FEMS Microbiology Letters, 365, 2018, fny240

doi: 10.1093/femsle/fny240
Advance Access Publication Date: 27 September 2018
Review

REVIEW – Biotechnology & Synthetic Biology

Genome mining for the search and discovery of bioactive compounds: the *Streptomyces* paradigm

A.C. Ward[1,*,†] and N.E.E. Allenby[2]

[1]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK and [2]Newcastle Biomedicine Bio-Incubators, Faculty of Medicine, NE2 4HH

*Corresponding author: School of Biology, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. Tel: +44-7814052950; E-mail: Alan.Ward@ncl.ac.uk
One sentence summary: Natural products re-discovery results from their production in frequently isolated (and frequently misidentified) strains, while nature's massive combinatorial biosynthesis experiment is under-explored in diverse, prolific strains with fickle expression.
Editor: Colin Harwood
†A.C. Ward, http://orcid.org/0000-0003-4102-8694

ABSTRACT

The need for new antimicrobials is indisputable. The flight from natural products in drug discovery was unfortunate; however, the revolution that is genome mining, enabled by the explosion in sequencing technology, is a cause for hope. Nevertheless, renewed search and discovery is still a challenge. We explore novel metabolite diversity and the challenges in *Streptomyces*. Estimating the extent of novel bioactive metabolites remaining to be discovered is an important driver for future investment. Frequent re-discovery of known natural products was a major factor in big pharma exiting search and discovery, and remains a reality. We explore whether this is due to exhaustive isolation and frequent lateral gene transfer. Analysing all biosynthetic gene clusters across all genomes is challenging. Therefore, representative examples of the patterns of secondary metabolite diversity suggest that re-discovery is linked to frequent expression in frequently isolated (and frequently misidentified) strains. Lateral gene transfer of complete biosynthetic clusters is less frequent than might be perceived but frequent gene exchange implies a massive combinatorial biosynthesis experiment. Genome sequencing emphasises rare expression of many secondary metabolite gene clusters and diversification at the finest levels of phylogenetic discrimination. In addition, we are only just beginning to unravel the impact of ecology. The hidden diversity suggests that cluster cloning and heterologous expression in microbial cell factories will explore this diversity more effectively.

Keywords: genome mining; natural products; *Streptomyces*; taxonomy

INTRODUCTION

The slowly unfolding increase in antimicrobial resistance threatens a return to a pre-antibiotic era. Antimicrobial resistance is a global challenge (Boucher *et al*. 2009; Butler, Blaskovich and Coope 2017; World Health Organization 2014) on a par with climate change, but with less agreement on the concerted action needed (from paying doctors to stop prescribing antibiotics to incentivizing the search for novel blockbuster, broad spectrum antimicrobials) and less publicity and public recognition. The public health benefit of antibiotics, since the development of penicillin, has been incalculable, illustrated by the story of avermectins (Ōmura 2016). Natural products were at the heart of this golden age of antibiotic discovery (Davies 2006), though resistance to penicillin in *Staphylococcus aureus* was observed in 1944 and common by the 1950s (Chambers and DeLeo 2009), and first, second, third and fourth generation semi-synthetic analogues created by chemists have been critical. Certainly, after the initial flurry of classic antibiotics, increasing technical, commercial and regulatory barriers made the introduction of new

Received: 25 May 2018; **Accepted:** 26 September 2018

antibiotics to the clinic technically challenging, expensive and protracted (House of Lords 1998; Payne *et al.* 2007) with multiple projects funded on the prospect of rare blockbuster successes, such as amoxicillin (Geddes, Klugman and Rolinson 2007). The derivatives of many of these antibiotics are still key antibiotics, e.g. azithromycin/clarithromycin (derived from erythromycin) and doxycycline/tigecycline (derived from tetracycline). From 1940–2007, either nature-derived or nature-inspired agents led to 155 drugs and 73% of the clinically approved anticancer agents (Newman and Craggs 2007).

Nevertheless, by the 1990s re-discovery of known antibiotics, bolstered by concepts of lateral gene transfer (Egan *et al.* 1998) and failure to discover new scaffolds, led to a view that prolific producers, such as *Streptomyces*, had been comprehensively screened (Strohl 2004), sustained by nomenclatural and taxonomic confusion, e.g. *Streptomyces hygroscopicus* (see Strohl 2004; Ward and Goodfellow 2004), fickle expression and flawed ecology, e.g. marine actinomycetes as soil wash-in (see Goodfellow and Williams 1983), which delayed funding for search and discovery in this ecological hotspot (Jensen, Dwight and Fenical 1991). So commercial judgements led to an exodus of big pharma and biotech from natural product drug discovery (Payne *et al.* 2007). High throughput screening, combinatorial chemistry and new target validation through genomics of pathogens (Fleischmann *et al.* 1995) led to completely new, non-natural product strategies for drug discovery (Payne *et al.* 2007).

Many academic groups never accepted that assessment (Jensen and Fenical 1994; Berdy 1995, 2005, 2012; Ward and Goodfellow 2004) and new bioactive secondary metabolites continued to be discovered (Newman and Cragg 2007, 2016; Butler, Blaskovich and Coope 2017). Whole-genome sequencing (wgs) of prolific producers (Bentley *et al.* 2002; Ikeda *et al.* 2003; Ohnishi *et al.* 2008; Doroghazi *et al.* 2014) and bioinformatic tools for genome mining (Cimermancic *et al.* 2014; Blin *et al.* 2017b; Skinnider *et al.* 2017) have validated that belief, in terms of both microbial diversity and secondary metabolite gene clusters per genome, and their variation across phylogenetic diversity. However, that has not translated into filling the antibiotic pipeline, as: big pharma's appetite and infrastructure for natural product development has largely gone; the technical challenges to natural product-led optimisation remain; and the commercial calculus has not changed.

Despite this evolutionary diversity it is not clear how many new scaffolds that may lead to new blockbuster antibiotics remain to be discovered. Walsh and Fischbach (2010) proposed that bioinformatics, genomics, ecology and systems biology are key to exploring how many more bioactive natural product scaffolds remain to be discovered and that a test of the power of bacterial genomics and bioinformatics would be to sequence 1000 new bacterial genomes from prolific producers. Doroghazi *et al.* (2014) added 344 actinobacterial genomes and analysed 830, by genome mining and mass spectrometry, identifying 'a strong phylogenetic signal' in the pattern of secondary metabolites, seeming to validate 'new bugs = new drugs', but without any dramatic effect on new drug discovery.

GENOMES, GENOME MINING AND SEARCH AND DISCOVERY

Academic groups and Small to Medium Enterprises continue to pursue isolation and screening for search and discovery of novel secondary metabolites (Jensen *et al.* 2014; Harvey, Edrada-Ebel and Quinn 2015; Kepplinger *et al.* 2018). Genomics and genome mining has revolutionised that approach (Bachmann, Van Lanen and Baltz 2014; Ziemert, Alanjary and Weber 2016; Machado, Tuttle and Jensen 2017). Search and discovery can be initiated by the identification of a biosynthetic gene cluster (BGC) by genome mining, as illustrated by the systematic search for the products of the cryptic biosynthetic gene clusters identified in the *Streptomyces coelicolor* genome (Challis 2014). Or genome mining may follow detection of bioactivity by screening, to understand its molecular, biosynthetic and functional basis (Fernández-Martínez *et al.* 2014; Kepplinger *et al.* 2018).

This depends upon identification of BGCs, essentially based upon homology with known secondary metabolite gene clusters. These homologies are derived from the classification of known BGCs (Medema *et al.* 2015) based on pathway types (e.g. polyketides synthetase -PKS I, PKS II, nonribosomal peptide synthetase - NRPS, or ribosomally synthesized and post-translationally modified peptides - RiPP), domain structure, conserved motifs, hidden Markov models and chemical structure classes of the natural product synthesised. Although essentially annotation, the complexity of the data and the diverse algorithms, e.g. PrediCAT, SANDPUMA, RODEO (see Blin *et al.* 2017b), mean that the development of user-friendly pipelines, optimising usability and computational efficiency, has been critical. The two most generic programs, antiSMASH 4.0 (Blin *et al.* 2017b) and PRISM 3 (Skinnider *et al.* 2017), are complementary: with independent detection strategies; in their output goals (with PRISM more focussed on structure prediction); and in their presentation and visualisation. The strategies, data and algorithms, and comparison with a range of other prior and more specialist programs, are well described elsewhere (Blin *et al.* 2017b; Skinnider *et al.* 2017).

Although powerful methods for the prediction of product structure from sequence exist, continue to be developed and are implemented in pipelines like antiSMASH and the GNP (Genomes to Natural Products) framework of PRISM (Challis and Ravel 2000; Johnston *et al.* 2015; Skinnider *et al.* 2015, 2016), this remains challenging and somewhat partial. Similarly, identifying tailoring enzymes with diverse structures, additional primary metabolic enzymes, recruited to gene clusters to ensure substrate supply and regulatory genes, on the periphery of clusters, makes defining the extent of BGCs difficult. And, of course, identifying novel BGCs depends upon them containing at least some homologies with elements of known BGCs.

Streptomyces as prolific producers

Streptomyces, from which the examples in this review are drawn, are not the only source of bioactive natural products, even excluding plants and fungi, e.g. Challinor and Bode (2015), other related filamentous *Actinobacteria* (Lazzarini *et al.* 2000), *Bacillus* (Fickers 2012; Harwood *et al.* 2018), *Burkholderia* species (Jones *et al.* 2016), *Myxobacteria* (Schäberle *et al.* 2014) or even organisms in the human microbiome (Wilson, Zha and Balskus 2017), but *Streptomyces* are prolific producers and arguably amongst the most sampled (Labeda *et al.* 2012) and best studied.

The discovery of 23 (*S. coelicolor*), 30 (*Streptomyces avermitilis*) and 34 (*Streptomyces griseus*) secondary metabolite gene clusters in the first streptomycete genomes (Bentley *et al.* 2002; Ikeda *et al.* 2003; Ohnishi *et al.* 2008) was completely unexpected. However, antiSMASH now finds 27, 37 and 40 clusters, respectively, which is typical. The 39 *Streptomyces* whole genomes (https://antismash-db.secondarymetabolites.org/; accessed March 2018) analysed by antiSMASH and deposited in the antiSMASH

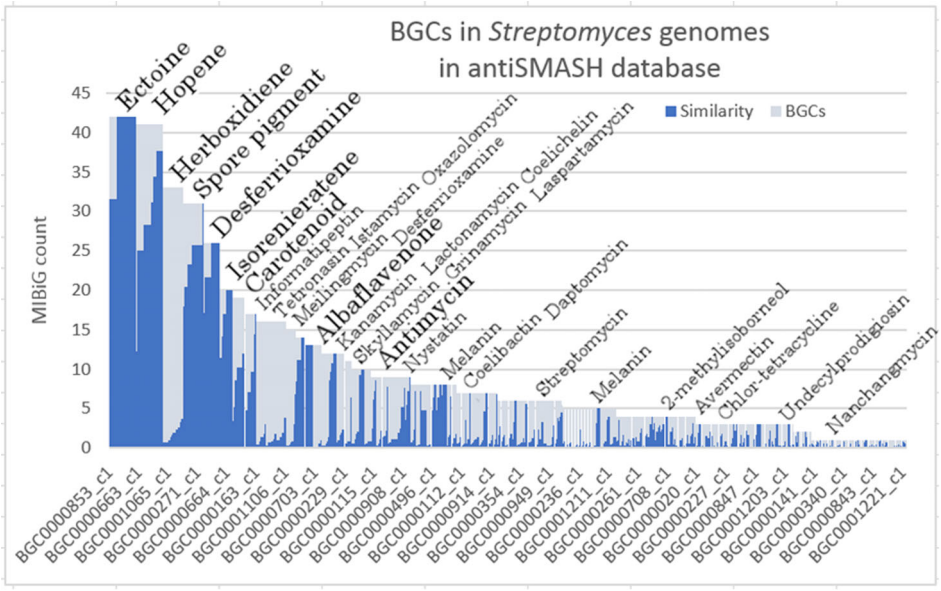


Figure 1. Distribution of BGCs linked to MIBiG ids across 39 genomes in the antiSMASH database (Blin *et al.* 2017a). Each BGC with a link to a MIBiG id is represented in a column. The column height and width is equal to the number of BGCs in that MIBiG-linked group (light blue), while the percentage of MIBiG BGC reference genes that are similar is shown (dark blue) in a stacked bar format, with that percentage of similar genes represented as the relative height in each column (e.g. for BGCs identified to MIBiG id BGC0000853_c1, whose product is identified as ectoine, there are 42 individual BGCs; 11 have 75% of the MIBiG0000853_c1 genes matching, while 31 have 100% of similar genes).

database (Blin *et al.* 2017a) have between 20 and 53 clusters (mean 32.5 ± 7.5).

If these BGCs were shared across microbial diversity (i.e. many organisms had the same BGCs, as suggested by rediscovery rates and predicted by extensive lateral gene transfer) these multiple BGCs/genomes would not reflect many more scaffolds to be discovered. The total number of BGCs is the product of the microbial diversity and the distribution of rare and unique BGCs. If every species in the Streptomycetaceae (869 validly described, http://www.bacterio.net/ accessed March 2018; Parte 2013) expressed 20 unique BCGs there would be 17,380 unique BGCs, which is not incompatible with the view that natural product diversity has been largely explored. Watve *et al.* (2001), from statistical analysis of search and discovery data, estimated that there would be 10^5 secondary metabolites from streptomycetes. We need to define the taxonomic level(s) at which BGCs diversify, estimate that microbial diversity (Bull 2004) and estimate how BGC diversity distributes across that microbial diversity, as well as determining how that cluster diversity is reflected in the structural diversity of the natural products.

This is important as natural products have been seen as challenging for modern approaches in drug discovery such as high throughput screening (Macarron *et al.* 2011) and virtual screening (Leung and Ma 2015), which require access to large numbers of compounds. The ZINC database (Irwin *et al.* 2012) describes more than 2 million chemical entities, with structures suitable for virtual docking, and links to commercial suppliers. There are 170,000 entries in the dictionary of natural products (http://dnp.chemnetbase.com) from all species, but many are only described in the literature.

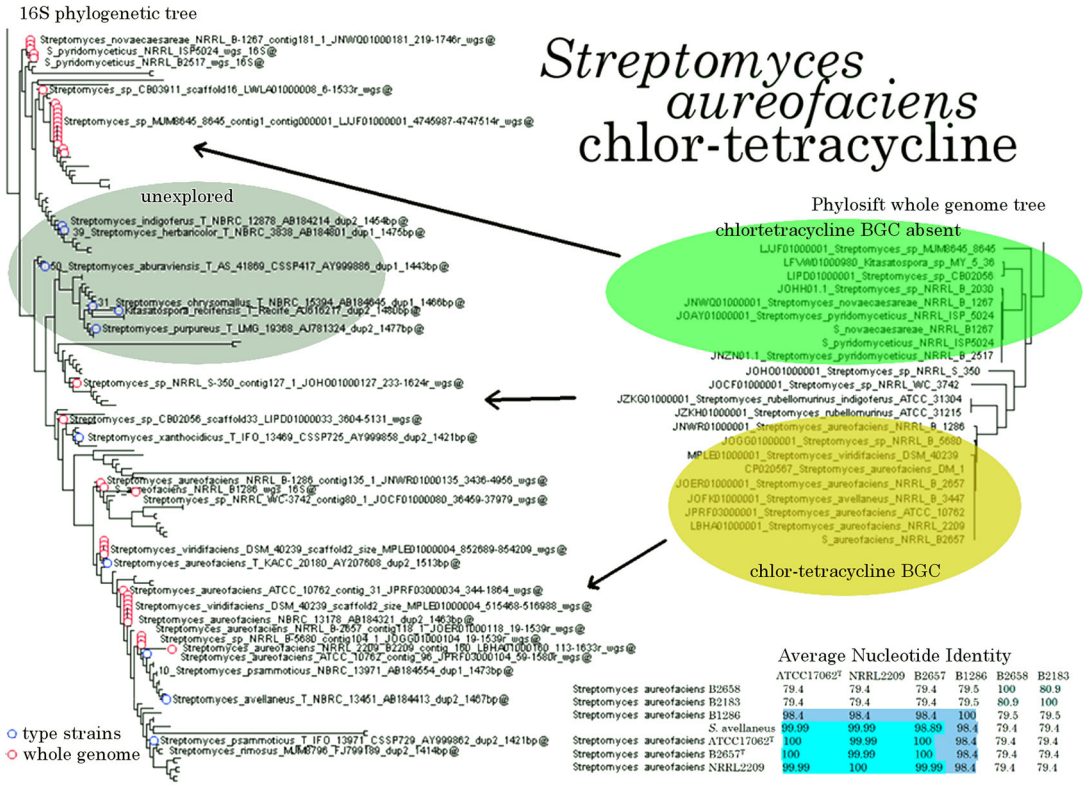
Genome mining in the Streptomycetaceae

Analysing the distribution of BGCs across diversity

Analysing the distribution of all BGCs across all genomes is challenging (Doroghazi *et al.* (2014). *Streptomyces* are not the only source of bioactive natural products, but they are prolific producers and consequently the genus is the largest in the prokaryotes, and arguably one of the most sampled and best studied. That very scale, >1000 genomes with, at least 20–30 BGCs for secondary metabolites/genome, makes a comprehensive analysis difficult.

AntiSMASH has processed 364,086 jobs, probably duplicating many analyses, and the antiSMASH database (Blin *et al.* 2017a) is building a curated, non-redundant dataset of results. The MIBiG database (https://mibig.secondarymetabolites.org/; Medema *et al.* 2015) contains reference BGCs. There are a total of 1346 BGCs detected by antiSMASH in the 39 streptomycete genomes in the antiSMASH database (https://antismash-db.secondarymetabolites.org/; accessed March 2018). This subset provides a tractable group for review to illustrate how secondary metabolites map to microbial diversity.

Some BGCs encode genes for products conserved across large taxonomic groups: ectoine, osmotic compatible solute (Sadeghi *et al.* 2014); hopene for prokaryotic membrane stability (Kannenberg and Poralla 1999; Ghimire, Koirala and Sohng 2015); siderophores like desferrioxime, for iron uptake (Barona-Gómez *et al.* 2004); melanin (Arai and Mikami 1972) and spore pigments (Novakova, Bistakova and Kormanec 2004); carotenoids like isorenieratene (Takano 2016); butyrolactones (Takano 2006), geosmin (Jiang, He and Cane 2007) and methylisoborneol (Wang and Cane 2008). Of the 1346 individual BGCs, 998 are identified as closest to 252 different MIBiG database entries while 348 (25.8%) have no significant similarity to a MIBiG BGC id. Some BGCs, like those most similar to ectoine and hopene, are present in all the genomes and show a very high percentage of similar genes to the reference BGC in MIBiG (Fig. 1). Ectoine and hopene account for 42 (Some BGCs are duplicated) and 39 of those 998 individual BGCs, respectively. Some, like the 33 BGCs that are most similar to the MIBiG entry for herboxidiene, show low similarity to the reference BGC and to one another (the similarity to MIBiG BGC0001065_c1 is based on similarity of a single, multidomain PKS, present in many different known biosynthetic



	ATCC17062T	NRRL2209	B2657	B1286	B2658	B2183
Streptomyces aureofaciens B2658	79.4	79.4	79.4	79.5	100	80.9
Streptomyces aureofaciens B2183	79.4	79.4	79.4	79.5	80.9	100
Streptomyces aureofaciens B1286	98.4	98.4	98.4	100	79.5	79.5
S. avellaneus	99.99	99.99	98.89	98.4	79.4	79.4
Streptomyces aureofaciens ATCC17062T	100	99.99	100	98.4	79.4	79.4
Streptomyces aureofaciens B2657T	100	99.99	100	98.4	79.4	79.4
Streptomyces aureofaciens NRRL2209	99.99	100	99.99	98.4	79.4	79.4

Figure 2. Chlortetracycline-producing strains and BGC-containing genomes in the *Streptomyces aureofaciens* clade. Blue circles, type strains; red circles, whole genomes; gold, chlortetracycline; green, *Streptomyces pyridomyceticus* (no chlortetracycline BGC); grey, wgs gap (16S diversity with no wgs representatives). Average nucleotide identities were calculated according to Goris *et al.* (2007).

clusters). However, most of the 252 clustered BGCs have few or no members that are unequivocally identified, and differences in the similarities to the reference MIBiG BGC indicate that many members of a cluster of BGCs are not only not unequivocally identified but must also differ from one another. Conversely, the 348 BGCs with no similarity to a reference MIBiG BGC are not identified but some may be similar to one another.

Fig. 1 shows an inverse hyperbolic curve, with a long tail (25% of the antiSMASH BGCs are off the end of the graph with no link to a reference MIBiG BGC). From the BGCs linked to MIBiG reference BGCs in Fig. 1, chlortetracycline, antimycin, albaflavenone and ectoine are chosen as examples to examine in detail, and *Streptomyces albidoflavus* as an example of genome mining in a putative species cluster.

Chlortetracycline/oxytetracycline

Chlortetracycline is a blockbuster antibiotic discovered in 1948 (Duggar 1948) and an example of a natural product that is often rediscovered through screening for antibacterial activity.

However, in the antiSMASH database set, although *Streptomyces* sp. 318 and *Streptomyces bingchenggensis* BCW-1 genomes contain a BGC with a link to the MIBiG reference for chlortetracycline, only 5% and 11% of genes show similarity. The chlortetracycline producer isolated by Duggar (1948) was *Streptomyces aureofaciens*^{T}. *Streptomyces avellaneus*^{T}, *Streptomyces psammoticus*^{T}, '*Streptomyces sayamaensis*' and '*Streptomyces viridifaciens*' are all are known chlortetracycline producers (US patent US3401088; Virgilio and Hengeller 1960; Villax 1963), reinforcing the idea of diverse producers.

There are seven whole genomes for *S. aureofaciens*, five are closely related but two differ from them and from one another (Fig. 2); the latter do not have a chlortetracycline-related BGC. *Streptomyces avellaneus*^{T} and '*S. viridifaciens*' also have whole genomes, contain the BGC for chlortetracycline and are closely related to *S. aureofaciens* with average nucleotide identities (ANI) >99% (Fig. 2). These strains fall into the same clade in a whole genome tree calculated using Phylosift (Darling *et al.* 2014).

There are 60 different strains in the *S. aureofaciens* 16S clade (bounded by *S. aureofaciens* NRRL B-1286 in Fig. 2) including the type strains for *S. avellaneus* and *S. psammoticus*, and '*S. sayamaensis*' and '*S. viridifaciens*'. In the 16S tree, *S. aureofaciens* is clearly on the periphery of the *Kitasatospora*, despite the long-accepted taxonomic analysis by Groth *et al.* (2003) disputing this. Their analysis was based on the type strain from the Institute for Micobiology and Experimental Therapy (IMET) culture collection, which differs from other culture collections. Labeda *et al.* (2017), have used multi-locus sequence typing to propose that these strains should be re-classified as *Kitasatospora* (The nomenclatural taxonomy in GenBank reflects this for *Kitasatospora aureofaciens* and *Kitasatopsora psammoticus*, but not *S. avellaneus*, and the new nomenclature enters the Organism and Taxonomy fields but not the title).

However, there are also 15 strains, including *S. aureofaciens*^{T} IMET 43577 (Groth *et al.* 2003), assigned the name *S. aureofaciens*, which fall into different 16S clades. Of these, *S. aureofaciens* B-2658 has a whole genome; it is in the *Streptomyces rimosus* 16S clade of oxytetracycline producers, contains an oxytetracycline BGC and an ANI of >99.9% to *S. rimosus*. All the whole genomes

Table 1. Demuris strains dereplicated as tetracycline producers.

DEM number	Identification		Lab strain	Potential analogue
DEM31230	*Streptomyces halstedii*	NRRL ISP5068	STS29	Tetracycline
DEM10311	*Streptomyces halstedii*	NRRL ISP5068	MU1598	Tetracycline
DEM10325	*Streptomyces halstedii*	NRRL ISP5068	MU1956	Tetracycline
DEM10341	*Streptomyces halstedii*	NRRL ISP5068	MU1561	Tetracycline
DEM10342	*Streptomyces halstedii*	NRRL ISP5068	MU1556	Tetracycline
DEM10328	*Streptomyces halstedii*	NRRL ISP5068	MU1617	Tetracycline
DEM10393	*Streptomyces halstedii*	NRRL ISP5068	MU1433	Tetracycline
DEM10422	*Streptomyces halstedii*	NRRL ISP5068	MU117	Tetracycline
DEM10430	*Streptomyces halstedii*	NRRL ISP5068	MU1618	Tetracycline
DEM10409	*Streptomyces halstedii*	NRRL ISP5068	MU1560	Tetracycline
DEM31515	*Streptomyces rimosus*		I56	Tetracycline
DEM31761	*Streptomyces rimosus*		ZT4	Tetracycline
DEM10401	*Streptomyces graminis*		MU1619	Tetracycline

with a chlortetracycline BGC fall in the *S. aureofaciens* species clade (in 16S and Phylosift whole-genome trees). In addition, the known chlortetracycline producers *S. avellaneus*, *S. psammoticus*, '*S. sayamaensis*' and '*S. viridifaciens*' also fall into the *S. aureofaciens* 16S species clade (as does a strain named *S. rimosus*). This is a pattern typical for many natural products, contributing significantly to the strong phylogenetic signal of Doroghazi *et al.* (2014). The apparent diversity of producing organisms is often the result of taxonomic confusion, or no taxonomy, which, if resolved, leaves a single species as the producing organism.

However, a strain of *Streptomyces lusitanus* is a known chlortetracycline producer (US patent US3401088) with limited sequence data from multiple strains. Based on available 16S sequence data they represent several different centres of variation and show significant differences between the two culture collection representatives of the type strain with a 16S sequence. Not all strains identified as *S. lusitanus* may produce chlortetracycline, but no strains correspond to *S. aureofaciens*.

Chlortetracycline and oxytetracycline are almost identical structures with highly similar bioactivity. Oxytetracycline was discovered, as Terramycin, in 1950 (Finlay *et al.* 1950) from *S. rimosus*. The BGCs contain homologous genes catalysing corresponding synthetic steps, nevertheless the gene clusters differ in organisation, gene content and sequence similarity (about 80% of the gene clusters match at ~80% or less sequence identity).

There is whole-genome sequence data for more than 35 genomes with an ANI >99.9% to *S. rimosus*^{T} and a 16S cluster of 70 strains. These strains include: *S. aureofaciens* (100% ANI to *S. rimosus*); *Streptomyces capuensis* (100% ANI); *Streptomyces cellulosae*; *Streptomyces griseoflavus* (96.3% ANI); *Streptomyces lavendulae* (100% ANI); *Streptomyces ochraceiscleroticus*; and *Streptomyces peucetius* (97.9% ANI). All these strains fall in the 16S clade for *S. rimosus*. All but *S. cellulosae* and *S. ochraceiscleroticus* have a whole genome, confirming both an oxytetracycline BGC and close identity with *S. rimosus*^{T}.

The frequent detection of tetracycline correlates with the high submission of 16S sequences for *S. aureofaciens*- and *S. rimosus*-related strains. So, much of the rediscovery arises from screening multiple isolates of common strains, like *S. aureofaciens* and *S. rimosus*, which readily express a powerful antibiotic, as illustrated, perhaps, by the putative prophylactic use of tetracycline by the ancient Nubians (https://www.wired.com/2010/09/antibiotic-beer/).

For oxytetracycline, strains phylogenetically distinct from *S. rimosus* produce oxytetracycline and/or contain BGCs homologous to the *S. rimosus* oxytetracycline BGC. *Streptomyces varsoviensis* is a known oxytetracycline producer, and at Demuris, screening a taxonomically dereplicated collection largely excludes *S. aureofaciens* and *S. rimosus*; however, dereplicating by screening with *tet*^{r} and *tet*^{s} mutants still detects tetracycline-positive strains (Table 1).

Most of these strains are related to *Streptomyces halstedii*. An oxytetracycline-related BGC is identified, by blastn and blastp, in the whole genomes of *S. halstedii* and *S. varsoviensis* (*Streptomyces halstedii* studied by Rong and Huang (2010) was *S. halstedii* CGMCC 4.1358^T (16S FJ405902) = *S. halstedii* LMG 19892^T (16S AJ781329), which differs from *S. halstedii* NRRL B-1238^T (16S EF178695) and is the same as NRRL ISP5068. If you accept CGMCC 4.1358/LMG 19892 as *S. halstedii*^{T} then *S. halstedii* NRRL B-1238 = *S griseolus*^{T} NRRL B-2925 (16S JOFC01000069 3472-4952) both with whole genomes and oxytetracycline BGC). The distinctiveness of these strains from *S. rimosus* is confirmed by 16S and whole-genome phylogeny. The gene content and cluster organisation match the oxytetracycline gene cluster (DQ143963) but the gene similarities differ (Fig. 3) with *S. halstedii*/*S. griseolus* (81.4% ANI to *S. rimosus*^{T}) having a highly homologous BGC (90.4% id), while *S. varsoviensis* (82.7% ANI) is more distinct (71.5% id), perhaps indicating relatively recent lateral gene transfer to *S. griseolus* (Fig. 4).

Demuris strain MU156 is an experimentally confirmed oxytetracycline producer; its whole genome has an oxytetracycline BGC >99% identical to *S. halstedii*/*S. griseolus*, a 1-base difference (1476/1477) in 16S and a whole-genome ANI of 98.2%.

Nevertheless, the distribution of the oxytetracycline is largely explained by vertical inheritance and more rarely by lateral gene transfer. Other structural analogues of tetracycline have been discovered (Fig. 5) including chelocardin (Lukezic *et al.* 2013), dactylocycline (Wang *et al.* 2012), polyketomycin (Daum *et al.* 2009) and SF2575 (Pickens *et al.* 2009) with diverse BGCs, structural variation and bioactivity. Based on the phylogeny of the conserved oxyA protein the chlortetracycline BGC differs from the oxytetracycline BGC similar to other tetracycline-related secondary metabolite BGCs (Fig. 5). Each structural analogue has been discovered in only a limited taxonomic radius (often only one strain). If these tetracycline-related analogues represent lateral gene transfer then it reinforces that it is relatively infrequent. It also indicates that the process

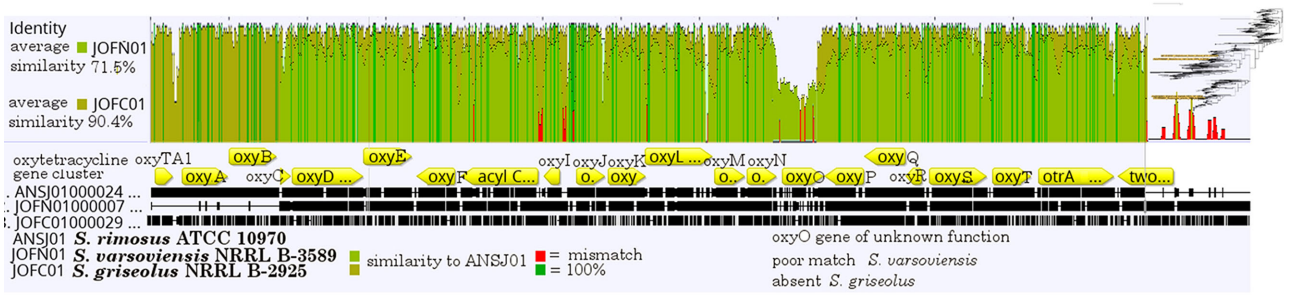


Figure 3. Similarity of oxytetracycline gene cluster in *Streptomyces griseolus* and *Streptomyces varsoviensis* compared with *Streptomyces rimosus* ATCC 10970. Insert: Phylosift whole-genome Streptomycetaceae tree showing *S. rimosus* close to *S. varsoviensis* while *S. griseolus*/ *Streptomyces halstedii* are more distant (also in 16S tree).

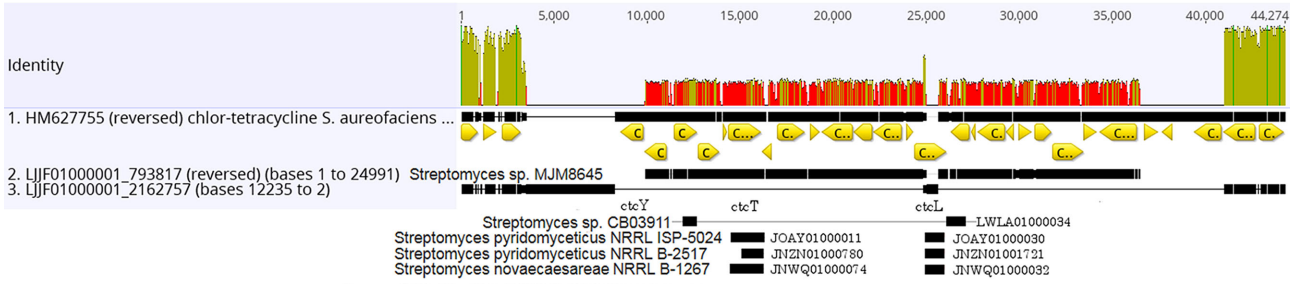


Figure 4. Alignment of chlortetracycline BGC (HM627755) to *Streptomyces* sp. MJM8645 genome (LJJF01000001 bases 793 817–818 807 and 2 162 757–2 174 991) and blast search for matches in the other members of the *Streptomyces pyridomyceticus* clade (Fig. 2): *Streptomyces novocaesaesareae* (NRRL B-1267), *Streptomyces* sp. CB03911 (LWLA01) *S. pyridomyceticus* ISP5024 (JOAY01) and NRRL B-2517 (JNZW01).

is linked to evolutionary diversification of genes and product structure.

There are many more oxyA-related genes in diverse strains than tetracycline-related BGCs. *Streptomyces albofaciens* has an oxyA homologue with high similarity to that in the oxytetracycline BGC of *S. rimosus* strains but its whole genome (BBHN01 with 3820 contigs) has only 32% of the oxytetracycline BGC, with 95% nucleotide identity but split across 14 different contigs.

Streptomyces sp. MJM8645 is present in the *Streptomyces pyridomyceticus* clade (Fig. 2) adjacent to the *S. aureofaciens* clade in both the 16S and whole-genome trees. In *Streptomyces* sp. MJM8645 the chlortetracycline BGC has diverged significantly and is not present in the genomes of the other members of the *S. pyridomyceticus* clade (ctcT and ctcL, asparagine synthetase and succinyl-diaminopimelate transaminase, are involved in asparagine and lysine biosynthesis).

Although many independently isolated strains contain the BGCs for chlortetracycline or oxytetracycline most of them are the same organism, with a very strong phylogenetic signal. Detection of tetracycline in screening reflects frequent isolation of a common organism readily expressing the tetracycline BGC, which is easily detected.

Antimycin

Antimycin-related BGCs are detected in nine *Streptomyces* genomes in the antiSMASH database, and have 100% or 93% of genes similar to MIBiG BGC0000958_c1, but antiSMASH identifies their types as nrps-t1pks (three) nrps-t1pks-lantipeptide (two) nrps-t1pks-lantipeptide-phenazine (one) nrps-terpene-t1pks (two) and indole (one—only 20% similar genes). In all Streptomycetaceae genomes, 73 were identified as containing an antimycin-related BGC (Joynt and Seipke 2018), which consists of 15, 16 or 17 genes in four polycistronic operons (Fig. 6). The antimycin BGC is found close to other putative secondary metabolite genes that are conflated into a single antiSMASH hit so the numbers of antiSMASH hits may under-estimate the real number of putative BGCs—a similar situation was observed for the iturin group BSC in *Bacillus amyloliquefaciens* (Harwood *et al.* 2018). In *S. albidoflavus*, antimycin and candicidin BGCs are adjacent and present in a single antiSMASH hit, labelled in the overview as similar to candicidin. Identifying the link to antimycin requires drilling down into the matches to homologous clusters and is easily interpretable only because both these BGCs are well defined.

Many genomes are multi-contig draft genomes, and PRISM analyses each contig separately. As illustrated in the PRISM cluster images (Fig. 6), BGC clusters may extend to contig ends. BGCs split across multiple contigs are only partially characterised, and may be detected as two BGCs or not detected at all.

Streptomyces albus J1074 is a model organism with an antimycin BGC and is a known producer. AntF is a CoA ligase that tethers anthranilate to peptidyl carrier protein antG for subsequent extension. *Streptomyces albus* NRRL B-2613^{T}, the type strain for the species, has a homologue to antF but it does not fall in the antF clade for known and putative antimycin producers (Fig. 6) and does not contain an antimycin BGC. The antF of *Streptomyces* sp. yr375 also falls outside the core antimycin-antF clade but is present in an antimycin-related genomic context (Fig. 6), though not identical in gene content or gene order, and adjacent to additional secondary metabolite genes (assigned to a cf-fatty acid type BGC by antiSMASH). *Streptomyces* sp. yr375 is representative of a small group of strains with a putative antimycin BGC, but is not known to produce antimycin.

There are 85 whole genomes in the Streptomycetaceae containing an antimycin BGC, based on searching the genomes for

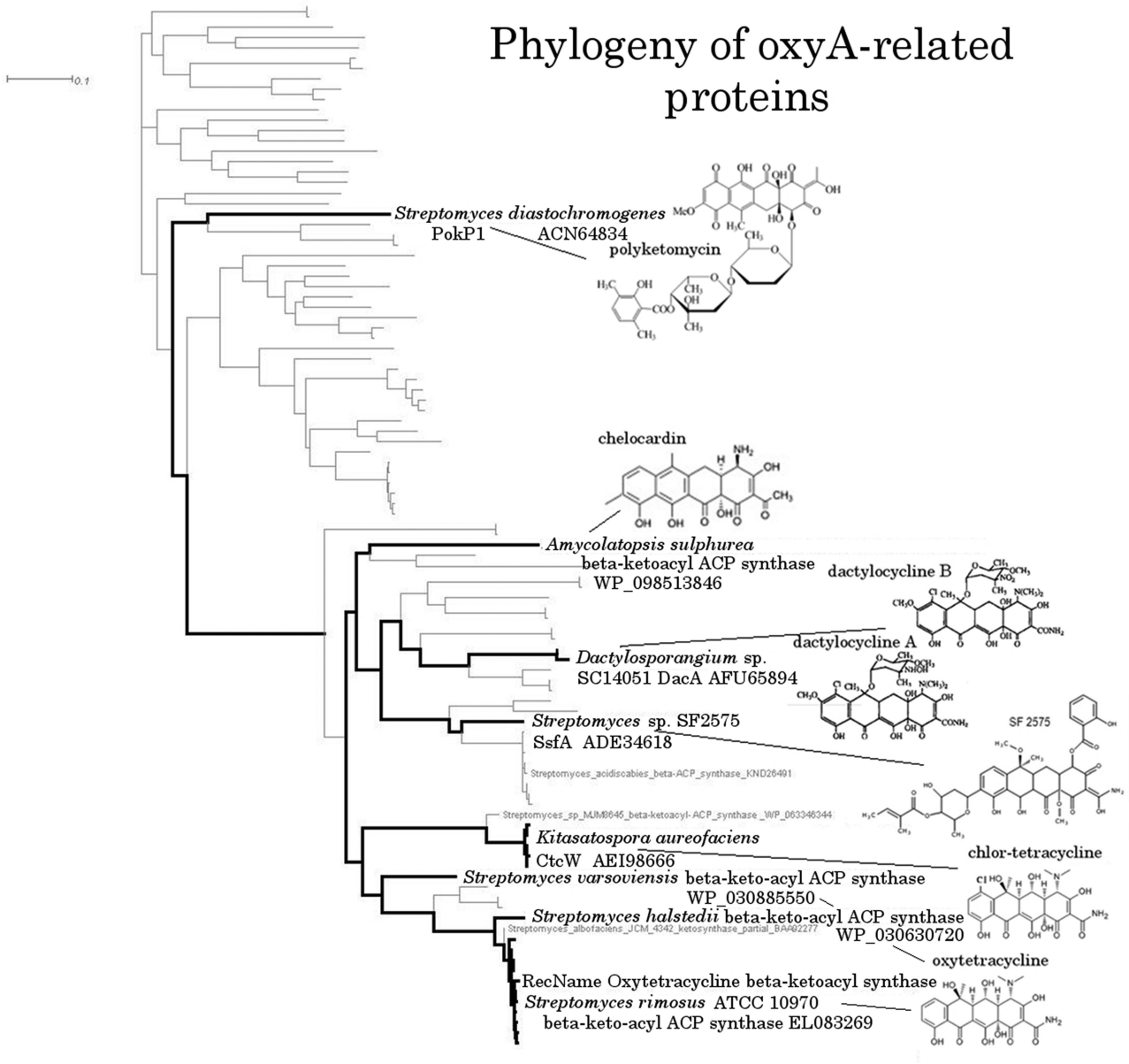


Figure 5. Phylogenetic relationship of oxyA protein sequences from tetracycline-related BGCs and tetracycline-related natural product structures (highlighted). Related beta-ketoacyl-ACP synthases not associated with tetracycline (greyed out).

the antimycin clade (Fig. 6). The BGCs for antimycin are distributed across diverse, small 16S clades (Fig. 7).

Six locations in the 16S tree contain only a single antimycin BGC strain, often with no closely related strains with whole genomes. There are four clades with multiple antimycin BGC-containing genomes, some like *S. albidoflavus*-related strains, are very closely related. Some clades with multiple antimycin BGC-containing genomes are interspersed with non-antimycin BGC-containing genomes. There are, perhaps, 10 phylogenetic centres of variation with antimycin BGCs, more extensive than tetracycline BGCs, but far less prevalent than might be suggested by 73, or 85, genomes and many different species names (Joynt and Seipke 2018). As for chlortetracycline (*S. aureofaciens*) and oxytetracycline (*S. rimosus*), 35 antimycin BGCs are present in 36 genomes for organisms that should be assigned to the *S. albidoflaus* species.

Antimycin is cytotoxic (inhibiting mitochondrial cytochrome c reductase) and anti-fungal, although of interest as a lead for development of anticancer agents. Its bioactivity is readily detected by screening. Nevertheless only six experimentally verified antimycin-producing strains are published (Joynt and Seipke 2018). *Streptomyces* sp. S4 isolated from attine ants is a known producer (Seipke *et al.* 2011), as is the closely related *S. albus* J1074 (Sandy *et al.* 2012), in mannitol-soy flour medium, but in the hands of Olano *et al.* (2014), on different media, it was silent, expressed after activation. Quezada *et al.* (2017) isolated closely related strains from cone snails (see section on *S. albidoflavus*). They screened for antifungal activity and

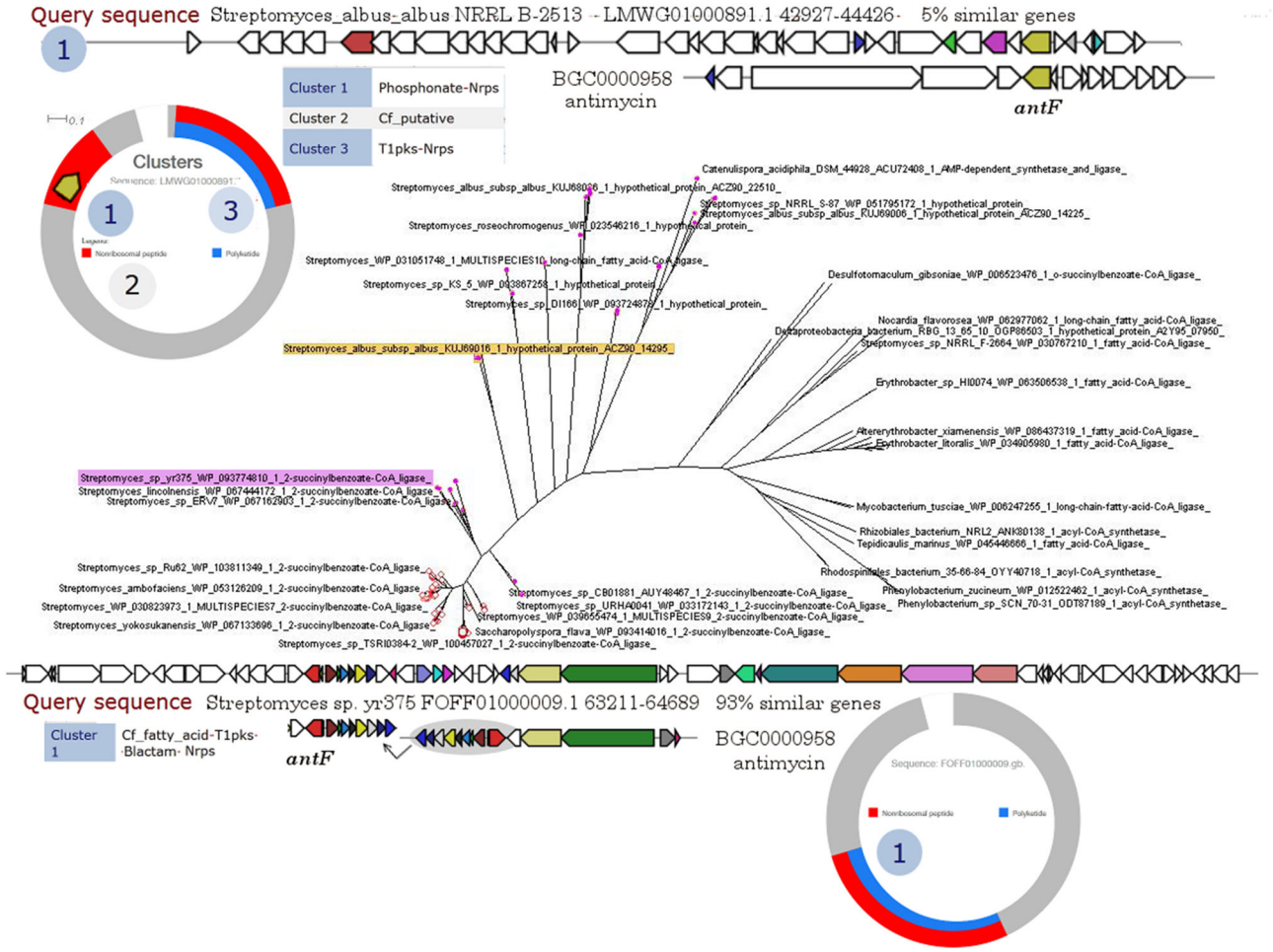


Figure 6. Identification of *antF*-related genes of the antimycin BGC in Streptomycetaceae whole genomes by blastp of *antF* from antimycin gene cluster from *Streptomyces* sp. NRRL 2288. Phylogenetic tree (FastTree) from gene sequences retrieved from GenBank, aligned with MAFFT. Red circles, annotated 2-succinylbenzoate CoA ligase and noted as *antF* (filled purple circle) annotated as 2-succinylbenzoate CoA ligase. Streptomyces sp. yr375 (highlighted purple) boundary of antimycin gene cluster (and see Joynt and Seipke 2018), and *Streptomyces albus* subsp. *albus* (highlighted yellow) NRRL B-2513 (KUJ69016) non-antimycin BGC-associated 2-succinylbenzoate CoA ligase. PRISM and antiSMASH gene clusters are shown for the wgs contigs containing the *antF*-related 2-succinylbenzoate CoA ligases from the organisms defining the boundary of the antimycin BGC clade.

detected polycyclic tetramic acid macrolactams (PTM) but not antimycin, though the antimycin BGC is present in the whole genomes. Conversely, *Streptomyces* sp. S4 was screened for antifungal activity, detecting candicidin and antimycin. A third antifungal activity remained after inactivation of these two BGCs but was unidentified, though the genome contains the BGC for a PTM.

Antimycin-related natural products are diverse, with 44 described (Liu *et al.* 2016). The original isolation of antimycin A from *Streptomyces* sp. NRRL 2288 (Dunshee *et al.* 1949) was later shown to be a mixture of antimycins A_1–A_4 and then minor amounts of A_5 and A_6. There are 29 different analogues, A_{1a}–A_{20}, with different combinations of alkyl/acyloxy groups attached as R1 and R2 to C7 and C8 on the nine-membered dilactone ring (Liu *et al.* 2016). Ten antimycin analogues were isolated and characterised from *Streptomyces* sp. THS-55 (Zhang *et al.* 2017). The rather short 16S (1386 bp KM103736) is identical to *Streptomyces sampsonii* KJ40 amongst other strains, a member of the *S. albidoflavus* species group, but no other sequence data are published. The splenocins contain benzyl/benzoyl groups at C7 and C8 and Strangman *et al.* (2009) isolated 10 splenocins, with 4 new analogues, from a marine *Streptomyces* sp. CNQ431 closely related to *S. albidoflavus*. This additional diversity around the major natural product from a BGC, both within a single strain and from different producers, is a common phenomenon. Antimycin-related natural products with more major structural variation fall into four classes based on ring size with 9-, 12-, 15- and 18-membered macrocyclic rings (Liu *et al.* 2016). Many of these natural products have been the focus for chemistry, not taxonomy, a trend that continues. The urauchimycins are antimycins with a free hydroxyl at C8. Two new urauchimycins, C and D, were characterised, both isolated from marine *Streptomyces* strain B1751 and terrestrial strain *Streptomyces* sp. AdM21. The antimycin BGC from *Streptomyces* sp. AdM21 (KJ920751) has been determined (Vanner *et al.* 2013) but there are no other data on the relationship of these strains to one another or to other streptomycetes. *Streptomyces* sp. ML55 produces 12-membered antimycins JBIR-06 and JBIR-52 (Ueda *et al.* 2008; Kozone *et al.* 2009). The BGC for JBIR-06 has been published (LC375135) and described (Li *et al.* 2013a) but the BGC is the sole sequence deposited. Neoantimycin has a 15-membered tetralactone ring structure (Li *et al.* 2013b), an antimycin BGC

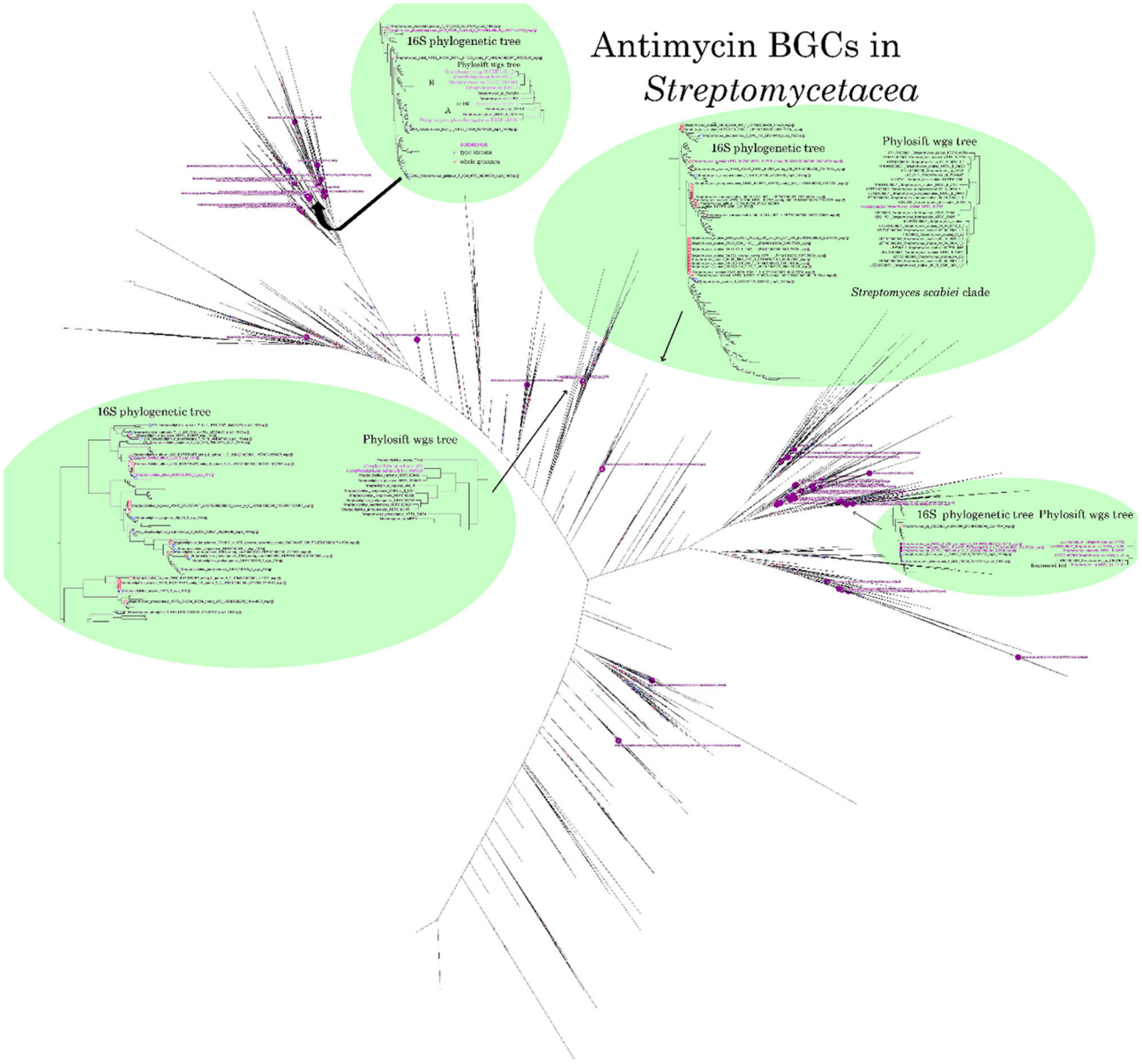


Figure 7. *Streptomyces* genomes with antimycin BGC strains mapped to a 16S phylogenetic tree and whole-genome tree (using the Phylosift pipeline; Darling *et al.* 2014). Insets in Supplementary Figs S1–6 available online.

sequence (LC375136), and a draft genome (PHNC01) for the producing organism *Streptomyces orinici* (Skyrud *et al.* 2018) has no close matches to the BGC sequence or the whole genome in the databases.

Albaflavenone/geosmin

Terpenes are structurally and biologically diverse, with perhaps 80 000 distinct natural products (Christianson 2017) largely from plants and fungi but increasingly in bacteria (Nakano *et al.* 2011, 2012), from the ubiquitous geosmin (Gerber and Lecheval 1965) to the golden pigment of *S. aureus* (Liu *et al.* 2005).

Geosmin, with a single bifunctional gene, sesquiterpene cyclase/geosmin synthase (Cane *et al.* 2006), is not identified in this antiSMASH data. Albaflavenone, identified as contributing to a camphor-like smell in *S. albidoflavus* (Gürtler *et al.* 1994), is a two-gene BGC, a sesquiterpene cyclase, epi-isozizaene synthetase, and a P450, epi-isozizaene mono-oxygenase. The albaflavenone BGC is found by antiSMASH in 12 of the 39 genomes (100% genes similar). Moody *et al.* (2012) identified the albaflavenone BGC as the most conserved secondary metabolite gene cluster in the genus *Streptomyces*, based on its presence in 11 genomes available in 2012 detected by blast of the '*S. coelicolor*' A3(2) genes *Sco5222* and *Sco5223*. Blastp of *Sco5222* against the GenBank nr database (accessed March 2018) gave 3288 hits. However, this includes hits to multiple sesquiterpene cyclases (Fig. 8).

There are about 382 blastp hits to actinobacterial genomes for an epi-isozizaene synthase with an adjacent P450 (blastp returns all protein hits from the nr database, the same protein may be present more than once: the original sequence assembly; copies in reference genomes; reference proteins; or may return a multispecies hit, a single identical protein present in multiple records). Almost all are *Streptomyces* except *Nonomuraea solani*

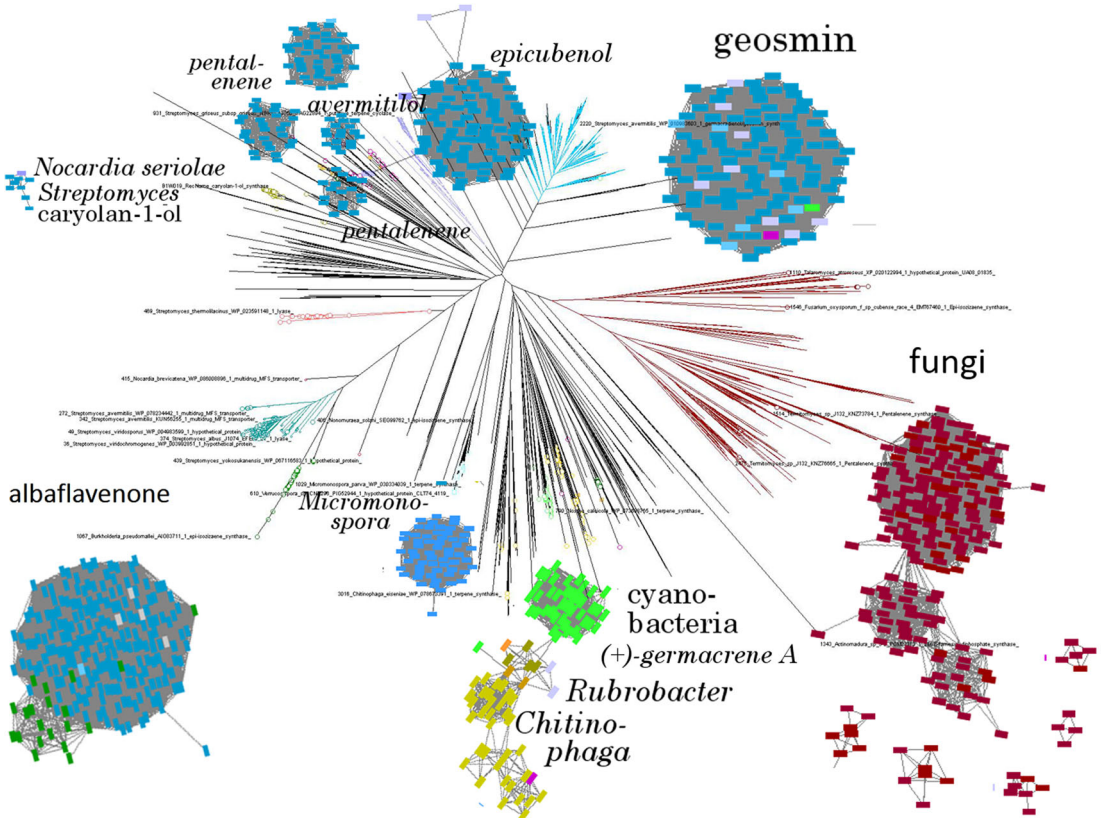


Figure 8. Phylogenetic tree (FastTree; Price, Dehal and Arkin 2010) of terpene cyclases (3288) retrieved from GenBank by blastp with SCO5222 (epi-isozizaene synthetase of albaflavenone BGC in '*S. coelicolor*' A3(2)) aligned with MAFFT (Kato *et al.* 2002; Kato and Toh 2008). SSN clusters at alignment threshold 10^{-75} (Supplementary data C1 available online) except albaflavenone (threshold 10^{-50}) to capture link with *Burhholderia* (dark green nodes).

and *Actinospica acidphilia*, which is misidentified and closely related to *Streptomyces griseorubens* JSD-1.

Representatives form a distinct clade in the phylogenetic tree of sesquiterpene cyclase-related sequences (Fig. 8). A clade of closely related epi-isozizaene synthases from *Burkholderia* strains suggests acquisition from *Streptomyces*, and a single hit to *Nocardia brevicatena*, more distantly related, also matches a sesquiterpene cyclase/P450 pair. A second sesquiterpene cyclase in *Streptomyces yokosukanensis*, not paired with a P450, marks the limits of the clade of albaflavenone-related epi-isozizaene synthases in the tree.

The albeflavenone clade is a small subset of the full sesquiterpene cyclase tree. A distinct clade of geosmin synthase is identified by sequence annotation and protein length of the bifunctional enzyme. A large diverse clade includes sesquiterpenes from eukaryotes, mostly fungi, but including an arthropod and a few plants. The prokaryotes are dominated by *Streptomyces* and other actinobacteria but include sesquiterpene cyclases from *Burkholderia*, pseudomonads and cyanobacteria (Supplementary Fig. S1 available online).

Streptomyces avermitilis sesquiterpene cyclases have been well studied and include: *sav2998*, known to encode a pentalenene synthase (Tetzlaff *et al.* 2006); *sav2163*, shown to be a germacradienol/geosmin synthase (Cane *et al.* 2006); *sav76*, which encodes a new sesquiterpene synthase for avermitilol synthesis (Chou *et al.* 2010); and *sav3032*, encoding the epi-isozizaene synthase for albaflavenone synthesis (Moody *et al.* 2012). Even after mapping these to the epi-isozizaene synthase tree (Supplementary data D1 available online) there is significant unassigned diversity. Analysis of the blast hits for Sco5222 retrieved 1160 sequences from the UniProt database by the web server at EFI-EST (Gerlt *et al.* 2015) to generate an enzyme sequence similarity network (SSN) at an alignment score threshold of 10^{-75} (Supplementary data D2 and Powerpoint slide P1 available online). The EFI-EST analysis annotates nodes in the network from multiple databases; the data set overlaps extensively with the proteins in the phylogenetic tree. Some SSN clusters from the EFI-EST analysis are mapped onto the phylogenetic tree (Fig. 8), and identification of the product is based on annotation and at least one experimentally verified producer in the cluster.

Analysis of a more extensive set of prokaryote sequences (2455 proteins retrieved using Sco5222 and *S. griseus* epicubenol synthetase SGR6065, merged, de-duplicated and eukaryotic sequences removed) at an alignment threshold of 10^{-120} identifies 199 clusters (at least two members) and 369 unclustered proteins in the SSN. There are 690 members in the putative geosmin enzyme functional cluster, 182 in the albaflavenone cluster, 56 in the epicubenol cluster and a very long tail of potentially diverse sesqiterpene cyclases with mostly unknown products (Fig. 9).

These albaflavenone BGC-containing streptomycete whole genomes map to two large clades in the 16S tree: strains related to *S. albidoflavus* and *Streptomyces violaceoruber* = '*S. coelciolor*' A3(2)'; and the 16S clade containing *S. avermitilis* (Supplementary Fig. S2 available online). In the whole-genome tree these 16S clades are in the same major branch of the tree (Fig. 9). Almost all genomes in the extended *S. albidoflavus* clade contain this BGC.

There are 300 sequenced genomes with a putative albaflavenone BGC (from a total of 1170 Streptomycetaceae

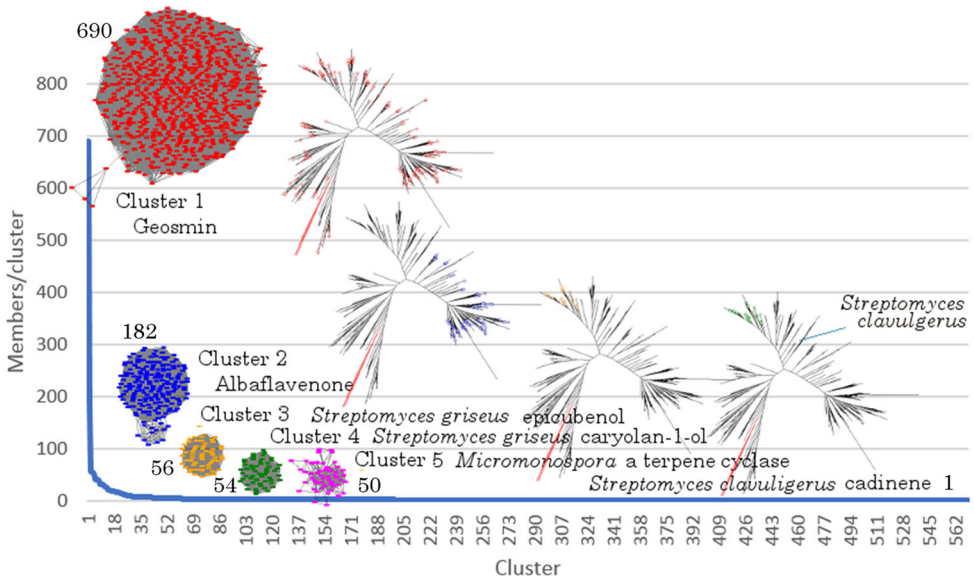


Figure 9. Numbers of members in SSN clusters. The SNN clusters and the distribution of putative natural products in clades in the Phylosift whole-genome tree are shown in the inserts for the first few clusters.

genomes). These are representative of more than 7000 strains in the albaflavenone 16S clades (out of >18 000 streptomycete 16S longer than 1200 bp in the GenBank nt database—accessed April 2017). Nevertheless, there are only six experimentally verified producers in the family Streptomycetaceae in the literature (Takamatsu *et al.* 2011; Moody *et al.* 2012). Moody *et al.* (2012) identified albaflavenone by extraction and gas chromatography–mass spectrometry analysis from *S. avermitilis*, while Takamatsu *et al.* (2011) found no detectable amounts of either epi-isozizaene or albaflavenone from the same strain. They verified the product of their bioinformatically detected BGC as albaflavenone by cloning and expression in *Escherichia coli*. This pattern for albaflavenone and geosmin illustrates a strong phylogenetic signal (Doroghazi *et al.* 2014) arising from vertical transmission or strong intra-species homogenisation by recombination (Doroghazi and Buckley 2014). The example of Takamatsu *et al.* (2011) shows the difficulties in either eliciting or detecting expression seen for antimycin.

These patterns are tracked by following the distribution of individual BGCs across biodiversity, but the individual organisms harbour multiple BGCs per genome.

Streptomyces albidoflavus—a host for heterologous BGC expression

Streptomyces albus J1074 is a widely used host for heterologous expression (Baltz 2010) with a small (for *Streptomyces*) genome of 6.9 Mb and a clean background, with secondary metabolites often unexpressed in cultivation. In the Phylosift whole-genome tree *S. albus* J1074 falls into a relatively large, closely related and distinct clade. This clade is the *S. albidoflavus* species clade, based on 16S, MLST, Phylosift whole genome phylogeny and ANI. Strains, like J1074, should be identified as *S. albidoflavus*, formally proposed by Labeda *et al.* (2017) but researchers in the field have largely ignored. The same set of strains (when they have a 16S) fall into a matching 16S clade (Fig. 10).

There are 30 whole genomes in the Phylosift wgs clade (Fig. 10) with ANI ≥96.0%, an ANI of 95% corresponding to 70% DNA–DNA similarity (Goris *et al.* 2007); there is a core of 20 strains with ANI >99%. The five cone snail isolates (Fig. 10) should be added to this clade. There are 924 16S sequences from strains in the 16S clade—there are 15 validly described type strains but only the type strain for *S. albidoflavus*^{T} has a whole genome. There are 20 different species names assigned to members of the 16S clade and 5 different species names for strains with whole genomes. Many of these species have been identified as synonyms of *S. albidoflavus* in multi-locus sequence typing (MLST) studies (Rong *et al.* 2009; Labeda *et al.* 2017).

In the Phylosift (Darling *et al.* 2014) whole-genome tree (based on conserved genes) *S. albus* J1074, *Streptomyces* sp. FR-008 and *S. sampsonii* KJ40 form a sub-cluster of strains >99% similar by ANI (Fig. 10). AntiSMASH finds 60, 61 and 67 putative BGCs, respectively. Aligning the genomes and mapping the antiSMASH clusters in these closely related strains shows the secondary metabolome is not identical. Comparing *S. albus* J1074 with *S. sampsonii* KJ40 there are 52 clusters that correspond (both in genome position and antiSMASH profile, Clusters 3 and 4 in *S. albus* J1074 match cluster 61 in *S. sampsonii* KJ40 and 2 clusters correspond but only partially match), but 7 clusters in J1074 and 15 clusters in KJ40 have no corresponding cluster in the other genome.

The first BGC that differs between *S. albus* J1074 and *S. sampsonii* KJ40 is cluster 9 (bases 601 269–629 178 in CP004370 for *S. albus* J1074) linked to MIBiG BGC0000392_c1, but in fact alignment of the two genomes shows that the *S. sampsonii* KJ40 sequence is 99% identical, but not annotated in the antiSMASH output. However, cluster 10 (bases 601 269–629 178 in CP004370) is missing from the *S. sampsonii* KJ40 genome (Fig. 11). AntiSMASH links it to MIBiG BGC0000025_c1 and the avermectin gene cluster (66% of genes show similarity) in the overview. However, the top hit to sequence in *Streptomyces paulus* NRRL 8115 (and *Streptomyces* sp. YN86) provides the link to its identification as the BGC for paulomycin(s) with 100% coverage at >99% identity by blastn of the paulomycin BGC (Li *et al.* 2015).

Paulomycin (U-43,120) was isolated from *S. paulus* in 1976 (Hanka and Dietz 1976; Wiley 1976). *Streptomyces paulus* NBRC 14877 has a 16S (AB184626), which is short (1380 bp) but identical to the core *S. albidoflaus* clade strains (Fig. 10) and derived from

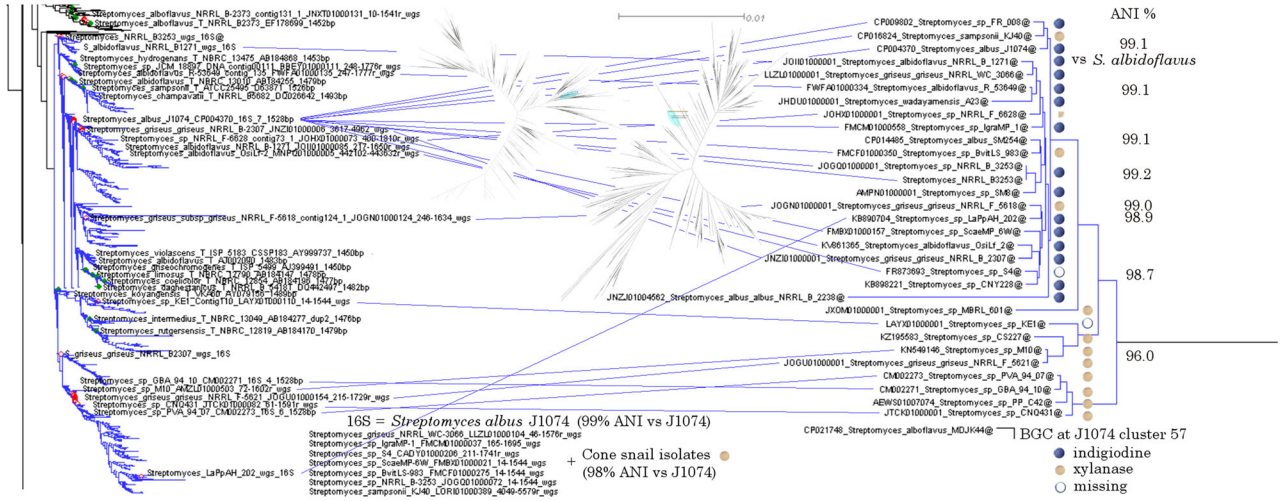


Figure 10. *Streptomyces albidoflavus* clade.

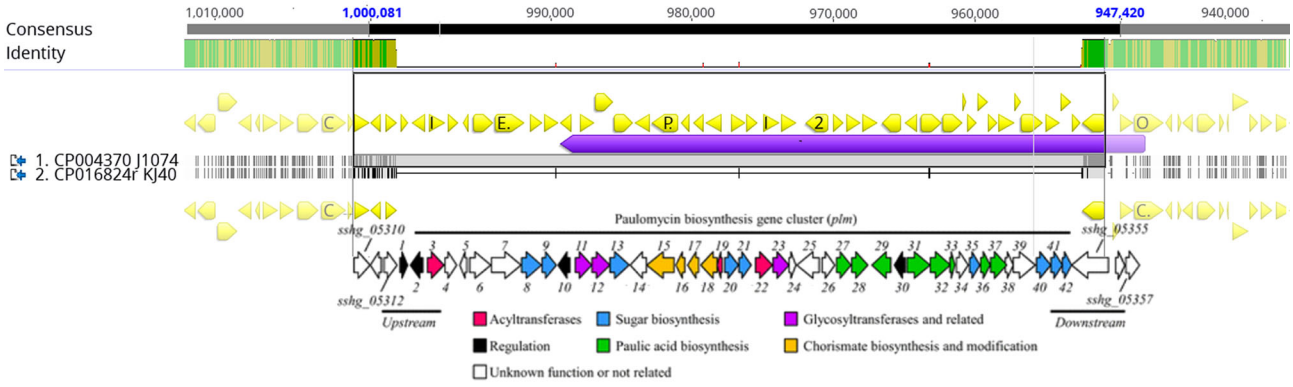


Figure 11. Paulomycin BGC (Li *et al.* 2015) in *Streptomyces albus* J1074 aligned with *Streptomyces sampsonii* KJ40.

strain Upjohn strain UC 5231. *Streptomyces paulus* NRRL 8115 is derived from UC 8560, which in turn was derived from the documented producer UC 5142 (Hanka and Dietz 1976). However, Marshall *et al.* (1984) confirm isolates UC 5142 and 5231 were identified as *S. paulus*, and both produced paulomycins.

Streptomyces albus J1074 is derived from parental strain *S. albus* G. Majer and Chater (1987) did not initially detect paulomycin in *S. albus* G but induced paulomycin production by mutagenesis with NTG. Olano *et al.* (2014) and Fernández-De la Hoz *et al.* (2017) did not detect paulomycins until they compared high performance liquid chromatography (HPLC) profiles of extracts from the parental strain with mutants in the paulomycin gene cluster to identify tiny HPLC peaks. Olano *et al.* (2014) note that the presence/absence of the five peaks linked to paulomycins was highly variable.

Blast analysis of the paulomycin BGC (KJ721164/KJ721165) identifies the cluster in only J1074, *Streptomyces* sp. CNY228 and *S. griseus* subsp. *griseus* NRRL F-5618, three members of the *S. albidoflavus* clade defined in Fig. 10, as is *S. paulus* NBRC 14877. Many BGCs detected by antiSMASH, in any one genome, do not contain enough information to identify its product and often the strain does not produce the natural products, or only in quantities insufficient to detect. Variable expression in related strains is typical, both quantitative and presence/absence. Failure to detect can only be linked to the presence of cryptic BGCs when both the presence of the gene cluster, often from a whole genome, and the natural product, with extensive screening, are associated by gene knockout experiments. González *et al.* (2016) expressed the paulomycin BGC in mutants of *S. albidoflavus* J1074 to update the genes involved in the BGC and generate novel derivatives.

Olano *et al.* (2014) set out to systematically activate cryptic clusters in *S. albidoflavus* J1074. These included antimycin and candicidin, two antifungals identified as being produced, without activation, by the attine ant-associated streptomycete *Streptomyces* sp. S4 (Seipke *et al.* 2011). This strain is a member of the *S. albidoflavus* whole-genome clade and the 16S sequence extracted from the whole genome falls in the 16S clade (note that there is a 16S for a different strain named *Streptomyces* sp. S4, AB591042, isolated from a termite mound).

In the case of antimycins, Joynt and Seipke (2018) identify three different variants of the BGC known to produce antimycin, in 73 genomes, but only six are known to express it ('Known knowns': there may be many 'unknown knowns'). Twelve of those 73 are *S. albidoflavus* (based on NRRL B-1271^{T}) and, in fact, of the 35 members of this clade (Fig. 10) with whole genomes, only *Streptomyces* sp. NRRL F-6628 lacks an antimycin BGC, although the gene cluster is scattered across multiple contigs for three strains (in one case across 24 different contigs). So, the perception that antimycin is scattered across many streptomycete species is distorted by being present in species that are commonly isolated, but frequently misidentified, at least partly

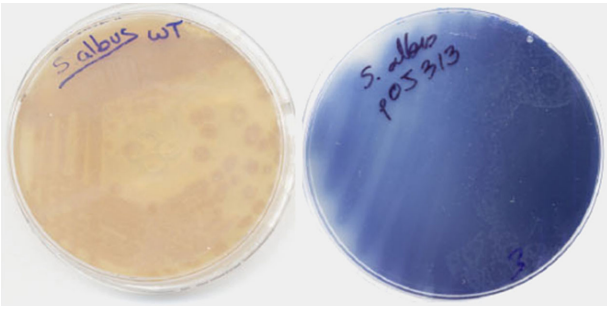


Figure 12. Expression of indigoidine blue pigment in *Streptomyces albus* J1074 (Olano *et al.* 2014).

because expression of their repertoire of secondary metabolites under laboratory conditions is so variable.

Olano *et al.* (2014) also identify the BGC for a blue pigment and induce its expression in J1074 (Fig. 12). In more than 20 years of isolating, sub-culturing and growing *Streptomyces*, I (A.C.W.) have never seen a strain identified as *S. albidoflavus* express this blue pigment. If I did see a strain this blue I would not easily identify it as *S. albidoflavus*.

However, *S. coelicolor* NBRC 12854, deposited as the type strain of *S. coelicolor*, which is *S. coelicolor* Müller, has a 16S (AB184196), placing it in the *S. albidoflavus* 16S clade. MLST (Kim 2004) also places it as member of this taxon. It was the blue colour of *S. coelicolor* Müller that famously deceived Waksman and Stanier to identify the parental strain of '*S. coelicolor*' A3(2) as *S. coelicolor* (Stanier 1942; Kutzner and Waksman 1959; Hopwood 1999) not *S. violaceoruber* (Waksman and Curtis 1916). Expression of the blue pigment by *S. coelicolor* Müller was famously fickle and difficult to observe (Habermehl and Christ 1977), while '*Streptomyces coelicolor*' A3(2) readily produces copious amounts of its blue pigment, actinorhodin (though not under all growth conditions). Diverse natural-product blue colours are widespread (Newsome, Culver and van Breemen 2014).

In J1074, cluster 57 in the antiSMASH analysis (Supplementary Table T2 available online) is the BGC for the blue pigment indigoidine. This BGC is absent from *S. sampsonii* KJ40, being replaced by genes for two xylanase degrading enzymes (Fig. 13). AntiSMASH annotates cluster 57 in the overview with similarity to the angucycline auricin, but the gene cluster for auricin from *S. aureofaciens* CCM3239 plasmid pSA3239 contains the indigoidine gene cluster hidden in the middle (Novakova *et al.* 2010) and cluster 57 matches the indigoidine cluster (idgA, idgB and indigoidine synthetase) and the uracylphosphotransferase and a drug/metabolite transporter, from the auricin BGC (only 5 out of 28 genes).

Streptomyces albidoflavus whole genomes either contain the BGC for indigoidine, at >99% similarity, or xylanase genes, which are much more variable, or a fragment of sequence in one or two strains (Fig. 10). The blue pigment from *S. coelicolor* Müller was previously identified as amylocyanin (Habermehl and Christ 1977), so clarification of *S. coelicolor* Müller as a later heterotypic synonym of *S. albidoflavus*, rather than NBRC 12854 as a contaminant (cf Kiss *et al.* 2008), is unclear.

However, calling '*S. coelicolor*' A3(2) *S. violaceoruber* (Hopwood 1999), or the recognition of *S. albus* J1074 as *S. albidoflavus* (Labeda *et al.* 2014), requires only that researchers 'choose to recognise' the proposals. Olano *et al.* (2014) note in their prelude to searching the secondary metabolome of J1074 that 'other *S. albus* strains were shown to produce the polyether salinomycin', but, of course, J1074 is unrelated to *S. albus*.

Olano *et al.* (2014) also identified a hybrid PKS-NRPS cluster related to BGCs producing polycyclic tetramate macrolactams (Blodgett *et al.* 2010) and elicited expression of two new peaks in HPLC of ethyl acetate extracts, identified as novel 6-epi-alteramides A and B, by inserting promoters before genes in the cluster.

Quezada *et al.* (2017) examined the microbiome associated with venomous cone snails and from the tissues (venom duct, stomach foot and hepatopancreas) isolated five streptomycetes that produced PTMs). One strain, *Streptomyces* CMB-CS038, was grown on four different media (ISP-4 agar, marine agar, nutrient agar and R2A agar) but only produced PTMs on ISP-4. They identified the PTMs as dihydromaltophilin (Graupner *et al.* 1997; Xu *et al.* 2015), xanthobaccin C (Hashidoko, Tahara and Nakayama 2000), frontalamide precursor F1–3 (Blodgett *et al.* 2010) and a novel product Δ^{30}-dihydromaltophilin (all related PTMs). They identified the five strains, isolated from five different cone snail species (>800 species worldwide), as related to *S. albus* J1074 and determined their whole genomes, which have an ANI of 98% to *S. albidoflavus* NRRL 1271^{T} (and *S. albus* J1074) and their 16S falls within the *S. albidoflavus* clade. The BGC for PTM biosynthesis in these genomes was identified by blastp searches for protein homology to a known PTM BGC (ikarugamycin MIBiG BGC0001435).

Quezada *et al.* (2017) correctly, we believe, submitted the whole-genome sequence data to GenBank as *S. albidoflavus* for *Streptomyces* sp. CMB-CS038, CMBCS132, CMB-CS138, CMB-CS143, CMB-CS145 (wgs: NVPZ01; NTFT01; NVPX01; NVPY01; NTFJ01), but then, in the paper, identify them as *S. albus*, based upon the undoubted similarity to *S. albus* J1074.

Quezada *et al.* (2017) have clearly linked their strains with a specific ecological niche, identified a bioactive natural product and its BGC, and identified the strains as potential snail-associated symbionts, undergoing genome reduction and conferring ecologically significant anti-fungal properties (Moree *et al.* 2014) by producing PTMs. They note that, compared with other samples from their environmental location, cone snail tissues were remarkably free from fungal contamination. The PTM BGC is cluster 2 in the antiSMASH analysis of J1074; it is present in the whole *S. albidoflavus* clade.

The evolution of PTM BGCs and the associated structural diversity of PTMs is explored in Quezada *et al.* (2017). However, genome comparisons are limited by following the nomen-

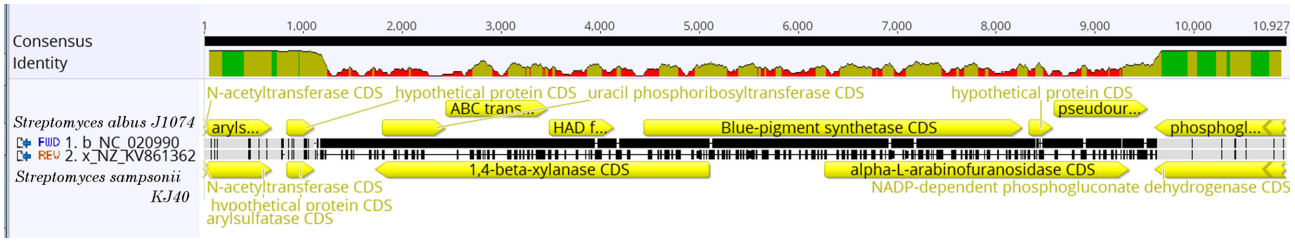


Figure 13. Indigoidine biosynthetic cluster of *Streptomyces albus* J1074 aligned with the xylanase cluster in *Streptomyces sampsonii* KJ40.

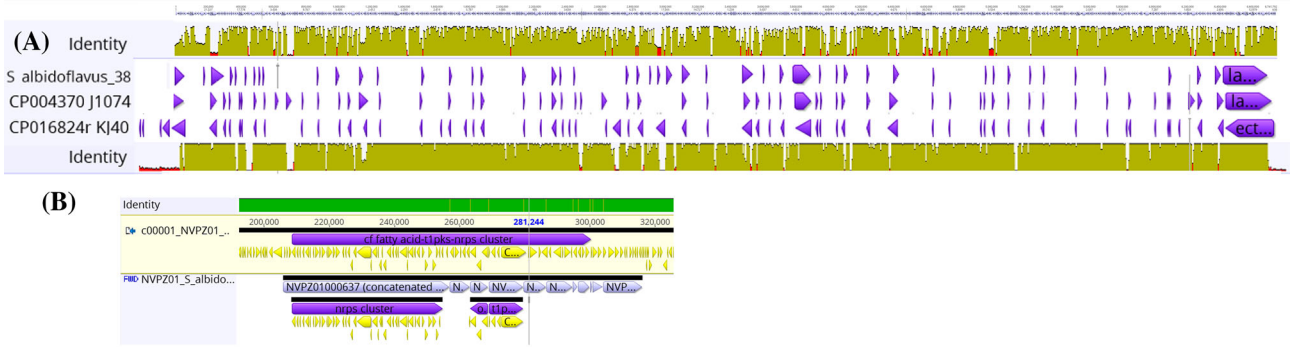


Figure 14. (**A**) Alignment of genomes, and antiSMASH clusters, for *Streptomyces sampsonii* KJ40 and *Streptomyces albidoflavus* 38 (contigs ordered to J1074 and concatenated) versus *Streptomyces albus* J1074. (**B**) AntiSMASH clusters: multi-contig *S. albidoflavus* 38 vs concatenated genome PTM BGC.

clatural taxonomic link to *S. albus*, and the whole-genome sequence data are in approximately 1000 contigs, so many BGCs are present on more than one fragment including the BGC for PTM. Of the 49 antiSMASH clusters identified in the concatenated genome (the *S. albidoflavus* 38 genome contigs can be aligned and ordered to J1074 because it is so similar), 14 are not detected by antiSMASH in the multi-contig analysis although 71 BGC clusters are found, with BGCs split across multiple contigs (Fig. 14B).

Clearly, even within a single species (by any current criteria), there can be evolution, seen in 16S, whole genomes (Phylosift and ANI) and BGC proteins, leading to expression of structural variation from an evolving BGC. In this, case 6-epi-alteramides A and B from J1074 (99% ANI to *S. albidoflavus* NRRL 1271^{T}), expressed after activation, and dihydromaltophilin, xanthobaccin C, frontalamide precursor F1–3 and Δ^{30}-dihydromaltophilin from *Streptomyces* sp. CMB-CS038 (98% ANI to *S. albidoflavus* NRRL 1271^{T}), expressed on ISP-4. Quezada *et al.* (2017), also cogently describe the difficulties with chemical structure assignment, nomenclature and scaffold numbering schemes, specifically for PTMs, but their analysis resonates more generally. Nomenclatural confusion is not limited to taxonomy.

Komaki *et al.* (2018) examine BGCs in a few related strains and conclude that strains in the same species may share secondary metabolite-biosynthetic pathways, but even closely related species have species specific pathways. However, Seipke (2015) studied six strains closely related to *S. albus* J1074 and identified significant intra-species diversity including paulomycins and alteramides in J1074, enterocin and kijanamycin in *Streptomyces* sp. PVA 94–07 and S. sp. GBA 94–10 (Ian *et al.* 2014), kendomycin and fredericamycin in *Streptomyces* sp. S4. This intra-species diversity is reflected in the *Salinispora* (Ziemert *et al.* 2014).

Discovery of novelty

Diverse PTMs have been isolated independently from multiple actinobacteria (Quezada *et al.* 2017). On the other hand, leinamycin (Hara *et al.* 1989) is a unique natural product structure, a 1,3-dioxo-1,2-dithiolane moiety spirofused to an 18-membered macrolactam (Pan *et al.* 2017), and a promising anticancer drug lead with a unique mechanism of action. However, it and a precursor, leinamycin E1 generated by combinatorial biosynthetic efforts, were the only known structural analogues, as judged by the limited success for rational combinatorial biosynthesis in a well-studied biosynthetic model (Huang *et al.* 2015). Using the domain of unknown function and cysteine lyase didomain (DUF-SH) sequence data, specific for sulfur incorporation in the known leinamycin BGC, a search by Pan *et al.* (2017) discovered 19 putative leinamycin gene clusters from the publicly available genome sequences, and another 30 from a PCR-based search of 5000 actinobacteria in their in-house collection. From fermentation optimisation and metabolite profiling, novel metabolites guangnanmycins from *Streptomyces* sp. CB01883 and weishanmycins from *Streptomyces* sp. CB02120–2 were characterised and a family of diverse metabolites showcased (see Fig. 4 in Pan *et al.* 2017).

Streptomyces sp. TSRI03842 has a putative leinomycin BGC (Pan *et al.* 2017); it also has a PTM BGC that is the most closely related to the PTM BGC in the *S. albidoflavus* clade. Its 16S places it in the centre of the *S. albidoflavus* 16S clade, as displayed in Fig. 10 (16S = *Streptomyces rutgersensis*^{T}), which is poorly sampled with publicly available genomes. It has an ANI of 91.2% to *S. albidoflavus* NRRL 1271^{T}, which makes it the closest genome outside that species clade. In this related species conserved secondary metabolite BGCs show comparable similarity to the PTM BGCs, but the overall profile of antiSMASH BGCs differs significantly from *S. albidoflavus*^{T}, including for example, similarity in the genomic context of the putative leinamycin BGC but the leinamycin BGC itself is absent in *S. albidoflavus*^{T}.

The strains with whole genomes encompassed within the *S. albidoflavus* species have been isolated from diverse environments: marine (sediments, marine sponges, marine cone snails); fresh water; insects (attine ants, bumble bees,); animals (human skin); plants (rhizosphere, endophytes); and soil. The link of BGC diversity to ecological niche has been explored in attine ants (Seipke *et al.* 2011) and cone snails (Quezada *et al.* 2017). Pan *et al.* (2017) note the diversity of strains producing leinamycin structural analogues—both the diversity of their ecological links but also how strains producing the same analogue are frequently closely related. Cheng *et al.* (2015) propose that MLST analysis of *S. albidoflavus* strains reveals habitat barriers to homologous recombination (Doroghazi and Buckley 2010, 2014). Doroghazi and Buckley (2010, 2014) also show that intraspecies homologous recombination is several orders of magnitude greater than interspecies recombination, factors favouring species homogenisation enabling population-level synergy and contingency, and rapid response to selective pressures in ecological niches.

Clearly, there is an inherent diversity of potential natural products with a strong phylogenetic signal. Related strains have a related repertoire of secondary metabolite gene clusters, but despite a high level of homologous recombination (Doroghazi and Buckley 2010, 2014) evidence prevalent, random lateral gene transfer is exaggerated by taxonomic confusion. But even the

most closely related strains exhibit BGC diversity, both in their presence in the genome and, more radically, in their expression. Once strains have diverged sufficiently to be recognised as different species, the secondary metabolome has often changed, except for a relatively small number of common metabolites like geosmin, ectoine, hopanoids and desferroximes. Although distantly related species may possess the same BGC this lateral gene transfer may happen often, but is relatively rarely conserved, perhaps the result of the two orders of magnitude difference in intra- vs inter-species homologous recombination (Doroghazi and Buckley 2010, 2014) and selective pressure. So related BGCs producing the same scaffold, or its analogues, often show significant sequence evolution in diverse species. However, what this high rate of recombination does seem to drive is gene transfer and a massive combinatorial biosynthetic experiment exploring natural product biodiversity and bioactivity (Pan *et al.* 2017).

This suggests an evolutionary strategy to maximise the secondary metabolome in the population but minimise the genetic load in individuals, but a drive, by natural selection, to optimise both diversity and expression to maximise survival in diverse ecological niches.

A strategy of screening for bioactivity expression can succeed by screening multiple isolates of the same species from diverse environments in different cultural conditions in order to find the strain that expresses a given BGC. This strategy is most successful for strains that are isolated often and BGCs that are often expressed—maximising rediscovery. However, multiple silent BGCs elude this strategy.

Fickle expression

The expression of amylocyanin (Habermehl and Christ 1977) and indigoidine is typical of what many involved in search and discovery know: expression is fickle. Also, we can see in the examples of *S. albidoflavus* and its expression of candicidin (*Streptomyces* sp. FR-008, Chen *et al.* 2003), candicidin and antimycin (*Streptomyces* sp. S4, Seipke *et al.* 2011), maltophilins (cone snail streptomycetes, Quezada *et al.* 2017) and cryptic clusters in J1074 (Olano *et al.* 2014). Similarly, in re-examining the biosynthetic gene cluster for chloramphenicol, Fernández-Martínez *et al.* (2014), with the known producing strain, were unable to elicit expression and were forced to use a host strain for heterologous expression. Far more of the secondary metabolome is encoded in the genomes of strains, than is expressed. For screening, in each organism, diversity is hiding behind the most readily expressed bioactive metabolites and the regulatory subtleties attuning the strain to its ecology, and the global diversity is hiding in the long tail of rare organisms and rare BGCs, hiding right next to the most commonly isolated.

Genome mining and cell factories for expression of cryptic biosynthetic clusters

From genome sequencing it seems that bioactive natural products discovered by screening are just the tip of an iceberg and contained in the DNA of these bacteria are a huge number of biosynthetic gene clusters that have not been previously tested for bioactivity. This conclusion is based on both many additional BGCs in known and screened strains but also, as illustrated in the examples, the additional diversity of strains, species and subspecies, at the taxonomic levels at which the secondary metabolome diversifies.

Most of these BGCs are often not expressed under the conditions used for screening, and although strategies to elicit expression of cryptic gene clusters are being pursued, they tend to be specific for individual gene clusters, and sporadically successful.

However, as illustrated, a more generic methodology that follows on naturally from genome mining and bioinformatics is cluster cloning and heterologous expression. However, wgs currently generates sequences with hundreds or thousands of contigs with assembly gaps at repeat regions, like rRNA, but also polyketide synthetases and non-ribosomal peptide synthetases. Long-read sequencing (PacBio or Nanopore) with high levels of indel errors potentially generate frameshift errors that will interfere with genome mining. Combining high throughput short-read and long-read technology provides an effective strategy (Gomez-Escribano *et al.* 2015; Wick *et al.* 2017); multiplex nanopore sequencing (Cao *et al.* 2017) combined with high accuracy short-read sequencing to correct errors may be cost effective for screening.

With high-quality genomes, high-throughput BGC detection, dereplication and prediction will enable prioritisation. Then host/vector improvement through advanced synthetic biology methods and rapid and efficient genetic manipulation of whole gene clusters will enable expression of potentially novel natural products (Nah *et al.* 2017).

Heterologous expression will potentially generate yields high enough to obtain sufficient material for drug-led discovery and to build natural product libraries with sufficient numbers and amounts to enable high-throughput screening. In addition. super-hosts for heterologous expression provide the basis for microbial cell factories for production of natural products as drug products or substrates for chemical modification.

CONCLUSION

If we are to learn by experience then cancer is a good model; the war on cancer (National Cancer Act 1971; Milestones in Cancer Research and Discovery) began, like the quest to get to the moon, as a single target goal. But, cancer is not like that—there are many different cancers, each with its own causes, mechanisms of action, consequences and potential treatments. The war against cancer has not been won but anyone visiting a child leukaemia ward in the 1960s knows each battle is worth winning—and natural product search and discovery has played its part and should continue to do so.

Infectious diseases, or even more broadly microbial activities that diminish or promote human health (because study of our microbiomes already demonstrates beyond doubt the all-encompassing influence of our microbes) are even more diverse than cancers. Yet Payne *et al.* (2007), in learning from the failed strategies of the last 20 years, seem to reflect current thinking and still set the goal of an ideal novel antibacterial as a blockbuster drug (Box 1).

The experience with *Clostridium difficile* infections following treatment with blockbuster drugs (Slimings and Riley 2014) and our growing knowledge of the essential roles that our indigenous microbial populations play in our overall health (Young 2017) should tell us that not only is that goal (Box 1) probably impossible, but also it is not desirable (Singh, Young and Silver 2017). In practice we already use our arsenal of blockbuster drugs in specific combinations for some infectious diseases such as TB or gonorrhoeae, but in search and discovery, and production, society and big pharma only seem to target 'ideal' antibiotics.

If we are to learn from nature, *Streptomyces* have an arsenal of antimicrobials that they (probably) deploy infrequently,

in minute amounts and only in response to specific cues. They invest a large amount of their overall genomic resources to the task. Tackling the problem of antimicrobial resistance is multifactorial, but anyone with children who reads the 1940s and 1950s editions of Spock (1957), whose experience came from the 1920s and 1930s, should come away with a sense of how the golden age of antibiotics changed life—going back is not an option. Potentially, microbial cell factories can provide a vehicle for both search and discovery of these 'gifts of the earth' and their production to enable modern high-throughput screening, in a new micro-pharma era.

Box 1

An ideal antibiotic

'Consider that an ideal novel antibacterial to treat community respiratory infections will need to cover seven key pathogens that cause such infections (*S. pneumoniae*, *H. influenzae*, *S. aureus*, *Moraxella catarrhalis*, *Mycoplasma pneumoniae*, *Chlamydia pneumoniae* and *Legionella pneumophila*). To treat such infections, a single compound must inhibit the growth of many different Gram-positive and Gram-negative bacterial species, all of which have different molecular targets, different membrane permeabilities and different metabolic pathways. Making such a compound is a profound chemical challenge: consider that Gram-positive *S. pneumoniae* and Gram-negative *H. influenzae* bacteria share less in common genetically than do humans and paramecia. Moreover, it must also demonstrate an acceptable side-effect profile at the high blood levels typically required to ensure effectiveness against the least susceptible organisms.' (Payne *et al.* 2007).

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSLE* online.

Conflict of interest. None declared.

REFERENCES

Arai T, Mikami Y. Chromogenecity of Streptomyces. *Appl Microbiol* 1972;**23**:402–6.

Bachmann BO, Van Lanen SG, Baltz RH. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol* 2014;**41**:175–84.

Baltz RH. *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J Ind Microbiol Biotechnol* 2010;**37**:759–72.

Barona-Gómez F, Wong U, Giannakopulos AE *et al.* Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J Am Chem Soc* 2004;**126**:16282–3.

Bentley SD, Chater KF, Cerdeño-Tárraga AM *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 2002;**417**:141–7.

Bérdy J. Are actinomycetes exhausted as a source of secondary metabolites? In: Debanov VG (ed). *Biotehnologija (ISSN 0234-2758) Proc. 9th Int. Symp. On the Biology of Actinomycetes, Moscow 1994*. 1995; pp. 13–34.

Berdy J. Bioactive microbial metabolites. *J Antibiot* 2005;**58**:1–26.

Bérdy J. Thoughts and facts about antibiotics: Where we are now and where we are heading. *J Antibiot* 2012;**65**:385–95.

Blin K, Medema MH, Kottmann R *et al.* The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res* 2017a;**45**:D555–9.

Blin K, Wolf T, Chevrette MG *et al.* AntiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 2017b;**45**:W36–41.

Blodgett JA, Oh DC, Cao S *et al.* Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc Natl Acad Sci USA* 2010;**107**:11692–7.

Boucher HW, Talbot GH, Bradley JS *et al.* Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis* 2009;**48**:1–12.

Bull AT. *Microbial Diversity and Bioprospecting*. Washington, DC: ASM, 2004.

Butler MS, Blaskovich MAT, Coope MA. Antibiotics in the clinical pipeline at the end of 2015. *J Antibiot* 2017;**70**:3–24.

Cane DE, He X, Kobayashi S *et al.* Geosmin biosynthesis in *Streptomyces avermitilis*. *J Antibiot* 2006;**59**:471–9.

Cao MD, Nguyen SH, Ganesamoorthy D *et al.* Scaffolding and completing genome assemblies in real-time with nanopore sequencing. *Nature Commun* 2017;**8**:14515.

Challinor VL, Bode HB. Bioactive natural products from novel microbial sources. *Ann N Y Acad Sci* 2015;**1354**:82–97.

Challis GL. Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J Indust Microbiol Biotech* 2014;**41**(S1):219–32.

Challis GL, Ravel J. 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* 2000;**187**:111–4.

Chambers HF, DeLeo FR. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat Rev Microbiol* 2009;**7**:629–41.

Chen S, Huang X, Zhou X *et al.* Organizational and mutational analysis of a complete FR-008/candicidin gene cluster encoding a structurally related polyene complex. *Chem Biol* 2003;**10**:1065–76.

Cheng K, Rong X, Pinto-Tomás AA *et al.* Population genetic analysis of *Streptomyces albidoflavus* reveals habitat barriers to homologous recombination in the diversification of streptomycetes. *Appl Environ Microbiol* 2015;**81**:966–75.

Chou WKW, Fanizza I, Uchiyama T *et al.* Genome mining in Streptomyces avermitilis: cloning and characterization of SAV_76, the synthase for a new sesquiterpene, Avermitilol. *J Am Chem Soc* 2010;**132**:8850–1.

Christianson DW. Structural and chemical biology of terpenoid cyclases. *Chem Rev* 2017;**117**:11570–648.

Cimermancic P, Medema MH, Claesen J *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 2014;**158**:412–21.

Darling AE, Jospin G, Lowe E *et al.* PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014;**2**:e243.

Daum M, Peintner I, Linnenbrink A *et al.* Organisation of the biosynthetic gene cluster and tailoring enzymes in the biosynthesis of the tetracyclic quinone glycoside antibiotic polyketomycin. *Chembiochem* 2009;**10**:1073–83.

Davies J. Where have all the antibiotics gone? *Can J Infect Dis Med Microbiol* 2006;**17**:287–90.

Doroghazi JR, Albright JC, Goering AW *et al*. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* 2014;**10**:963–8.
Doroghazi JR, Buckley DH. Widespread homologous recombination within and between *Streptomyces* species. *ISME J* 2010;**4**:1136–43.
Doroghazi JR, Buckley DH. Intraspecies comparison of *Streptomyces pratensis* genomes reveals high levels of recombination and gene conservation between strains of disparate geographic origin. *BMC Genomics* 2014;**15**:970.
Duggar BM. Aureomycin: a product of the continuing search for new antibiotics. *Ann N Y Acad Sci* 1948;**51**:177–81.
Dunshee BR, Leben C, Keitt GW *et al*. The isolation and properties of antimycin A. *J Am Chem Soc* 1949;**71**:2436–7.
Egan S, Wiener P, Kallifidas D *et al*. Transfer of streptomycin biosynthesis gene clusters within streptomycetes isolated from soil. *Appl Environ Microbiol* 1998;**64**:5061–3.
Fernández-De la Hoz J, Méndez C, Salas JA *et al*. Novel bioactive paulomycin derivatives produced by *Streptomyces albus* J1074. *Molecules* 2017;**22**:1758.
Fernández-Martínez LT, Borsetto C, Gomez-Escribano JP *et al*. New insights into chloramphenicol biosynthesis in *Streptomyces venezuelae* ATCC 10712. *Antimicrob Agents Chemother* 2014;**58**:7441–50.
Fickers P. Antibiotic compounds from *Bacillus*: why are they so amazing? *Am J Biochem Biotechnol* 2012;**8**:38–43.
Finlay AC, Hobby GL, P'an SY *et al*. Terramycin, a new antibiotic. *Science* 1950;**111**:85.
Fleischmann RD, Adams MD, White O *et al*. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;**28**:496–512.
Geddes AM, Klugman KP, Rolinson GN. Introduction: historical perspective and development of amoxicillin/clavulanate. *Int J Antimicrob Agents* 2007;**30**(Suppl 2):S109–12.
Gerber NN, Lecheval HA. Geosmin, an earthy-smelling substance isolated from actinomycetes. *Appl Microbiol* 1965;**13**:935–8.
Gerlt JA, Bouvier JT, Davidson DB *et al*. Enzyme function Initiative-Enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* 2015;**1854**:1019–37.
Ghimire GP, Koirala N, Sohng JK. Activation of cryptic hop genes from *Streptomyces peucetius* ATCC 27952 involved in hopanoid biosynthesis. *J Microbiol Biotechnol* 2015;**25**:658–61.
Gomez-Escribano JP, Castro JF, Razmilic V *et al*. The *Streptomyces leeuwenhoekii* genome: *de novo* sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. *BMC Genomics* 2015;**16**:485.
González A, Rodríguez M, Braña AF *et al*. New insights into paulomycin biosynthesis pathway in *Streptomyces albus* J1074 and generation of novel derivatives by combinatorial biosynthesis. *Microb Cell Fact* 2016;**15**:56.
Goodfellow M, Williams ST. Ecology of actinomycetes. *Ann Rev Microbiol* 1983;**37**:189–216.
Goris J, Konstantinidis KT, Klappenbach JA *et al*. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;**57**:81–91.
Graupner PR, Thornburgh S, Mathieson JT *et al*. Dihydromaltophilin; a novel fungicidal tetramic acid containing metabolite from *Streptomyces* sp. *J Antibiot* 1997;**50**:1014–9.
Groth I, Schultze B, Boettcher T *et al*. *Kitasatospora putterlickiae* sp. nov., isolated from rhizosphere soil, transfer of *Streptomyces kifunensis* to the genus *Kitasatospora* as *Kitasatospora kifunensis* comb. nov., and emended description of *Streptomyces aureofaciens* Duggar 1948. *Int J Syst Evol Microbiol* 2003;**53**:2033–40.
Gürtler H, Pedersen R, Anthoni U *et al*. Albaflavenone, a sesquiterpene ketone with a zizaene skeleton produced by a streptomycete with a new rope morphology. *J Antibiot* 1994;**47**:434–9.
Habermehl G, Christ BG. Amylocyanin, the blue pigment of *Streptomyces coelicolor*. *Naturwissenschaften* 1977;**64**:97–98.
Hanka LJ, Dietz A. U-43,120, a new antitumor antibiotic I. production, biological activity, microbiological assay and taxonomy of the producing micro-organism. *J Antibiot* 1976;**29**:611–7.
Hara M, Asano K, Kawamoto I *et al*. Leinamycin, a new antitumor antibiotic from *Streptomyces*: Producing organism, fermentation and isolation. *J Antibiot* 1989;**42**:1768–74.
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**:111–29.
Harwood CR, Mouillon J-M, Pohl S *et al*. Secondary metabolite production and the safety of industrially important members of the *Bacillus subtilis* group. *FEMS Microbiol Rev* 2018;**42**:721–38.
Hashidoko Y, Tahara S, Nakayama T. Xanthobaccin antibiotics. World Patent WO 2000020418, 2000.
Hopwood DA. Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. *Microbiology* 1999;**145**:2183–202.
House of Lords Select Committee on Science and Technology Seventh Report. *Resistance to Antibiotics and Other Antimicrobial Agents*. London, UK: Stationary Office, 1998, [online].
Huang SX, Yun BS, Ma M *et al*. Leinamycin E1 acting as an anticancer prodrug activated by reactive oxygen species. *Proc Natl Acad Sci USA* 2015;**112**:8278–83.
Ian E, Malko DB, Sekurova ON *et al*. Genomics of sponge-associated *Streptomyces* spp. closely related to *Streptomyces albus* J1074: Insights into marine adaptation and secondary metabolite biosynthesis potential. *PLoS One* 2014;**9**:e96719.
Ikeda H, Ishikawa J, Hanamoto A *et al*. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotech* 2003;**21**:526–31.
Irwin JJ, Sterling T, Mysinger MM *et al*. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* 2012;**52**:1757–68.
Jensen PR, Fenical W. Strategies for the discovery of secondary metabolites from marine bacteria – ecological perspectives. *Ann Rev Microbiol* 1994;**48**:559–84.
Jensen PR, Dwight R, Fenical W. Distribution of actinomycetes in near-shore tropical marine sediments. *Appl Environ Microbiol* 1991;**57**:1102–8.
Jensen PR, Chavarria KL, Fenical W *et al*. Challenges and triumphs to genomics-based natural product discovery. *J Indust Microbiol Biotech* 2014;**41**:203–9.
Jiang J, He X, Cane DE. Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat Chem Biol* 2007;**3**:711–5.
Johnston CW, Skinnider MA, Wyatt MA *et al*. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat Commun* 2015;**6**:8421.
Jones C, Bull MJ, Jenner M *et al*. Genome mining and characterisation of multiple bioactive compounds from a *Burkholderia gladioli* isolate collection. *Planta Med* 2016;**82**(S 01):S1–S381.
Joynt R, Seipke RF. A phylogenetic and evolutionary analysis of antimycin biosynthesis. *Microbiology* 2018;**164**:28–39.
Kannenberg EL, Poralla K. Hopanoid biosynthesis and function in bacteria. *Naturwissenschaften* 1999;**86**:168–76.

Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008;**9**:286–98.

Katoh K, Misawa K, Kuma K-I *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66.

Kepplinger B, Morton-Laing S, Seistrup KH *et al.* Mode of action and heterologous expression of the natural product antibiotic vancoresmycin. *ACS Chem Biol* 2018;**13**:207–14.

Kim WY. The *Streptomyces violaceoruber* clade: a model system for defining a post-genomic species concept for Streptomycetes. Thesis, University of Newcastle upon Tyne. School of Biology. University of Newcastle upon Tyne. 2004.

Kiss Z, Ward AC, Birkó Z *et al.* *Streptomyces griseus* 45H, a producer of the extracellular autoregulator protein factor C, is a member of the species *Streptomyces albidoflavus*. *Int J Syst Evol Microbiol* 2008;**58**:1029–31.

Komaki H, Sakurai K, Hosoyama A *et al.* Diversity of nonribosomal peptide synthetase and polyketide synthase gene clusters among taxonomically close Streptomyces strains. *Sci Rep* 2018;**8**:6888.

Kozone I, Ueda JY, Takagi M *et al.* JBIR-52, a new antimycin-like compound, from Streptomyces sp. ML55. *J Antibiot* 2009;**62**:593–5.

Kutzner HJ, Waksman SA. *Streptomyces coelicolor* Muüller and *Streptomyces violaceoruber* Waksman and Curtis, two distinctly different organisms. *J Bacteriol* 1959;**78**:528–38.

Labeda DP, Goodfellow M, Brown R *et al.* Phylogenetic study of the species within the family Streptomycetaceae. *Antonie Van Leeuwenhoek* 2012;**101**:73–104.

Labeda DP, Doroghazi JR, Ju K-S *et al.* Taxonomic evaluation of *Streptomyces albus* and related species using multilocus sequence analysis and proposals to emend the description of *Streptomyces albus* and describe *Streptomyces pathocidini* sp. nov. *Int J Syst Evol Microbiol* 2014;**64**:894–900.

Labeda DP, Dunlap CA, Rong X *et al.* Phylogenetic relationships in the family Streptomycetaceae using multi-locus sequence analysis. *Antonie Van Leeuwenhoek* 2017;**110**:563–83.

Lazzarini A, Cavaletti L, Toppo G *et al.* Rare genera of actinomycetes as potential sources of new antibiotics. *Antonie Van Leeuwenhoek* 2000;**78**:399–405.

Leung C-H, Ma D-L. Recent advances in virtual screening for drug discovery. *Methods* 2015;**71**:1–3.

Li J, Xie Z, Wang M *et al.* Identification and Analysis of the paulomycin biosynthetic gene cluster and titer improvement of the paulomycins in *Streptomyces paulus* NRRL 8115. *PLoS One* 2015;**10**:e0120542.

Li X, Zvanych R, Torchia J *et al.* Structures and biosynthesis of 12-membered macrocyclic depsipeptides from *Streptomyces* sp. ML55. *Bioorg Med Chem Lett* 2013a;**23**:4150–3.

Li X, Zvanych R, Vanner SA *et al.* Chemical variation from the neoantimycin depsipeptide assembly line. *Bioorg Med Chem Lett* 2013b;**23**:5123–7.

Liu GY, Essex A, Buchanan JT *et al.* *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J Exp Med* 2005;**202**:209–15.

Liu J, Zhu X, Kim SJ *et al.* Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivities. *Nat Prod Rep* 2016;**33**:1146–65.

Lukezic T, Lesnik U, Podgorsek A *et al.* Identification of the chelocardin biosynthetic gene cluster from *Amycolatopsis sulphurea*: a platform for producing novel tetracycline antibiotics. *Microbiology* 2013;**159**:2524–32.

Macarron R, Banks MN, Bojanic D *et al.* Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* 2011;**10**:188–95.

Machado H, Tuttle RN, Jensen PR. Omics-based natural product discovery and the lexicon of genome mining. *Curr Opin Microbiol* 2017;**39**:136–42.

Majer J, Chater KF. *Streptomyces albus* G produces an antibiotic complex identical to paulomycins A and B. *J Gen Microbiol* 1987;**133**:2503–7.

Marshall VP, Cialdella JI, Fox JA *et al.* Precursor directed biosynthesis of paulomycins A and B. The effects of valine, isoleucine, isobutyric acid and 2-methylbutyric acid. *J Antibiot* 1984;**37**:923–5.

Medema MH, Kottman R, Yilmaz P *et al.* Minimum information about a biosynthetic cluster. *Nat Chem Biol* 2015;**11**:625–31.

Milestones in Cancer Research and Discovery https://www.cancer.gov/research/progress/250-years-milestones (March 2018, date last accessed).

Moree WJ, McConnell OJ, Nguyen DD *et al.* Microbiota of healthy corals are active against fungi in a light-dependent manner. *ACS Chem Biol* 2014;**9**:2300–8.

Moody SC, Zhao B, Lei L *et al.* Investigating conservation of the albaflavenone biosynthetic pathway and CYP170 bifunctionality in Streptomycetes. *FEBS J* 2012;**279**:1640–9.

Nah H-J, Pyeon H-R, Kang S-H *et al.* Cloning and heterologous expression of a Large-sized natural product biosynthetic gene cluster in streptomyces species. *Front Microbiol* 2017;**8**:394.

Nakano C, Kudo F, Eguchi T *et al.* Genome mining reveals two novel bacterial sesquiterpene cyclases: (-)-Germacradien-4-ol and (-)-epi-a-Bisabolol synthases from streptomyces citricolor. *Chembiochem* 2011;**12**:2271–5.

Nakano C, Tezuka T, Horinouchi S *et al.* Identification of the SGR6065 gene product as a sesquiterpene cyclase involved in (+)-epicubenol biosynthesis in *Streptomyces griseus*. *J Antibiot* 2012;**65**:551–8.

National Cancer Act, 1971. https://www.cancer.gov/about-nci/legislative/history/national-cancer-act-1971 (March 2018, date last accessed).

Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. *J Nat Prod* 2007;**70**:461–77.

Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* 2016;**79**:629–61.

Newsome AG, Culver CA, van Breemen RB. Nature's palette: the search for natural blue colorants. *J Agric Food Chem* 2014;**62**:6498–511.

Novakova R, Bistakova J, Kormanec J. Characterization of the polyketide spore pigment cluster whiESa in *Streptomyces aureofaciens* CCM3239. *Arch Microbiol* 2004;**182**:388–95.

Novakova R, Odnogova Z, Kutas P *et al.* Identification and characterization of an indigoidine-like gene for a blue pigment biosynthesis in *Streptomyces aureofaciens* CCM 3239. *Folia Microbiol (Praha)* 2010;**55**:119–25.

Ohnishi Y, Ishikawa J, Hara H *et al.* Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* 2008;**190**:4050–60.

Olano C, García I, González A *et al.* Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microbial Biotech* 2014;**7**:242–56.

Ōmura S. A splendid gift from the Earth: the origins and impact of the avermectins (Nobel Lecture). *Angew Chem Int Ed* 2016;**55**:1521–3773.

Pan G, Xu Z, Guo Z *et al*. Discovery of the leinamycin family of natural products by mining actinobacterial genomes. *Proc Natl Acad Sci USA* 2017;**114**:E11131–40.

Parte AC. LPSN — list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* 2013;**42**:D613–6.

Payne DJ, Gwynn MN, Holmes DJ *et al*. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* 2007;**6**:29–40.

Pickens LB, Kim W, Wang P *et al*. Biochemical analysis of the biosynthetic pathway of an anticancer tetracycline SF2575. *J Am Chem Soc* 2009;**131**:17677–89.

Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood trees for large alignments. *PLoS One* 2010;**5**:e9490.

Quezada M, Licona-Cassani C, Cruz-Morales P *et al*. Diverse Cone-Snail species harbor closely related streptomyces species with conserved chemical and genetic profiles, including polycyclic tetramic acid macrolactams. *Front Microbiol* 2017;**8**:2305.

Rong X, Huang Y. Taxonomic evaluation of the *Streptomyces griseus* clade using multilocus sequence analysis and DNA–DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. *Int J Syst Evol Microbiol* 2010;**60**:696–703.

Rong X, Gou Y, Huang Y. Proposal to reclassify the *Streptomyces albidoflavus* clade on the basis of multi-locus sequence analysis and DNA-DNA hybridization, and taxonomic elucidation of *Streptomyces griseus* subsp. *solvifaciens*. *Syst Appl Microbiol* 2009;**32**:314–22.

Sadeghi A, Soltani BM, Nekouei MK *et al*. Diversity of the ectoines biosynthesis genes in the salt tolerant Streptomyces and evidence for inductive effect of ectoines on their accumulation. *Microbiol Res* 2014;**169**:699–708.

Sandy M, Riu Z, Gallagher J *et al*. Enzymatic synthesis of dilactone scaffold of antimycins. *ACS Chem Biol* 2012;**7**:1956–61.

Schäberle TF, Lohr F, Schmitz A *et al*. Antibiotics from myxobacteria. *Nat Prod Rep* 2014;**31**:953–20.

Seipke RF. Strain-level diversity of secondary metabolism in *Streptomyces albus*. *PLoS One* 2015;**10**:e0116457.

Seipke RF, Barke J, Brearley C *et al*. A single *Streptomyces* symbiont makes multiple antifungals to support the fungus farming ant *Acromyrmex octospinosus*. *PLoS One* 2011;**6**:e22028.

Singh SB, Young K, Silver LL. What is an "ideal" antibiotic? Discovery challenges and path forward. *Biochem Pharmacol* 2017;**133**:63–73.

Skinnider MA, Dejong CA, Rees PN *et al*. Genomes to natural products prediction informatics for secondary metabolomes (PRISM). *Nucleic Acids Res* 2015;**43**:9645–62.

Skinnider MA, Johnston CW, Edgar RE *et al*. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc Natl Acad Sci USA* 2016;**113**:E6343–51.

Skinnider MA, Merwin NJ, Johnston CW *et al*. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* 2017;**45**:W49–54.

Skyrud W, Liu J, Thankachan D *et al*. Biosynthesis of the 15-membered ring depsipeptide neoantimycin. *ACS Chem Biol* 2018;**13**:1398–406.

Slimings C, Riley TV. Antibiotics and hospital-acquired *Clostridium difficile* infection: update of systematic review and meta-analysis. *J Antimicrob Chemother* 2014;**69**:881–91.

Spock B. *Baby and Child Care* (2nd ed.). New York, NY: Pocket Books, 1957.

Stanier RY. Agar-decomposing strains of the *Actinomyces coelicolor* species-group. *J Bacteriol* 1942;**44**:555–70.

Strangman WK, Kwon HC, Broide D *et al*. Potent inhibitors of pro-inflammatory cytokine production produced by a marine-derived bacterium. *J Med Chem* 2009;**52**:2317–27.

Strohl WR. *Pharmacologically active agents of microbial origin, Chapter 32. In: Bull AT (ed). Microbial Diversity and Bioprospecting*. Washington, DC: ASM, 2004.

Takamatsu S, Lin X, Nara A *et al*. Characterization of a silent sesquiterpenoid biosynthetic pathway in *Streptomyces avermitilis* controlling epi-isozizaene albaflavenone biosynthesis and isolation of a new oxidized epi-isozizaene metabolite. *Microb Biotechnol* 2011;**4**:184–91.

Takano E. γ-Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol* 2006;**9**:287–94.

Takano H. The regulatory mechanism underlying light-inducible production of carotenoids in nonphototrophic bacteria. *Biosci Biotech Biochem* 2016;**80**:1264–73.

Tetzlaff CN, You Z, Cane DE *et al*. A gene cluster for biosynthesis of the sesqui-terpenoid antibiotic pentalenolactone in *Streptomyces avermitilis*. *Biochemistry* 2006;**45**:6179–86.

Ueda JY, Nagai A, Izumikawa M *et al*. A novel antimycin-like compound, JBIR-06, from Streptomyces sp. ML55. *J Antibiot*. 2008;**61**:241–4.

Vanner SA, Li X, Zvanych R *et al*. Chemical and biosynthetic evolution of the antimycin-type depsipeptides. *Mol Biosyst* 2013;**9**:2712–9.

Villax I. *Streptomyces lusitanus* and the problem of classification of the various tetracycline-producing *Streptomyces*. *Antimicrob Agents Chemother* 1963;**1962**:661–8.

Virgilio A, Hengeller C. Produzione di Tetraciclina con *Streptomyces psammoticus*. *Farmaco*, Edizione Scientifica 1960;**15**:164–74.

Waksman SA, Curtis R. The actinomyces of the soil. *Soil Sci* 1916;**1**:99–134.

Walsh CT, Fischbach MA. Natural products Version 2.0: Connecting genes to molecules. *J Am Chem Soc* 2010;**132**:2469–93.

Wang CM, Cane DE. Biochemistry and molecular genetics of the biosynthesis of the earthy odorant methylisoborneol in *Streptomyces coelicolor*. *J Am Chem Soc* 2008;**130**:8908–9.

Wang P, Kim W, Pickens LB *et al*. Heterologous expression and manipulation of three tetracycline biosynthetic pathways. *Angew Chem Int Ed* 2012;**51**:11136–40.

Ward AC, Goodfellow M. *Phylogeny and functionality: taxonomy as a roadmap to genes Chapter 28*. In: Bull AT (ed). *Microbial Diversity and Bioprospecting*. Washington, DC: ASM, 2004.

Watve MG, Tickoo R, Jog MM *et al*. How many antibiotics are produced by the genus *Streptomyces*? *Arch Microbiol* 2001;**176**:386–90.

Wick RR, Judd LM, Gorrie CL *et al*. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genomics* 2017;**3**:e000132.

Wiley PF. A new antibiotic, U-43, 120 (NSC-163500). *J Antibiot* 1976;**29**:587–9.

Wilson MR, Zha L, Balskus EP. Natural product discovery from the human microbiome. *J Biol Chem* 2017;**292**:8546–52.

World Health Organization. Antimicrobial resistance: global report on surveillance. 2014. http://www.who.int/drugresistance/documents/surveillancereport/en/.

Xu L, Wu P, Wright SJ *et al*. Bioactive polycyclic tetramate macrolactams from Lysobacter enzymogenes and their absolute configurations by theoretical ECD calculations. *J Nat Prod* 2015;**78**:1841–7.

Young VB. The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ* 2017;**356**:j831.

Zhang W, Che Q, Tan H *et al.* Marine Streptomyces sp. derived antimycin analogues suppress HeLa cells via depletion HPV E6/E7 mediated by ROS-dependent ubiquitin-proteasome system. *Sci Rep* 2017;**7**:42180.

Ziemert N, Lechner A, Wietz M *et al.* Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc Natl Acad Sci USA* 2014;**111**:E1130–9.

Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes - a review. *Nat Prod Rep* 2016;**33**:988–1005.