J Ind Microbiol Biotechnol (2014) 41:211–217
DOI 10.1007/s10295-013-1337-8

MINI-REVIEW

SIMB
Society for Industrial Microbiology and Biotechnology

Evolutionary concepts in natural products discovery: what actinomycetes have taught us

Janko Diminic · Antonio Starcevic · Mohamed Lisfi · Damir Baranasic · Ranko Gacesa · Daslav Hranueli · Paul F. Long · John Cullum · Jurica Zucko

Received: 23 May 2013 / Accepted: 2 September 2013 / Published online: 24 September 2013

Abstract Actinomycetes are a very important source of natural products for the pharmaceutical industry and other applications. Most of the strains belong to *Streptomyces* or related genera, partly because they are particularly amenable to growth in the laboratory and industrial fermenters. It is unlikely that chemical synthesis can fulfil the needs of the pharmaceutical industry for novel compounds so there is a continuing need to find novel natural products. An evolutionary perspective can help this process in several ways. Genome mining attempts to identify secondary metabolite biosynthetic clusters in DNA sequences, which are likely to produce interesting chemical entities. There are often technical problems in assembling the DNA sequences of large modular clusters in genome and metagenome projects, which can be overcome partially using information about the evolution of the domain sequences. Understanding the evolutionary mechanisms of modular clusters should allow simulation of evolutionary pathways in the laboratory to generate novel compounds.

Keywords Polyketide synthase · Non-ribosomal peptide synthetase · *Streptomyces* · Genome mining

J. Diminic · A. Starcevic · D. Baranasic · R. Gacesa · D. Hranueli · J. Zucko (✉)
Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, 10000 Zagreb, Croatia
e-mail: jzucko@pbf.hr

M. Lisfi · D. Baranasic · J. Cullum
Department of Genetics, University of Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany

P. F. Long
Institute of Pharmaceutical Science, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

P. F. Long
Department of Chemistry, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

Secondary metabolite biosynthesis genes occur as clusters and there are many secondary metabolite clusters in each genome

David Hopwood and collaborators [12] developed genetic systems for the model species *Streptomyces coelicolor* A3(2) and produced a genetic map of the chromosome based on conjugation. This showed that *Streptomyces* are very different from the ruling *E. coli* paradigm of the time. For instance, the genes for each amino acid biosynthesis pathway (e.g. *his* genes) were scattered throughout the chromosome instead of being clustered as in *E. coli*. However, the genes for the biosynthesis of a secondary metabolite (e.g. the actinorhodin genes, *act*) were all in a cluster. Later, they developed cloning methods for *Streptomyces* strains and were able to show that the *act* genes formed a physical cluster on the chromosome [17]. The cluster could be transferred to other strains, enabling them to also synthesize actinorhodin.

The rule that all the genes in a secondary metabolite biosynthesis pathway are present in a single cluster has proved surprisingly resilient despite the subsequent accumulation of vast amounts of data. This has translated into standard practice in the laboratory. In order to clone the genes for the biosynthesis of a new secondary metabolite, a strategy is developed to clone one gene and it is assumed that the other genes are surrounding it in a cluster. The clustering is not essential for the functioning of the genes

as numerous complementation experiments involve splitting clusters. In general, the selective pressures controlling whether genes in a common pathway are clustered in the chromosome are not very clear [2]. However, if a pathway can be considered as "selfish DNA", which can increase its evolutionary success by horizontal gene transfer, there will be selection for clustering so the complete pathway can be transferred. A striking example is the distribution in nature of the secondary metabolite clusters responsible for β-lactam production i.e. with the non-ribosomal peptide synthetase (NRPS) producing the tripeptide precursor. Comparison of the sequences of actinomycete and fungal genes showed that they are very similar suggesting that horizontal gene transfer has occurred [15, 27]. Different lines of evidence support the idea that β-lactam clusters evolved in bacteria and were subsequently transferred into fungi [7]. As transfer of clusters to other species in the laboratory usually results in secondary metabolite production, it seems likely that clusters can be considered as selfish DNA and that horizontal gene transfer could be the main reason for clustering.

A second striking observation is that a single *Streptomyces* strain can produce many secondary metabolites [19]. Genetic analysis of the model strain *S. coelicolor* A3(2) identified gene clusters for three antibiotics (see references in [12]). When the genome sequence of the strain was obtained, it was possible to recognize at least 22 secondary metabolite biosynthesis clusters [5]. Other species have even more clusters (e.g. at least 30 in *S. avermitilis* [19] and 48 for *S. rapamycinicus* [4]). The clusters present in different species are very different. The large number of compounds and their diversity mean that there must be strong selection for novel compounds.

What are the roles of secondary metabolites in nature?

There has been much debate about the roles of secondary metabolites in nature. As many have antibiotic activity it is attractive to believe that they are produced to inhibit competitors, but it is very difficult to demonstrate this in soil systems [16]. It is likely that some function as signal molecules to coordinate the activities of different individuals of the species [14]. The fact that the biosynthesis clusters often contain many genes and are highly regulated suggests that significant selection is acting on them. Both the antibiotic and signal molecule hypotheses would explain the diversity of secondary metabolites. Resistance develops to any widespread antibiotic so that novel antibiotics would give their producer strains a selective advantage i.e. development of new weapons for the "war in the soil". Signal pathways are a target for competing microorganisms, so that there would be selection for the development of novel signal molecules, which are not yet targeted i.e. development of new codes to prevent "espionage in the soil".

Understanding the evolution of secondary metabolite clusters should help screening programs for novel natural products. There are two critical questions governing the evolution of chemical diversity. The first one is whether organisms can evolve biosynthesis genes for any particular chemical entity, e.g. whether there are any classes of compounds of pharmaceutical interest, for which biological limitations would prevent synthesis by bacteria. It would be an interesting challenge to find an approach to defining the theoretical range of biological compounds and to see if there are gaps, which might be interesting for chemical synthesis approaches. If an organism evolves a cluster for a new secondary metabolite, it will only be found in nature if it provides a selective advantage. In the pharmaceutical industry, much screening of natural products tests activity against human targets and it is perhaps surprising that screening has been so successful considering that the compounds have mostly undergone selection for targets in very different organisms. This suggests that it would be useful to simulate evolution by manipulating clusters to produce chemical structures, which may have been produced during evolution, but have not been successfully selected in the soil environment. This approach would require both a good understanding of the evolution of secondary metabolites and tools to manipulate the clusters.

Evolution of modular biosynthetic clusters

Any detailed analysis of the evolution of secondary metabolite biosynthetic clusters has to consider the interplay between DNA sequence and the structure of the compounds which are under selection. The best analysed cases are of modular clusters, because they are constructed on a building block principle with a simple relationship between genetic structure and chemical structure. The modular clusters can be divided into polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS) and mixed clusters containing modules of both types [9]. Each extension reaction is usually catalysed by a single module. The modules are constructed from domains and the extender unit incorporated can usually be deduced by analysing the activity of the individual domains of the module [25]. The minimal module in each case consists of three domains (Fig. 1). There is a domain which is covalently attached to the chemical product by a thiol linkage [acyl carrier protein (ACP) or peptidyl carrier protein (PCP) for PKS and NRPS modules respectively], a domain which selects the substrate to be incorporated [acyl

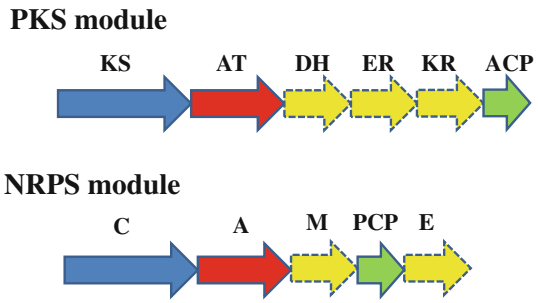

Fig. 1 Domain structures of PKS and NRPS extender modules. In each case the minimal module has three domains, which select the substrate (AT or A respectively), bind the new substrate (ACP or PCP) and couple the growing chain from the previous module to the new substrate (KS or C). There may be further domains present such as reduction domains for PKS or methylation or epimerization domains for NRPS

transferase (AT) or adenylation (A)] and a domain which couples the growing chain to the new substrate [keto synthase (KS) or condensation (C)]. There may be additional domains, which modify the extender unit. For PKS modules the unmodified extender unit carries a keto group. Reduction domains can modify the keto group in up to three reactions: ketoreductase (KR) produces a hydroxyl group, which may be reduced to a double bond by dehydratase (DH) and this may be fully reduced by enoyl-reductase (ER). These domains also control the stereochemistry. An NRPS module may contain an epimerase domain (E) or a methyltransferase domain (M), which modifies the substrate amino acid [26].

Differences in the numbers of modules and their specificities allow modular systems to synthesize a vast array of different chemical structures including many important pharmaceuticals (e.g. activities ranging from antibiotic and antiparasitic to immunosuppressant and anticancer and even lowering cholesterol). The question arises as to how new modular clusters can evolve. Most work has been carried out with PKS clusters. Phylogenetic analysis of domains shows an interesting pattern. The KS domains of clusters usually group together on phylogenetic trees. However, the AT domains usually group in clades with domains from other clusters with the same substrate specificity. As KS and AT domains are adjacent in modules (Fig. 1), this indicates strongly that recombination must play an important role in the evolution of clusters. An interesting model was proposed by Jenke-Kodama and collaborators [13]. They proposed (Fig. 2a) that novel clusters could be created by amplification of a single module. Subsequent recombination events with other clusters in a genome could replace the AT and reduction domains and give rise to the diversity of module types present in a cluster. As the KS domains are derived from a

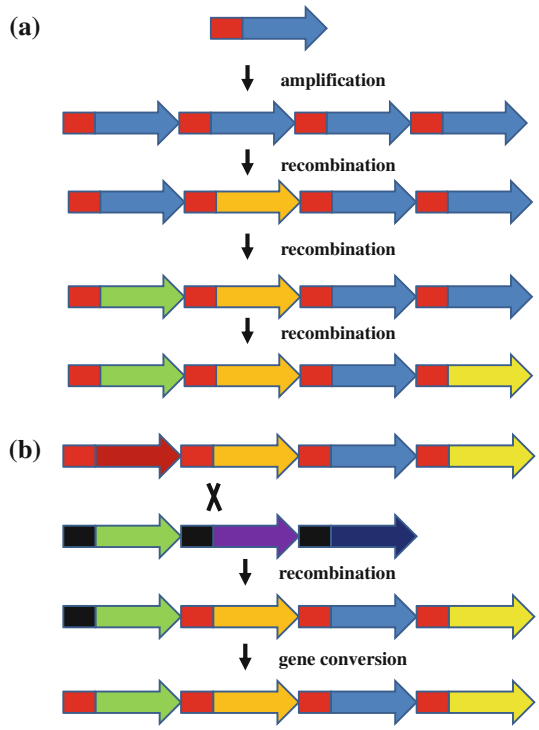

Fig. 2 Models for the evolution of new PKS modular clusters. **a** The amplification model [13]. A single module is duplicated to form a modular PKS with identical modules. Subsequently, in a series of recombination events with other PKS clusters, the AT domains and/or reduction domains are replaced by ones with different specificity. **b** The single-crossover recombination model [30]. Two clusters undergo a single crossover to produce a recombinant—this will occur in a region of high sequence similarity such as a KS domain. Subsequently, gene conversion between KS domains will result in highly similar KS domain sequences

single module, the clustering pattern in phylogenetic trees would be explained.

Zucko and collaborators [30] suggested an alternative explanation for the patterns of the phylogenetic trees. They investigated the evolution of whole clusters rather than individual domains. The clusters were considered as "organisms" and it was assumed that clusters of very similar genetic structure were closely related. By comparing such clusters, it was possible to identify orthologous modules. Comparison of the orthologous modules showed that gene conversion occurred frequently within PKS clusters. Gene conversion would account for the grouping of the KS domains of a cluster in phylogenetic trees. Gene conversion also occurs for AT domains, but usually occurs between AT domains with the same substrate; careful examination of the phylogenetic trees shows that AT domains with the same substrate from a cluster do group together closely. When gene conversion involves AT domains with different substrates, the whole domain is replaced, probably because a hybrid domain would be inactive. An alternative model for the evolution of new

clusters is that recombination between homologous domains in two clusters generates a new cluster (Fig. 2b) [23]. Subsequent gene conversion events would produce the observed phylogenetic patterns for KS and AT domains. Comparison of the structure of the PKS genes in different clusters has revealed three examples of cluster pairs which are candidates for such recombination events [6, 30].

Any model for evolution of secondary metabolite clusters has to explain selection of the new products. The size and complex organisation of most clusters makes it likely that the products are under selection. Otherwise, clusters would be expected to accumulate mutations, which would prevent production of functional modular PKSs. When high quality genome sequences are analysed, the cluster genetic structures are compatible with production, with no apparent pseudogenes [5, 20]. When the coding sequences of modular PKS genes are compared, most differences involve synonymous codons, i.e. changes in the DNA sequence not the protein sequence [29]. This indicates strong purifying selection, i.e. most amino acid replacements would be deleterious. In the amplification model (Fig. 2a) the initial amplification product as well as the products from each recombination event would have to produce a product with a selective advantage. Changes in the reduction domains after some of the recombination events are likely to make large differences in the final chemical structure (e.g. ring formation). Thus, a series of selected products of very different chemical structures would be needed to reach the final product and this seems somewhat implausible. In contrast, the recombination model (Fig. 2b) yields a new selected product in a single recombination step. In most cases, "tailoring" enzymes modify the initial polyketide product, e.g. glycosylation, hydroxylation or methylation. These enzymes are usually encoded in the biosynthetic cluster and are often important for the biological activity of the product [9]. In the recombination model, tailoring enzyme genes would be present in the parental clusters and might also use the recombinant product as a substrate; recombination would yield products with sections of the polyketide chain having identical chemical structures to those of one of the parent molecules [23]. In the amplification model, the parental module could also be derived from a cluster with tailoring enzymes. However, the amplification process would yield an initial product chemical structure with little relation to that of the parental cluster. In either model, subsequent small modifications of the chemical structure would be selected for optimal product activity. These modifications might include module duplication or domain replacement as envisaged in the amplification model [21].

Much less is known about evolution of NRPS clusters. Although the genetic structures of the PKS and NRPS clusters are similar, there is a significant difference in how diversity of modules is achieved. In PKS clusters, there are four different substrates for extender modules and this affects the nature of side chains in the product. Most modules incorporate malonyl-CoA or methyl-malonyl-CoA and much of the chemical diversity is achieved by combinations of reduction domains, which determine the group incorporated (keto group, alcohol, double bond or fully reduced) as well as the stereochemistry. These groups are important for the possibilities of forming ring structures and play an important role in the activity of polyketides. In contrast to PKS modules, which usually contain reduction domains, most NRPS modules are minimal modules only containing the three domains C–A–PCP (Fig. 1). Thus, although epimerization domains, which allow the synthesis of peptides with D-amino acids, are important for the biological activity of NRPS products, most of the chemical diversity is generated by the large number of substrates (>500) incorporated by the different A domains. Structural studies of the A domain combined with bioinformatics analyses defined 8–10 amino acid residues in the binding pocket which are responsible for substrate specificity [10, 22]. Little is known as to how the A domains evolved such a diversity of substrates, but there is some evidence suggesting co-evolution of C and A domains [3, 26]. Many of the C domains of NRPS clusters group together in phylogenetic trees [6], whereas the sequences of A domains mainly reflect the substrate specificity. These data could be explained by gene conversion similar to that observed in PKS clusters [30].

Use of evolutionary data to assemble modular cluster DNA sequences

The development of DNA sequencing methods has resulted in an exponentially increasing number of sequences and a rapid drop in sequencing costs. In contrast, the labour and cost of isolating and characterizing secondary metabolites have only made modest improvements. This has led to the idea of using data mining i.e. recognizing promising clusters in DNA sequences using bioinformatics and focusing efforts on promising clusters. The structures of the products can to a large extent be predicted by analysing the sequences of the different domains and "adding" the effects of the different activities (e.g. in the ClustScan program [25]). Several computer programs have been developed to analyse the DNA sequences of modular clusters and make predictions about the chemical structures of the products. Many of these tools have been put together in the recent release of anti SMASH, which is capable of searching whole genomes for natural product biosynthesis gene clusters (including those other than coding for PKS

and NRPS), predicting both substrate selection and the associated product structure [18]. These and other programs function fairly well for good quality DNA sequences from actinomycetes. However, the cost of producing good quality sequences is very high compared to lower quality sequences. Present technology produces large numbers of relatively short reads (50–500 b depending on the system used [1]). The reads are assembled into contigs by assembly programs, which recognize overlapping reads. This process is potentially error-prone, because the assembly programs must cope with sequencing errors and allow overlap of non-identical sequence reads and parameters are chosen to make this unlikely. The assembly of simple reads usually results in hundreds or thousands of contigs rather than a single contiguous sequence. As modular clusters are usually large (50–150 kb) they are often present on several contigs. The modules have conserved sequences and are easily recognized, but it is difficult to decide which modules belong to each cluster as most *Streptomyces* strains have several different modular PKS and NRPS clusters.

One approach is to carry out more sequencing experiments such as paired-end reads. In this method, random DNA fragments with a particular size (e.g. 5 kb) are isolated and both ends sequenced. This allows contigs to be linked in scaffolds for which some sequences are missing, but the orientation and distance between contigs are known. This approach allows assembly of modular clusters, but is considerably more expensive than simple shotgun sequencing. As mentioned above, the KS domains of the PKS modules of each cluster are usually grouped together in phylogenetic trees [13, 30]. It is possible to use this fact to assign contigs containing PKS modules to clusters [6]. As the C domains of NRPS also show such a phylogenetic grouping, it is also possible to recognize the contigs belonging to different NRPS clusters. This approach has potential limitations for NRPS clusters as there are three distinct classes of C domains [3]. The most common class is responsible for coupling L-amino acids to L-amino acids and successful assembly depends on the presence of such C domains on each contig. The other two classes are needed for coupling D-amino acids to L-amino acids and coupling fatty acids to L-amino acids. The ClustScan program [25] is a semi-automatic annotation program with a graphical user interface, which allows the user to recognize the correct assembly of the clusters from the contigs by the distribution of modules and domains. It is also possible to detect sequencing errors, which result in frame shifting or false stop codons. This allows enough cluster annotation using simple shotgun reads to decide if the cluster is likely to produce a product of chemically interesting structure. The assumptions about assembly of the clusters and sequencing errors can then be easily tested by sequencing PCR products and the transcription of the cluster can be studied to suggest suitable production conditions. At present, the assembly of the clusters and error detection are done using manual intervention in ClustScan. In principle, the information generated by ClustScan and the internal data format could be developed to allow automatic assembly and error detection.

Is it possible to model PKS evolution in the laboratory to produce novel compounds?

Much work has been carried out trying to manipulate the building block principle of modular clusters to generate novel structures as the desired modifications in chemical structure are not easily accessible with the methods of organic chemistry. Targeted changes in PKS genes often result in products with the expected chemical structures, which should allow the generation of a vast range of novel products by mixing and matching different combinations of modules i.e. combinatorial schemes [28]. Unfortunately, although this is fascinating for understanding the basic science, it has proved disappointing for industrial applications as nearly every manipulation results in drastic loss of product yield making it unviable.

The situation with NRPS clusters is rather different. There has been considerable success in manipulating cluster structure to produce predicted products with reasonable yield (see Table 3 of Baltz [3]). This success was dependent on identifying suitable linker regions between domains as the junctions for constructing recombinants. Attempts to replace A domains alone were unsuccessful, possibly because of co-evolution of A and C domains. A combinatorial synthesis strategy was developed, which allowed the generation of many lipopeptide antibiotics related to the important commercial antibiotic daptomycin [3].

Natural selection acts on the products of modular clusters, so that, in nature, there must be mechanisms to generate novel PKS clusters, which retain reasonable product yields. Otherwise, the novel clusters would be equivalent to non-producers from the point of view of selection. The most likely mechanisms for generating novel clusters would be some sort of recombination.

We have suggested that recombinants generated by homologous recombination may be more likely to give good yields of products than the junctions used for in vitro constructs which do not take into account the local sequence environment [23, 24]. The selection pressures acting on protein sequences can be estimated by looking at the nature of nucleotide substitutions between related genes. If the ratio of non-synonymous/synonymous substitutions is low, i.e. most nucleotide changes do not

change the amino acid sequence, it is taken as evidence that there is strong purifying selection i.e. most amino acid changes are deleterious and selected against. In contrast, a high ratio suggests positive selection for changes which improve function. When such an analysis was carried out for PKS modules [29] it was found that the ratios were low, especially in KS domains. This suggests that in the evolution of PKS clusters it is more important to conserve functional sequences than to undergo extensive adaptation to fine-tune the cluster and increase yield.

It would be conceivable that, in nature, only a small proportion of recombinants give good yields and that the number of recombinants generated over evolutionary time is high enough to account for the observed clusters. Traditionally, evolutionary theory has considered that recombination occurs so frequently that it will produce any combinations needed for selection. However, much of this theory was developed for eukaryotes, which undergo frequent meiosis; bacterial recombination is much rarer. Even in human populations, where reproduction involves extensive recombination each generation, it has recently been realized that linkage disequilibrium (i.e. evidence of restricted recombination) is common in populations. There is, as yet, little experimental evidence for how good the yield of product after homologous recombination is. There is one example where, during experiments to manipulate the nystatin cluster, homologous recombination within the cluster resulted in deletion of a single module [8]. The strain produced a product with corresponding smaller size at about 25 % of the yield of the parent cluster. This example of efficient production by a homologous recombination product gives hope that homologous recombination will give products at high enough yields to be industrially interesting. Theoretical studies using a recombination model show that many recombination events produce a cluster structure compatible with the production of a novel compound and that the predicted compounds show considerable chemical diversity [11].

If it is possible to simulate evolution of clusters in the laboratory, it may well generate compounds which are useful for the pharmaceutical industry, but are not successful in nature. It is likely that many compounds which interact with human target proteins do not have a selective advantage for the organism in a soil environment so would not be retained by evolution. The use of evolutionary concepts should improve the ability of genome mining to yield important new lead compounds for the pharmaceutical industry.

Acknowledgments This work was supported by the grant 09/5 (to D.H.) from the Croatian Science Foundation, Republic of Croatia and by a cooperation grant of the German Academic Exchange Service (DAAD) and the Ministry of Science, Education and Sports, Republic of Croatia (to D.H. and J.C.). It was also supported by King's College London (to P.F.L.).

References

1. Ansorge WJ (2009) Next-generation DNA sequencing techniques. Nat Biotechnol 25:95–203
2. Ballouz S, Francis AR, Lan R, Tanaka MM (2010) Conditions for the evolution of gene clusters in bacterial genomes. PLoS Comp Biol 6:e1000672
3. Baltz RH (2012) Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. ACS Synth Biol. doi:10.1021/sb3000673
4. Baranasic D, Gacesa R, Starcevic A, Zucko A, Blažič M, Horvat M, Gjuračić K, Fujs Š, Hranueli D, Kosec G, Cullum J, Petković H (2013) Draft genome sequence of streptomyces rapamycinicus NRRL 5491, the producer of the immunosuppressant rapamycin. Gen Ann 1:e00581-13
5. Bentley SD, Chater KF, Cerdeño-Tárraga 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 417:141–147
6. Blazic M, Starcevic A, Lisfi M, Baranasic D, Goranovic D, Fujs S, Kuscer E, Kosec G, Petkovic H, Cullum J, Hranueli D, Zucko J (2012) Annotation of the modular polyketide synthase and non-ribosomal peptide synthetase gene clusters in the genome of *Streptomyces tsukubaensis* NRRL18488. Appl Environ Microbiol 78:8183–8190
7. Brakhage AA, Thön M, Spröte P, Scharf DH, Al-Abdallah Q, Wolke SM, Hortschansky P (2009) Aspects on evolution of fungal beta-lactam biosynthesis gene clusters and recruitment of trans-acting factors. Phytochemistry 70:1801–1811
8. Brautaset T, Bruheim P, Sletta H, Hagen L, Ellingsen TE, Strøm AR, Valla S, Zotchev SB (2002) Hexane derivatives of nystatin produced as a result of an induced rearrangement within the *nysC* polyketide synthase gene in *S. noursei* ATCC 11455. Chem Biol 9:367–373
9. Cane DE, Walsh CT, Khosla C (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. Science 282:63–68
10. Challis GL, Ravel J, Townsend CA (2000) Predictive, structure-based model of amino acid recognition by non-ribosomal peptide synthetase adenylation domains. Chem Biol 7:211–224
11. Diminic J, Zucko J, TrninicRuzic I, Gacesa R, Hranueli D, Long PF, Cullum J, Starcevic A (2013) Databases of the thiotemplate modular systems (CSDB) and their in silico recombinants (r-CSDB). J Ind Microbiol Biotechnol 40:653–659
12. Hopwood DA (1999) Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico. Microbiology 145:2183–2202
13. Jenke-Kodama H, Börner T, Dittmann E (2006) Natural biocombinatorics in the polyketide synthase genes of the actino bacterium *Streptomyces avermitilis*. PLoS Comp Biol 2:e132
14. Kitani S, Miyamoto KT, Takamatsu S, Herawati E, Iguchi H, Nishitomi K, Uchida M, Nagamitsu T, Omura S, Ikeda H, Nihira T (2011) Avenolide, a *Streptomyces* hormone controlling

antibiotic production in *Streptomyces avermitilis*. Proc Natl Acad Sci U S A 108:16410–16415
15. Kovacevic S, Weigel BJ, Tobin MB, Ingolia TD, Miller JR (1989) Cloning, characterization, and expression in Escherichia coli of the *Streptomyces clavuligerus* gene encoding deacetoxycephalosporin C synthetase. J Bacteriol 171:754–760
16. Laskaris P, Tolba S, Calvo-Bado L, Wellington EM (2010) Coevolution of antibiotic production and counter-resistance in soil bacteria. Environ Microbiol 12:783–796
17. Malpartida F, Hopwood DA (1984) Molecular cloning of the whole biosynthetic pathway of a streptomyces antibiotic and its expression in a heterologous host. Nature 309:462–464
18. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) Anti SMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucl Acids Res 39:W339–W346
19. Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. Nat Prod Rep 26:1362–1384
20. 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 U S A 98:12215–12220
21. Ridley CP, Lee HY, Khosla C (2008) Evolution of polyketide synthases in bacteria. Proc Natl Acad Sci U S A 105:4595–4600
22. Stachelhaus T, Mootz HD, Marahiel MA (1999) The specificity-conferring code of adenylation domains in non-ribosomal peptide synthetases. Chem Biol 6:493–505
23. Starcevic A, Diminic J, Zucko J, Elbekali M, Schlosser T, Lisfi M, Vukelic A, Long PF, Hranueli D, Cullum J (2011) A novel docking domain interface model predicting recombination between homologous modular biosynthetic gene clusters. J Ind Microbiol Biotechnol 13:1295–1304
24. Starcevic A, Wolf K, Diminic J, Zucko J, Ruzic IT, Long PF, Hranueli D, Cullum J (2012) Recombinational biosynthesis of polyketides. J Ind Microbiol Biotechnol 39:503–511
25. 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. Nucl Acids Res 36:6882–6892
26. Strieker M, Tanovic A, Marahiel MA (2010) Non-ribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol 20:234–240
27. Weigel BJ, Burgett SG, Chen VJ, Skatrud PL, Frolik CA, Queener SW, Ingolia TD (1988) Cloning and expression in *Escherichia coli* of isopenicillin N synthetase genes from *Streptomyces lipmanii* and *Aspergillus nidulans*. J Bacteriol 170:3817–3826
28. Wong FT, Khosla C (2012) Combinatorial biosynthesis of polyketides: a perspective. Curr Opin Chem Biol 16:117–123
29. Zucko J, Cullum J, Hranueli D, Long PF (2011) Evolutionary dynamics of modular polyketide synthases, with implications for protein design and engineering. J Antibiot (Tokyo) 64:89–92
30. Zucko J, Long PF, Hranueli D, Cullum J (2012) Horizontal gene transfer and gene conversion drive evolution of modular polyketide synthases. J Ind Microbiol Biotechnol 39:1541–1547