilkinson B. Engineering enzymatic assembly lines to produce new antibiotics. Curr Opin Microbiol. 2019;51:88–96.
29. Cane DE, Walsh CT, Khosla C. Harnessing the biosynthetic code: combinations, permutations, and mutations. Science. 1998;282(5386):63–8.
30. Challis GL, Naismith JH. Structural aspects of non-ribosomal peptide biosynthesis. Curr Opin Struct Biol. 2004;14(6):748–56.
31. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev. 2006;106(8):3468–96.
32. Helfrich EJ, Piel J. Biosynthesis of polyketides by trans-at polyketide synthases. Nat Prod Rep. 2016;33(2):231–316.
33. Keatinge-Clay AT. The structures of type i polyketide synthases. Nat Prod Rep. 2012;29(10):1050–73.
34. Mootz HD, Schwarzer D, Marahiel MA. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. ChemBioChem. 2002;3(6):490–504.
35. Fisch KM. Biosynthesis of natural products by microbial iterative hybrid pks-nrps. RSC Adv. 2013;3(40):18228–47.
36. Herbst DA, Townsend CA, Maier T. The architectures of iterative type i pks and fas. Nat Prod Rep. 2018;35(10):1046–69.
37. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJ, Van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, et al. Mibig 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 2020;48(D1):454–8.
38. Zhang JJ, Tang X, Moore BS. Genetic platforms for heterologous expression of microbial natural products. Nat Prod Rep. 2019;36(9):1313–32 (Publisher: The Royal Society of Chemistry).
39. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. Cobrapy: constraints-based reconstruction and analysis for python. BMC Syst Biol. 2013;7(1):74.
40. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, et al. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. Nat Protoc. 2019;14(3):639–702.
41. Matilla MA, Stöckmann H, Leeper FJ, Salmond GP. Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. J Biol Chem. 2012;287(46):39125–38.
42. Blin K, Kim HU, Medema MH, Weber T. Recent development of antismash and other computational approaches to mine secondary metabolite biosynthetic gene clusters. Brief Bioinform. 2019;20(4):1103–13.
43. Medema MH, Fischbach MA. Computational approaches to natural product discovery. Nat Chem Biol. 2015;11(9):639.
44. Helfrich EJ, Ueoka R, Dolev A, Rust M, Meoded RA, Bhushan A, Califano G, Costa R, Gugger M, Steinbeck C, et al. Automated structure prediction of trans-acyltransferase polyketide synthase products. Nat Chem Biol. 2019;15(8):813–21.
45. Kjærbølling I, Mortensen UH, Vesth T, Andersen MR. Strategies to establish the link between biosynthetic gene clusters and secondary metabolites. Fungal Genet Biol. 2019;130:107–21.
46. Wang B, Guo F, Huang C, Zhao H. Unraveling the iterative type i polyketide synthases hidden in streptomyces. Proc Natl Acad Sci. 2020;117(15):8449–54.
47. Kloosterman AM, Cimermancic P, Elsayed SS, Du C, Hadjithomas M, Donia MS, Fischbach MA, van Wezel GP, Medema MH. Expansion of RIPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lantibiotics. PLoS Biol. 2020;18(12):3001026.
48. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. Plantismash: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. Nucleic Acids Res. 2017;45(W1):55–63.
49. Li YF, Tsai KJ, Harvey CJ, Li JJ, Ary BE, Berlew EE, Boehman BL, Findley DM, Friant AG, Gardner CA, et al. Comprehensive curation and analysis of fungal biosynthetic gene clusters of published natural products. Fungal Genet Biol. 2016;89:18–28.
50. Nützmann H-W, Huang A, Osbourn A. Plant metabolic clusters-from genetics to genomics. New Phytologist. 2016;211(3):771–89.
51. Duboc P, Mollet B. Applications of exopolysaccharides in the dairy industry. Int Dairy J. 2001;11(9):759–68.

52. Moscovici M. Present and future medical applications of microbial exopolysaccharides. Front Microbiol. 2015;6:1012.
53. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. Prism 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res. 2017;45(W1):49–54.
54. Mendoza SN, Olivier BG, Molenaar D, Teusink B. A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biol. 2019;20(1):1–20.
55. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. Antismash: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39(suppl–2):339–46.
56. Floss HG, Yu T-W, Arakawa K. The biosynthesis of 3-amino-5-hydroxybenzoic acid (ahba), the precursor of mc 7 n units in ansamycin and mitomycin antibiotics: a review. J Antibiot. 2011;64(1):35–44.
57. Fritzler JM, Zhu G. Functional characterization of the acyl-[acyl carrier protein] ligase in the cryptosporidium parvum giant polyketide synthase. Int J Parasitol. 2007;37(3–4):307–16.
58. Zhang F, He H-Y, Tang M-C, Tang Y-M, Zhou Q, Tang G-L. Cloning and elucidation of the fr901464 gene cluster revealing a complex acyltransferase-less polyketide synthase using glycerate as starter units. J Am Chem Soc. 2011;133(8):2452–62.
59. Gu L, Geders TW, Wang B, Gerwick WH, Håkansson K, Smith JL, Sherman DH. Gnat-like strategy for polyketide chain initiation. Science. 2007;318(5852):970–4.
60. Chan YA, Podevels AM, Kevany BM, Thomas MG. Biosynthesis of polyketide synthase extender units. Nat Prod Rep. 2009;26(1):90–114.
61. Kraas FI, Helmetag V, Wittmann M, Strieker M, Marahiel MA. Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation. Chem Biol. 2010;17(8):872–80.
62. Scherlach K, Lackner G, Graupner K, Pidot S, Bretschneider T, Hertweck C. Biosynthesis and mass spectrometric imaging of tolaasin, the virulence factor of brown blotch mushroom disease. ChemBioChem. 2013;14(18):2439–43.
63. Steller S, Sokoll A, Wilde C, Bernhard F, Franke P, Vater J. Initiation of surfactin biosynthesis and the role of the srfd-thioesterase protein. Biochemistry. 2004;43(35):11331–43.
64. Zhang W, Fortman JL, Carlson JC, Yan J, Liu Y, Bai F, Guan W, Jia J, Matainaho T, Sherman DH, et al. Characterization of the bafilomycin biosynthetic gene cluster from streptomyces lohii. Chembiochem Eur J Chem Biol. 2013;14(3):301.
65. Nara A, Hashimoto T, Komatsu M, Nishiyama M, Kuzuyama T, Ikeda H. Characterization of bafilomycin biosynthesis in kitasatospora setae km-6054 and comparative analysis of gene clusters in actinomycetales microorganisms. J Antibiot. 2017;70(5):616–24.
66. Li Z, Du L, Zhang W, Zhang X, Jiang Y, Liu K, Men P, Xu H, Fortman JL, Sherman DH, et al. Complete elucidation of the late steps of bafilomycin biosynthesis in streptomyces lohii. J Biol Chem. 2017;292(17):7095–104.
67. Patel K, Piagentini M, Rascher A, Tian Z-Q, Buchanan GO, Regentin R, Hu Z, Hutchinson C, McDaniel R. Engineered biosynthesis of geldanamycin analogs for hsp90 inhibition. Chem Biol. 2004;11(12):1625–33.
68. Rascher A, Hu Z, Viswanathan N, Schirmer A, Reid R, Nierman WC, Lewis M, Hutchinson CR. Cloning and characterization of a gene cluster for geldanamycin production in streptomyces hygroscopicus nrrl 3602. FEMS Microbiol Lett. 2003;218(2):223–30.
69. Rascher A, Hu Z, Buchanan GO, Reid R, Hutchinson CR. Insights into the biosynthesis of the benzoquinone ansamycins geldanamycin and herbimycin, obtained by gene sequencing and disruption. Appl Environ Microbiol. 2005;71(8):4862–71.
70. Chen X-H, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, et al. Structural and functional characterization of three polyketide synthase gene clusters in bacillus amyloliquefaciens fzb 42. J Bacteriol. 2006;188(11):4024–36.
71. Piel J. Biosynthesis of polyketides by trans-at polyketide synthases. Nat Prod Rep. 2010;27(7):996–1047.
72. Zhao C, Ju J, Christenson SD, Smith WC, Song D, Zhou X, Shen B, Deng Z. Utilization of the methoxymalonyl-acyl carrier protein biosynthesis locus for cloning the oxazolomycin biosynthetic gene cluster from streptomyces albus ja3453. J Bacteriol. 2006;188(11):4142–7.
73. Kopp M, Irschik H, Gemperlein K, Buntin K, Meiser P, Weissman KJ, Bode HB, Müller R. Insights into the complex biosynthesis of the leupyrrins in sorangium cellulosum so ce690. Mol BioSyst. 2011;7(5):1549–63.
74. Rouhiainen L, Jokela J, Fewer DP, Urmann M, Sivonen K. Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in anabaena (cyanobacteria). Chem Biol. 2010;17(3):265–73.
75. Fell DA, Small JR. Fat synthesis in adipose tissue. an examination of stoichiometric constraints. Biochem J. 1986;238(3):781–6.
76. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28(3):245–8.

Publisher's Note