J Ind Microbiol Biotechnol (2014) 41:357–369
DOI 10.1007/s10295-013-1360-9

ORIGINAL ARTICLE

SIMB
Society for Industrial Microbiology and Biotechnology

# MbtH homology codes to identify gifted microbes for genome mining

**Richard H. Baltz**

Received: 21 July 2013 / Accepted: 30 September 2013 / Published online: 7 November 2013


**Abstract** Advances in DNA sequencing technologies have made it possible to sequence large numbers of microbial genomes rapidly and inexpensively. In recent years, genome sequencing initiatives have demonstrated that actinomycetes with large genomes generally have the genetic potential to produce many secondary metabolites, most of which remain cryptic. Since the numbers of new and novel pathways vary considerably among actinomycetes, and the correct assembly of secondary metabolite pathways containing type I polyketide synthase or nonribosomal peptide synthetase (NRPS) genes is costly and time consuming, it would be advantageous to have simple genetic predictors for the number and potential novelty of secondary metabolite pathways in targeted microorganisms. For secondary metabolite pathways that utilize NRPS mechanisms, the small chaperone-like proteins related to MbtH encoded by *Mycobacterium tuberculosis* offer unique probes or beacons to identify gifted microbes encoding large numbers of diverse NRPS pathways because of their unique function(s) and small size. The small size of the *mbtH*-homolog genes makes surveying large numbers of genomes straight-forward with less than ten-fold sequencing coverage. Multiple MbtH orthologs and paralogs have been coupled to generate a 24-mer multiprobe to assign numerical codes to individual MbtH homologs by BLASTp analysis. This multiprobe can be used to identify gifted microbes encoding new and novel secondary metabolites for further focused exploration by extensive DNA sequencing, pathway assembly and annotation, and expression studies in homologous or heterologous hosts.

R. H. Baltz (✉)
CognoGen Biotechnology Consulting, 7636 Andora Drive, Sarasota, FL 34238, USA
e-mail: rbaltz923@gmail.com

**Keywords** Actinomycetes · Genome mining · Gifted microbes · Glycopeptide · Lipopeptide · MbtH · Mixed PKS–NRPS · NRPS · *Streptomyces*

## Introduction

Secondary metabolites produced by actinomycetes, other eubacteria, and fungi have had an enormous impact on the discovery, development, manufacturing, and commercialization of compounds for human medicine, animal health, and plant crop protection [20, 28, 51]. However, the productivity of the natural product discovery process began to decline in the 1980s, largely because of the costly nature of the process which had already harvested the "low hanging fruit" [3, 4]. Since the numbers of microbes capable of producing particular secondary metabolites are distributed in a quasi-exponential fashion in soils sampled over the past many decades, ranging in frequency over six orders of magnitude [3, 4], a major confounding issue is the rediscovery and dereplication of known secondary metabolites that are produced in relatively high frequencies among actinomycetes. Another issue is that secondary metabolites are produced in quantities that range over several orders of magnitude, and the most abundantly produced compounds can mask the activities of the less abundant compounds. Furthermore, it has been known for decades in the pharmaceutical industry that expression of secondary metabolite pathways is media dependent [31, 79, 81], and more recent genomic studies indicate that many secondary metabolite biosynthetic gene clusters are not expressed in sufficient quantities for detection under standard fermentation conditions [12, 15, 18, 50]. Genome mining addresses all of these issues by bypassing fermentation, pathway expression, compound dereplication, and screening for desired

activities, and by focusing on the potential novelty of secondary metabolite pathways by comparative bioinformatic analyses [13]. Candidate new and novel pathways and their encoding strains become focal points for expression studies as outlined in other reviews [1, 6, 8, 9, 29, 41, 43, 54, 83, 89]. In this report, I present evidence that a combination of 24 diverse MbtH homolog sequences can be used as a multiprobe to identify gifted actinomycetes and to predict the potential novelty of cryptic NRPS biosynthetic gene clusters without extensive editing and annotation of the complete gene clusters. Thus it can be coupled with low pass DNA sequencing to prioritize individual strains, pools of strains, and metagenomic libraries to identify those that merit further more intensive study for drug discovery.

## Materials and methods

BLASTp analysis [2] was carried out on the National Center for Bioinformatic Information (NCBI) server (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The 24-mer MbtH probe was constructed by excising 60 amino acid segments from 24 MbtH protein sequences starting at position three N of the conserved NPF relative to DptG and ending at the conserved S at position 62 (see [9] for alignments of typical MbtH proteins). This region encompasses the most conserved region between MbtH homologs. The 60 amino acid segments were lined up (ligated) in silico to generate the MbtH 24-mer multiprobe.

## Results

### Gifted microbes

With the emergence of genome mining as a new approach to enhance the discovery of new secondary metabolites from microorganisms, it will become increasingly important to focus DNA sequencing resources on the most promising phylogenetic groups and individual species. That raises the question of how to identify such groups and the most metabolically gifted individuals within groups. When we think of *Homo sapiens*, it is relatively easy to identify gifted individuals. Albert Einstein and the 1992 US Olympic basketball team (the "Dream Team") are two examples, although they represent very different types of talents. Likewise, there are many types of gifted strains within the microbial world. However, only a small subset of microbes have the capability to produce secondary metabolites with potential applications in human medicine, animal health, and crop protection [4, 5, 9, 21, 88], and most of these are cultivatable cosmopolitan bacteria with relatively large genomes. Within this subset, there will be a distribution gifted, average, and not-so-gifted (metabolically challenged) strains as will be discussed in more detail below.

So how can we define gifted microorganisms? A good starting point is to examine the historical record. There are many important secondary metabolites, and their properties and uses have been reviewed [20, 28, 51]. Table 1 shows a list of 60 secondary metabolites produced by microorganisms selected because of their importance primarily to human medicine, animal health, or crop protection; a few are included that are important research tools that have aided the discovery process. Three things are apparent from examination of this list: (1) actinomycetes have provided the majority of important secondary metabolites (77 %); (2) among the actinomycetes, *Streptomyces* species account for 76 %; and (3) >60 % of the compounds are produced by NRPS, PKS, or mixed NRPS–PKS mechanisms. It has been observed from whole genome sequencing that actinomycetes with large genomes encode multiple, but variable numbers of NRPS, PKS, and mixed NRPS–PKS pathways [9, 10, 19, 21, 33, 50, 55, 57, 65, 87]. So with the advent of rapid, low-cost DNA sequencing, how can we identify the most gifted species within groups such as the actinomycetes before expending substantial resources on the pathway expression, fermentation, isolation, and characterization of new molecules related to known secondary metabolites, or novel molecules that can serve as scaffolds for future medicinal chemical or combinatorial biosynthetic modification for product development?

Whole genome sequencing and finishing has been used to identify and annotate many cryptic secondary metabolite gene clusters [e.g., see 21, 33, 50, 55, 57, 65, 87]. However, this approach is expensive, and not feasible for large-scale analysis of thousands to millions of microbes in a timeframe compatible with pharmaceutical and agricultural discovery and product development. What is needed is a methodology that can sort through very large numbers of microbes with inexpensive low-pass sequencing to identify the potentially gifted strains for focused discovery research.

### MbtH (DptG) homologs as beacons for gifted strains encoding multiple NRPS gene clusters

NRPS and PKS genes *per se* are not ideal as general probes to identify related and novel gene clusters because of their large size, and sequence similarities in modular domains that confound correct assemblies with low pass sequencing. Ideal probes would be small genes or gene fragments associated with NRPS or PKS clusters that can: (1) predict the number of gene clusters; (2) identify gene clusters related

**Table 1** Microbial sources of important secondary metabolites

| Secondary metabolite | Biosynthetic origin | Major use | Producing organism | Type of organism |
|---|---|---|---|---|
| Acarbose | Glycoside | Antidiabetic (HM) | *Actinoplanes* sp. | Actinomycete |
| Actinomycin | NRPS | Antitumor (HM) | *Streptomyces anulatus* | Actinomycete |
| Adriamycin | PKS II | Antitumor (HM) | *Streptomyces peucetius* | Actinomycete |
| Amphotericin B | PKS I | Antifungal (HM) | *Streptomyces nodosus* | Actinomycete |
| Apramycin | Aminoglycoside | Research tool | *Streptoalloteicus hindustanus* | Actinomycete |
| Ascomycin | NRPS-PKS I | Immunomodulator (HM) | *Streptomyces hygroscopicus* | Actinomycete |
| Avermectin | PKS I | Anthelmintic (HM, AH) | *Streptomyces avermitilis* | Actinomycete |
| Bialaphos | NRPS | Herbicide (CP) | *Streptomyces viridochromogenes* | Actinomycete |
| Bleomycin | NRPS-PKS | Antitumor (HM) | *Streptomyces verticillus* | Actinomycete |
| Calcimycin (A23187) | PKS I | Research tool | *Streptomyces chartreusis* | Actinomycete |
| Capreomycin | NRPS | Antitubercular (HM) | *Saccharothrix mutabilis* | Actinomycete |
| Cephalosporin C | NRPS | Antibacterial (HM) | *Cephalosporium acremonium* | Fungus |
| Chloramphenicol | Shikimate modification | Antibacterial (HM) | *Streptomyces venezuelae* | Actinomycete |
| Clavulanic acid | Other | Antibacterial (HM) | *Streptomyces clavuligerus* | Actinomycete |
| Compactin | PKS | Cardiovascular (HM) | *Penicillium brevicompactum* | Fungus |
| Cyclosporin A | NRPS | Immunomodulator (HM) | *Tolypocladium inflatum* | Fungus |
| D-Cycloserine | Other | Antitubercular (HM) | *Streptomyces lavendulae* | Actinomycete |
| Daptomycin | NRPS | Antibacterial (HM) | *Streptomyces roseosporus* | Actinomycete |
| Dalbavancin | NRPS | Antibacterial (HM) | *Nonomuraea* sp. | Actinomycete |
| Daunorubicin | PKS II | Antitumor (HM) | *Streptomyces peucetius* | Actinomycete |
| Echinocandin B | NRPS | Antifungal (HM) | *Aspergillus nidulans* | Fungus |
| Epothilone | NRPS-PKS I | Anti-tumor (HM) | *Sorangium cellulosum* | Myxobacteria |
| Ergometrine | Ergoline alkaloid | Cardiovascular (HM) | *Claviceps purpurea* | Fungus |
| Erythromycin | PKS I | Antibacterial (HM) | *Saccharopolyspora erythraea* | Actinomycete |
| Fusidic acid | Terpine | Antibacterial (HM) | *Fusidium coccineum* | Fungus |
| Gentamicin | Aminoglycoside | Antibacterial (HM) | *Micromonospora purpurea* | Actinomycete |
| Hygromycin | Aminoglycoside | Research tool (AH) | *Streptomyces hygroscopicus* | Actinomycete |
| Kanamycin | Aminoglycoside | Antibacterial (HM) | *Streptomyces kanamyceticus* | Actinomycete |
| Lincomycin | Other | Antibacterial (HM) | *Streptomyces lincolnensis* | Actinomycete |
| Lipiarmycin (fidaxomicin) | PKS I | Antibacterial (HM) | *Dactylosporangium aurantiacum* | Actinomycete |
| Lipstatin | Fatty acyl-lactone | Antiobesity (HM) | *Streptomyces toxytricini* | Actinomycete |
| Lovastatin | PKS | Cardiovascular (HM) | *Aspergillus terreus* | Fungus |
| Mitomycin C | Quinone | Antitumor (HM) | *Streptomyces lavendulae* | Actinomycete |
| Monensin | PKS I | Coccidiostat (AH) | *Streptomyces cinnamonensis* | Actinomycete |
| Mycophenolic acid | PKS | Immunomodulator (HM) | *Penicillium brevicompactum* | Fungus |
| Narasin | PKS I | Coccidiostat (AH) | *Streptomyces aureofaciens* | Actinomycete |
| Nisin | Ribosomal peptide | Food preservative | *Lactococcus lactis* | Lactobacillales |
| Nystatin | PKS I | Antifungal (HM) | *Streptomyces noursei* | Actinomycete |
| Oxytetracycline | PKS II | Antibacterial (HM) | *Streptomyces rimosus* | Actinomycete |
| Paclitaxel | Isoprenoid | Antitumor (HM) | Several endophytic fungi | Fungus |
| Penicillin | NRPS | Antibacterial (HM) | *Penicillium chrysogenum* | Fungus |
| Phosphomycin | Phosphone | Antibacterial (HM) | *Streptomyces wedmorensis* | Actinomycete |
| Pneumocandin | NRPS | Antifungal (HM) | *Glarea lozoyensis* | Fungus |
| Polymyxin B and E | NRPS | Antibacterial (HM) | *Paenibacillus polymyxa* | Bacillales |
| Pristinamycin IA | NRPS | Antibacterial (HM) | *Streptomyces pristinaespiralis* | Actinomycete |
| Pristinamycin IIA | NRPS-PKS I | Antibacterial (HM) | *Streptomyces pristinaespiralis* | Actinomycete |
| Rapamycin | NRPS-PKS I | Immunomodulator (HM) | *Streptomyces hygroscopicus* | Actinomycete |
| Rifamycin | PKS I | Anti-tubercular (HM) | *Amycolatopsis mediterranei* | Actinomycete |

**Table 1** continued

| Secondary metabolite | Biosynthetic origin | Major use | Producing organism | Type of organism |
|---|---|---|---|---|
| Sinefungin | Nucleoside | Research tool | *Streptomyces griseolus* | Actinomycete |
| Spinosyns | PKS I | Insecticidal (CP) | *Saccharopolyspora spinosa* | Actinomycete |
| Staurosporine (aglycone) | Alkaloid | (Antitumor) | *Streptomyces staurosporeus* | Actinomycete |
| Streptomycin | Aminoglycoside | Anti-tubercular (HM) | *Streptomyces griseus* | Actinomycete |
| Streptozotocin | Glucosamine-nitrosourea | Antitumor | *Streptomyces achromogenes* | Actinomycete |
| Tacrolimus (FK-506) | NRPS-PKS I | Immunomodulator (HM) | *Streptomyces tsukubaensis* | Actinomycete |
| Teicoplanin | NRPS | Antibacterial (HM) | *Actinoplanes teichomyceticus* | Actinomycete |
| Tetracycline | PKS II | Antibacterial (HM) | *Streptomyces rimosus* | Actinomycete |
| Thienamycin | Other | Antibacterial (HM) | *Streptomyces cattleya* | Actinomycete |
| Tobramycin | Aminoglycoside | Antibacterial (HM) | *Streptoalloteicus hindustanus* | Actinomycete |
| Tunicamycin | Nucleoside | Research tool | *Streptomyces chartreusis* | Actinomycete |
| Tylosin | PKS I | Antibacterial (AH) | *Streptomyces fradiae* | Actinomycete |

*HM* human medicine, *AH* animal health, *CP* crop protection

to known pathways; and (3) predict novel secondary metabolic pathways. For NRPS pathways, the *mbtH* superfamily may provide the solution [9]. The *mbtH* gene, the founding member of a large superfamily, is located in the mycobactin biosynthetic gene cluster in *M. tuberculosis*, and encodes a small protein of 70 amino acids [63]. The *dptG* gene is a *mbtH* homolog found in the daptomycin biosynthetic gene cluster [47], and it was used to analyze the prevalence of *mbtH* (*dptG*) homologs in microbial groups [9]. MbtH proteins have been shown to facilitate adenylation reactions [9, 19, 30, 86], and single *mbtH*(*dptG*)-like genes are associated with most of the important functional NRPS pathways, including those for vancomycin, teicoplanin, dalbavancin, daptomycin, pristinamycin, and capreomycin [9]. *mbtH* homologs are not observed in all pathways that use NRPS mechanisms; for example, they are missing in beta-lactam and rubradirin biosynthetic gene clusters. In the latter case, the MbtH-like function is provided by a different type of protein fused upstream of the A-PCP bi-domain in RubC1 [14].

MbtH homologs have the important feature of having orthologous functions within similar biosynthetic pathways (e.g., the glycopeptides vancomycin, balhimycin, teicoplanin, and dalbavancin), but paralogous functions in dissimilar pathways [9]. This is not surprising, since MbtH homologs interact with different (paralogous) adenylation domains in different NRPS pathways. The small size of *mbtH* paralogs (generally 186–240 bp), makes incorrect assembly by low-pass sequencing highly unlikely.

## Correlation between *mbtH* (*dptG*) homologs and NRPS gene clusters in actinomycetes

Because of the evolutionary relatedness of *mbtH* orthologs and paralogs, the number of *mbtH*-like genes per microbial genome can be accurately counted by BLASTp analysis [9]. Figure 1 shows the relationship between MbtH (DptG) hits and annotated NRPS gene clusters in several finished actinomycete genomes. Although the correlation is not strictly 1:1, it is close enough to use the number of MbtH hits as a surrogate to rank strains for NRPS gene cluster abundance.

## Distribution of *mbtH* (*dptG*) homologs in actinomycetes and other microorganisms with different genome sizes

Microbes with small genomes generally have little capacity to produce complex secondary metabolites by NRPS or PKS mechanisms, and they lack significant numbers of NRPS and PKS genes [5, 21]. Furthermore, they lack meaningful numbers of *mbtH* homologs [9]. Among these, the *Archaea* are devoid of *mbtH* homologs and NRPS genes. Among the bacterial groups, the Bacteroidetes/Chlorobi and Firmicutes are nearly devoid of *mbtH* (*dptG*) homologs. The exception among the Firmicutes is the Bacillales, where *mbtH* homologs are found primarily in *Bacillus* species at a prevalence of 0.4 per fully sequenced strain. However, no *mbtH* genes were observed in the sequenced genomes of Clostridia, Lactobacillales or Molecutes. Within the Proteobacteria, *mbtH* genes are most prevalent within the beta and gamma groups at prevalences of 0.6 and 0.4 per sequenced genome, primarily in cosmopolitan *Burkholderia* and *Pseudomonas* species, respectively. *mbtH* genes are most prevalent within the Actinobacteria (~1 per sequenced genome), where the highest prevalence is within the streptomycetes (~3 per sequenced genome) [9].

Figure 2 shows the distribution of *mbtH*(*dptG*) homologs among actinomycetes with different genome sizes, ranging from 3 to 11 MB [9]. There is a rough

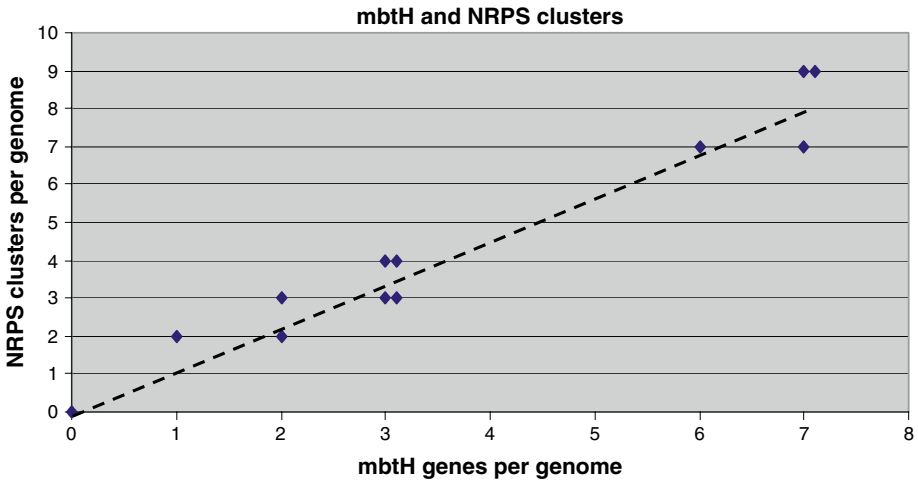


**Fig. 1** Correlation between the numbers of *mbtH* homologs and NRPS gene clusters in actinomycetes. The number of *mbtH* homologs was determined by BLASTp analysis and shown in () for the complete, fully annotated genomes of *Streptosporangium roseum* (7), *Actinosynnema mirum* (7), *Streptomyces griseus* (7), *Amycolatopsis mediterranei* (6), *Saccharopolyspora erythraea* (3), *Salinispora arenicola* (3) *Streptomyces avermitilis* (3), *Streptomyces scabiei* (3), *Streptomyces coelicolor* (2), *Saccharomonospora viridis* (2), *Micromonospora aurantiaca* (1), and *Thermobispora bispora* (0). Data from [9]

correlation between the number of *mbtH* homologs and genome size, but the degree of scatter indicates a wide disparity between different strains, including those with the largest genomes. For example, although the streptomycetes have genomes ranging from 6.6 to 11 MB, and averaging ~3 *mbtH* homologs per strain, the number of *mbtH* homologs range from 0 (*Streptomyces sviceus*; genome size = 9.1 MB) to 7 (*Streptomyces griseus*; genome size = 8.5 MB). It is clear that some actinomycetes are gifted (6–7 *mbtH* homologs), many are average (2–4 *mbtH* homologs), and many are not-so-gifted (0–1 *mbtH* homologs). Having a large genome is important, but not sufficient to achieve gifted status.

## Orthologs, paralogs, and MbtH homology codes

It has been shown that MbtH homologs within the glycopeptide pathways for vancomycin, balhimycin, teicoplanin, and dalbavancin, which are produced by diverse actinomycete genera, show 77–88 % amino acid identities, similar to the level of amino acid identities for conserved orthologous primary metabolic proteins [9]. Furthermore, analysis of the ratios of non-synonymous amino acid substitutions to synonymous substitutions ($dN/dS = 0.5$) indicated that these proteins are orthologs [9]. In contrast, other MbtH homologs encoded by unrelated NRPS pathways showed lower amino acid identities and $dN/dS$ ratios of ~ 1.0, indicating that they are paralogs [9]. This dichotomy provides predictive probes for BLASTp (and BLASTn) analysis to sort NRPS clusters into: (1) known characterized pathways; (2) pathways related to known characterized pathways; (3) known but uncharacterized pathways; and (4) unknown, novel pathways. Since the MbtH proteins are small (usually 65–75 amino acids), and have a core 60 amino acids that show the most conservation across orthologs and paralogs, multiple diverse MbtH 60-mers can be ligated (electronically) to generate a multiprobe for BLASTp analysis of individual MbtH homologs, whole genomes, or pools of whole genome sequences. In principle, each type of MbtH homolog should give a different pattern of homologies to the individual MbtH homologs in the multiprobe. To test this concept, 24 individual MbtH 60-mers were fused (see Table 2), and the multiprobe was used to carryout BLASTp analyses. Figure 3 shows an example of the BLASTp readout for *S. griseus*. The degree of sequence similarity is expressed by color (pink > green > blue > black) which can be converted into numerical codes (3 > 2 > 1 > 0).

## MbtH codes for MbtH homologs from characterized secondary metabolite pathways

Table 3 shows the MbtH codes generated by BLASTp with the MbtH multiprobe against MbtH homologs encoded by well-characterized secondary metabolite biosynthetic gene clusters that employ NRPS mechanisms for assembly.

It is noteworthy that MbtH codes for highly related biosynthetic pathways are identical or nearly identical. For instance, the related glycopeptides vancomycin, balhimycin, and dalbavancin have identical codes that differ only at position 7 from another related glycopeptide (teicoplanin), even though these glycopeptides are produced by species from three different actinomycete genera. The

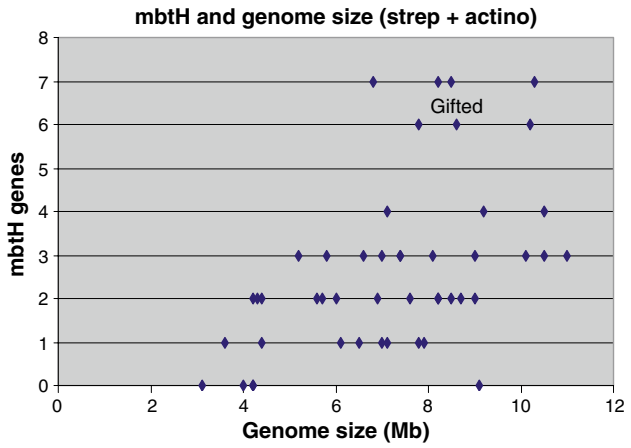


**Fig. 2** Correlation between the numbers of *mbtH* homologs and genome size in actinomycetes. The *Streptomyces* (*S.*) strains and other actinomycetes with (genome size; MbtH homologs) are: *Aeromicrobium marinum* (3.1; 0), *Thermobifida fusca* (3.6; 1), *Intrasporangium calvum* (4; 0), *Thermobispora bispora* (4.2; 0), *Tsukamurella paurometabola* (4.2; 2), *Janibacter* sp. HTCC2649 (4.2; 0), *Saccharomonospora viridis* (4.3; 2), *Mycobacterium tuberculosis* (4.4; 2), *Arthrobacter chlorophenolicus* (4.4; 1), *Salinispora tropica* (5.2; 3), *Thermomonospora curvata* (5.6; 2), *Nocardiopsis dassonvillei* (5.7; 2), *Salinispora arenicola* (5.8; 3), *Nocardia farcinica* (6; 2), *Nakamurella multipartita* (6.1; 1), *Rhodococcus erythropolis* (6.5; 1), *S. albus* (6.6; 3), *S. clavuligerus* (6.8 [+1.8 linear plasmid]; 7), SPB78 (6.9; 2), *Micromonospora aurantiaca* (7; 1), *Mycobacterium smegmatis* (7; 3), *S.* sp. MG1 (7.1; 1), *S.* sp. E14 (7; 4), *S. griseoflavus* (7.4; 3), *S.* sp. ACTE (7.4; 3), *S. pristinaespiralis* (7.6; 2), *S. roseosporus* (7.8; 6), *Rhodococcus jostii* (7.8; 1), *S.* sp. C (7.9; 1), *Saccharopolyspora erythraea* (8.1; 3), *Actinosynnema mirum* (8.2; 7), *S. lividans* (8.2; 2), *S. ghanaensis* (8.2; 2), *S. griseus* (8.5; 7), *S. viridochromogenes* (8.5; 2), *S.* sp. Act-1 (8.6; 6), *S. coelicolor* (8.7; 2), *S. avermitilis* (9; 3), *Frankia* sp EAN1pec (9; 2), *S. svicius* (9.1; 0), *S* sp. AA4 (9.2; 4), *S. scabiei* (10.1; 3), *Amycolatopsis mediterranei* (10.2; 6), *Streptosporangium roseum* (10.3; 7), *Catenulispora acidiphila* (10.5; 3), *S. hygroscopicus* (10.5; 4), and *S. violaceusniger* (11; 3). Data from [9]

quinomycins from three different *Streptomyces* species have identical codes that differ in only two positions (9 and 10) from the related thiocoraline from a *Micromonospora* species. The lipopeptides daptomycin, CDA, A54145, and friulimicin, which have ten-membered rings with similar stereochemistry, but differ substantially in amino acid sequence, have related codes, but the structurally related laspartomycin has a totally different code, suggesting that its MbtH homolog may have a very different function in the producing strain.

The uridylpeptides pacidamycin, sansanmycin, and napsamycin have very distinct and highly related codes, as do those for bleomycin and the structurally related tallysomycin and zorbamycin. The *mbtH* genes from the bleomycin-like pathways are substantially larger than average, and will be discussed in more detail below. In contrast, although capreomycin and viomycin have similar structures [23, 71], their MbtH codes differ substantially. This may be a reflection of the mechanistically distinct NRPSs used for their assembly [24]. The capreomycin MbtH code is closely related to that of the bleomycin family, but the capreomycin MbtH protein is rather small (62 amino acids). Other codes for MbtH homologs associated with other secondary metabolic pathways for which additional family members are not available are also shown in Table 3. The code least related to others is that for saframycin, demonstrating the wide divergence in codes.

## MbtH codes for unknown secondary metabolite pathways from sequenced genomes

Having defined a substantial number of MbtH codes for known secondary metabolic pathways, the MbtH multiprobe can also be used to survey the number, diversity, and potential novelty of MbtH homologs in sequenced genomes. It can also be used to screen for new related family members of known secondary metabolite pathways. Table 4 shows the MbtH codes generated for six potentially gifted actinomycetes that have six or seven MbtH homologs. There is substantial diversity in MbtH codes within this group, and only six of the 39 MbtH homologs have been assigned to pathways so far. The best annotated strain is *Streptomyces roseosporus*, where three of the MbtH homologs have been assigned to daptomycin, arylomycin, and pacidamycin biosynthetic pathways.

## MbtH homologs from the bleomycin and related pathways

Bleomycin, tallysomycin, and zorbamycin are related antitumor agents produced by *Streptomyces verticillus*, *Streptoalloteichus hindustanus*, and *Streptomyces flavoviridis*, respectively [22, 26, 27, 70]. The MbtH multiprobe analysis indicated that they have functionally similar MbtH homologs (Table 3). However, the *mbtH* homolog in the bleomycin pathway (*blm-orf13*) encodes a protein of 187 amino acids, which is over twice the size of typical MbtH homologs (65–75 amino acids). The N-terminal region of *blm-orf13* contains the typical *mbtH* conserved sequences [9], whereas the C-terminal region contains no sequence motifs related to other proteins in GenBank by BLASTp analysis. When Blm-orf13 was used as a probe in BLASTp analysis, five hits to large MbtH homologs (187–203 amino acids) were obtained (Table 5). Two of the hits came form the tallysomycin and zorbamycin biosynthetic gene clusters, and the others were from *Streptomyces mobaraensis*, *Mycobacterium abscessus*, and *Actinosynnema mirum*. The bleomycin, tallysomycin, and zorbamycin pathways have several

**Table 2** MbtH homologs comprising MbtH multi-probe

| Probe position | MbtH homolog | Microorganism | Pathway | Reference |
|---|---|---|---|---|
| 1 | AAX31560 (DptG) | *Streptomyces roseosporus* | Daptomycin | [47] |
| 2 | BAH04160 (TrsH) | *Streptomyces triostinicus* | Triosin A | [60] |
| 3 | YP_001822424 | *Streptomyces griseus* | Unknown | [55] |
| 4 | AAL90876 | *Amycolatopsis orientalis* | Vancomycin | [82] |
| 5 | CAE53354 | *Actinoplanes teichomyceticus* | Teicoplanin | [67] |
| 6 | ABD65966 | *Streptomyces fungicidicus* | Enduracidin | [80] |
| 7 | CBH31049 (MbtY) | *Streptomyces pristinaespiralis* | Pristinamycin | [45, 46] |
| 8 | AAP92504 (VioN) | *Streptomyces vinaceus* | Viomycin | [71] |
| 9 | ABR67757 (CmnN) | *Saccharothrix mutabilis* | Capreomycin | [23] |
| 10 | AAU34213 (MppT) | *Streptomyces hygroscopicus* | Mannopeptimycin | [44] |
| 11 | AAN65223 (CloY) | *Streptomyces roseochromogenes* | Clorobiocin | [59] |
| 12 | AAG29779 (CouY) | *Streptomyces rishiriensis* | Coumermycin | [76] |
| 13 | ADN26355 (AcmR) | *Streptomyces chrysomallus* | Actinomycin D | [35] |
| 14 | ADN26246 (PacJ) | *Streptomyces coeruleorubidus* | Pacidamycin | [85, 86] |
| 15 | ZP_04709435 | *Streptomyces roseosporus* | Pacidamycin | [64, 85] |
| 16 | CBA63677 | *Streptomyces sp.* ATCC700974 | Griseobactin | [58] |
| 17 | ZP_09402445 | *Streptomyces sp.* W007 | Griseobactin | [61], This report |
| 18 | AAT09800 (NocI) | *Nocardia uniformis* | Nocardicin | [19] |
| 19 | CAC11137 (NikP1) | *Streptomyces tendae* | Nikkomycin | [16, 37] |
| 20 | CBA11570 | *Streptomyces lydicus* | Streptolydigin | [30, 56] |
| 21 | MbtH | *Mycobacterium tuberculosis* H37Rv | Mycobactin | [63] |
| 22 | SCO0489 | *Streptomyces coelicolor* | Coelichelin | [39] |
| 23 | YP_00824102 | *Streptomyces griseus* | Unknown | [55] |
| 24 | YP_00626397 | *Actinoplanes sp.* SE50/110 | Unknown | [65] |

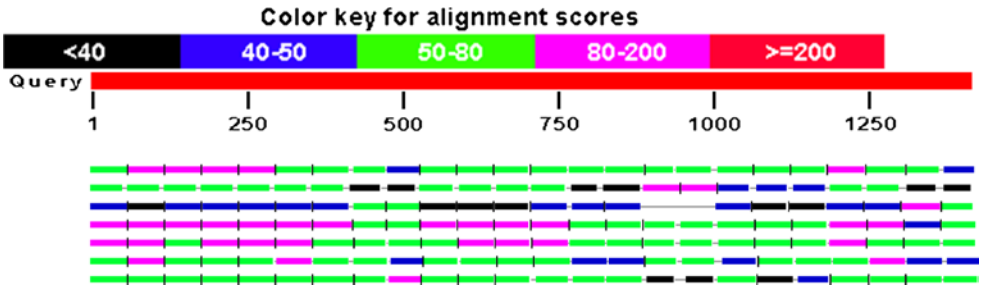


**Fig. 3** MbtH homology codes for *S. griseus*

conserved genes, including *blmX*, *tlmX*, and *zbmX*, which encode dimodular (CAT–CAT) NRPSs [27]. When BlmX (2140 amino acids) was used as a query for BLASTp analysis, the five top hits were to proteins ranging is size from 2109 to 2140 amino acids from the same five actinomycetes identified using Blm-orf13 (Table 5). These data demonstrate that an initial screen by MbtH multiprobe BLASTp followed by BLASTp with a specific MbtH homolog and a secondary metabolite pathway-specific protein can be a productive approach to predict the existence of related pathways, and suggest that *S. mobaraensis*, *M. abscessus*, and *A. mirum* may encode bleomycin-like antitumor agents.

## Discussion

Janus, the ancient Roman god of beginnings and transitions, had two faces, one to look to the past and one to the future. Natural products discovery is in a transition that can proceed in many different directions, many of which will be non-productive if the wrong organisms are chosen for focused attention and/or if non-robust discovery methods are employed. When we look to the past, it is clear that the most productive source of medically and commercially important secondary metabolites has been the actinomycetes, most notably the *Streptomyces* species. The producers of commercially important secondary metabolites

**Table 3** MbtH codes for known secondary metabolite biosynthetic pathways

| Microorganism | Pathway type | Pathway | MbtH homolog | MbtH barcode | Reference |
|---|---|---|---|---|---|
| *Amycolatopsis orientalis* | Glycopeptide | Vancomycin | AAL90876 | 333,333,322,333,322,222,223,322 | [82] |
| *Amycolatopsis balhimycina* | | Balhimycin | CAC48363 | 333,333,322,333,322,222,223,322 | [68] |
| *Nonomuraea* sp. ATTC39727 | | Dalbavancin | CAD91210 | 333,333,322,333,322,222,223,322 | [66] |
| *Actinoplanes teichomyceticus* | | Teicoplanin | CAE53354 | 333,333,222,333,322,222,223,322 | [67] |
| *Actinoplanes teichomyceticus* | | Teicoplanin | CAE53358 | 333,333,222,333,322,222,223,322 | [67] |
| *Streptomyces griseovariabilis* | Quinomycin | Echinomycin | AET98904 | 333,333,222,333,222,222,223,311 | [84] |
| *Streptomyces triostinicus* | | Triosin A | BAH04260 | 333,333,222,333,222,222,223,311 | [60] |
| *Streptomyces sp.* SNA15896 | | SW-163C | BAI63290 | 333,333,222,333,222,222,223,311 | Unpublished |
| *Micromonospora sp.* ML1 | | Thiocoraline | CAJ34376 | 333,333,221,233,222,222,223,311 | [42] |
| *Streptomyces roseosporus* | Cyclic lipopeptide | Daptomycin | AAX31560 | 332,333,322,333,322,222,223,322 | [47] |
| *Saccharomonospora viridis* | | (Daptomycin-like) | YP_003133693 | 332,333,332,333,322,122,223,322 | [7] |
| *Streptomyces coelicolor* | | CDA | AAD18046 | 332,333,332,333,322,222,223,312 | [32, 39] |
| *Streptomyces fradiae* | | A54145 | AAZ23079 | 332,333,222,333,322,222,223,312 | [48] |
| *Actinoplanes friuliensis* | | Friulimicin | CAM56772 | 332,323,322,333,322,222,223,312 | [49] |
| *Streptomyces viridochromogenes* | | Laspartomycin | AEL88630 | 212,211,122,111,122,002,011,132 | [75] |
| *Streptomyces roseochromogenes* | Aminocoumarin | Clorobiocin | AAN65223 | 332,333,322,333,322,222,223,312 | [59] |
| *Streptomyces rishiriensis* | | Coumermycin | AAG29779 | 332,333,222,333,322,222,223,312 | [76] |
| *Streptomyces coeruleorubidus* | Uridylpeptide | Pacidamycin | ADN26246 | 222,222,222,222,233,002,112,222 | [64, 85, 86] |
| *Streptomyces roseosporus* | | (Pacidamycin) | ZP_04709435 | 222,222,222,122,233,002,112,222 | [85] |
| *Streptomyces sp.* SS | | Sansanmycin | AGG82464 | 222,222,222,222,233,002,112,222 | [74] |
| *Streptomyces sp.* DSM 5940 | | Napsamycin | ADY76676 | 222,222,222,222,233,002,112,222 | [34] |
| *Nocardia uniformis* | Beta-lactam | Nocardicin A | ATT09800 | 222,222,222,222,222,113,212,222 | [19] |
| *Actinosynnema mirum* | | Nocardicin A | YP_003102272 | 222,222,222,222,222,113,212,222 | [19, 36] |
| *Streptomyces sp.* ATCC700974 | Catechol-peptide | Griseobactin | CBA63677 | 222,222,200,222,200,331,112,200 | [58] |
| *Streptomyces sp.* W007 | | Griseobactin | ZP_09404220 | 222,222,200,222,200,331,122,200 | [61], This report |
| *Streptomyces griseus griseus* | | Griseobactin | YP_00828250 | 222,222,200,222,200,331,112,200 | [53] |
| *Streptomyces griseus* XvlebKG-1 | | Griseobactin | ZP_08240466 | 222,222,200,222,200,331,122,200 | [53, 58] |
| *Streptomyces globisporus* | | Griseobactin | ZP_11380315 | 222,222,200,222,200,331,112,200 | [73], This report |
| *Streptomyces tendae* | Peptidyl nucleoside | Nikkomycin | CAC11137 | 222,222,101,222,211,112,322,201 | [16, 37, 38] |
| *Streptomyces ansochromogenes* | | Nikkomycin | AAO73548 | 222,222,101,222,211,112,322,201 | [72] |
| *Streptomyces vinaceus* | Cyclic peptide | Viomycin | AAP92504 | 222,222,133,122,322,222,223,312 | [71] |
| *Saccharothrix mutabilis* | | Capreomycin | ABR67757 | 221,222,233,222,222,002,112,223 | [23] |
| *Streptomyces verticillus* | Mixed PKS-glycopeptide | Bleomycin | AAG02368 | 221,222,223,222,222,002,112,223 | [22] |
| *Streptoalloteichus hindustanus* | | Tallysomycin | ABL74949 | 222,222,223,222,222,002,112,222 | [70] |
| *Streptomyces flavoviridis* | | Zorbamycin | ACG60748 | 222,222,233,222,222,003,312,222 | [26] |
| *Streptomyces mobaraensis* | | Bleomycin-like | ZP_23079825 | 221,222,223,222,222,002,112,223 | [78], This report |
| *Actinosynnema mirum* | | Bleomycin-like | YP_003101402 | 222,222,233,222,222,002,122,223 | [36], This report |

**Table 3** continued

| Microorganism | Pathway type | Pathway | MbtH homolog | MbtH barcode | Reference |
|---|---|---|---|---|---|
| *Mycobacterium abscessus* | | Bleomycin-like | ZP_14236827 | 222,222,222,222,222,002,112,222 | [17], This report |
| *Streptomyces roseosporus* | Cyclic lipoglycopeptide | Arylomycin | ZP_04712039 | 333,333,222,233,322,222,223,312 | [41] |
| *Streptomyces coelicolor* | Peptide siderophore | Coelichelin | Sco0489 | 333,333,332,333,332,222,223,312 | [39] |
| *Streptomyces flaveolus* | Mixed PKS-NRPS | Sanglifehrin | ACY06300 | 333,333,222,333,322,222,223,312 | [62] |
| *Streptomyces fungicidicus* | Cyclic lipopeptide | Enduracidin | ABD65966 | 332,333,222,333,322,222,223,312 | [80] |
| *Streptomyces virginiae* | Cyclic peptide | Virginiamycin S | ABD65966 | 332,333,222,333,322,222,223,312 | [52] |
| *Streptomyces hygroscopicus* | Cyclic lipoglycopeptide | Mannopeptimycin | AAU34213 | 332,333,322,333,321,222,223,311 | [44] |
| *Streptomyces collinus* | Mixed PKS-NRPS | Kirromycin | CAN89660 | 323,233,222,222,222,222,222,212 | [77] |
| *Streptomyces chrysomallus* | Chromopeptide lactone | Actinomycin D | ADG27355 | 322,333,322,332,322,222,222,322 | [35] |
| *Streptomyces pristinaespiralis* | Cyclic peptide | Pristinamycin | CBH31049 | 322,322,322,332,322,222,222,322 | [45, 46] |
| *Streptomyces antibioticus* | Aminocoumarin | Simocyclinone | AAG34186 | 233,332,222,222,222,222,223,322 | [25] |
| *Mycobacterium tuberculosis* | Mixed PKS-NRPS | Mycobactin | NP_216893 (MbtH) | 222,222,222,222,222,222,223,322 | [63] |
| *Streptomyces viridochromogenes* | PT-tripeptide[a] | Bialaphos | AAU00076 | 222,222,222,222,222,002,102,222 | [11] |
| *Streptomyces lydicus* | Mixed PKS-NRPS | Streptolydigin | CBA11570 | 222,221,222,222,211,221,232,211 | [30, 56] |
| *Streptomyces lavendulae* | Tetrahydroisoquinoline | Saframycin | ABI22136 | 101,101,022,000,121,002,002,122 | [40] |

[a] Phosphenothricin-tripeptide

can be considered as gifted microbes by default, but only a fraction of actinomycetes are gifted by these retrospective criteria. When looking prospectively to the future, it would be very useful to be able to predict which microbes are gifted to optimize resource allocations. Since we cannot predict commercial successes prospectively, we need surrogate markers to predict which microbes have the highest likelihood of yielding commercially successful products. One likely predictor is the number and novelty of NRPS and PKS pathways encoded by candidate microbes. In this report I have shown that the number and variety of NRPS gene clusters can be estimated by BLASTp analysis with a diverse MbtH 24-mer multiprobe. Since *mbtH* genes can have orthologous functions in related pathways, and paralogous functions in unrelated pathways [9], the probing with a 24-mer yields relatedness information that can be translated into a 24 number code. The 24 MbtH homologs used in this study were chosen primarily from well-characterized pathways. As more NRPS pathways are characterized, additional MbtH homologs can be incorporated into new multiprobes to supplement the utility of this multiprobe.

The initial MbtH codes have been defined for many secondary metabolite pathways that employ NRPS mechanisms. It is noteworthy that MbtH codes for highly similar secondary metabolites are generally identical or highly similar, while those for unrelated pathways are dissimilar. Thus, as more NRPS gene clusters are assigned to specific products, the robustness of code differences can be used more efficiently for dereplication of known pathways, identification of important new members of known families, and prediction of novel pathways and products for future focus.

It was shown previously that as a group the actinomycetes on average have more MbtH homologs per sequenced organism than any other bacterial group, and among the actinomycetes the *Streptomyces* had the highest average (~3). None-the-less, the number of MbtH homologs ranged from 0 to 7 per individual species. Actinomycetes with small genomes tended to have 0 or 1 *mbtH* gene. If we use the number of *mbtH* homologs per cell as the initial screen, only a small number of actinomycete genomes encode 6–7 *mbtH* homologs, and can be considered as potentially gifted. The MbtH multiprobe can be used to differentiate between the individual MbtH homologs, and six potentially gifted actinomycetes were analyzed for MbtH codes in this study. It is noteworthy that three of the strains would be designated as gifted by retrospective and prospective criteria, as they are the commercial producers of daptomycin (*S. roseosporus*), rifamycin (*A. mediterranei*), and streptomycin (*S. griseus*). Of the three products, only daptomycin is

**Table 4** MbtH codes for gifted actinomycetes

| Microorganism | Genome size (MB) | MbtH homolog | MbtH homolog codes | Secondary metabolite pathway | Genome reference |
|---|---|---|---|---|---|
| *Actinosynnema mirum* | 8.2 | YP_003101336 | 323,333,322,333,322,222,223,312 | | [36] |
| | | YP_003102731 | 233,333,222,232,222,222,223,311 | | |
| | | YP_003102279 | 223,222,222,222,222,112,223,221 | | |
| | | YP_003101402 | 222,222,233,222,222,002,122,223 | Bleomycin-family | This report |
| | | YP_003102192 | 222,222,122,222,222,002,112,222 | | |
| | | YP_003101315 | 222,222,222,222,222,002,112,222 | | |
| | | YP_003102272 | 222,222,222,222,222,113,212,222 | Nocardicin | [19] |
| *Streptomyces griseus* | 8.5 | YP_00821964 | 333,333,322,333,322,222,223,312 | | [55] |
| | | YP_00824768 | 332,333,222,233,322,222,223,322 | | |
| | | YP_00822424 | 233,332,221,222,222,222,223,221 | | |
| | | YP_00822166 | 232,223,221,222,211,221,222,311 | | |
| | | YP_00822088 | 222,222,223,222,222,002,012,222 | | |
| | | YP_00828250 | 222,222,200,222,200,331,112,200 | Griseobactin | [58] |
| | | YP_00824102 | 101,110,122,000,111,001,001,132 | | |
| *Streptosporangium roseum* | 10.3 | YP_003342011 | 322,333,322,333,322,222,223,322 | | [53] |
| | | YP_003339148 | 333,333,322,333,322,222,223,322 | | |
| | | YP_003342413 | 332,333,221,233,222,222,223,212 | | |
| | | YP_003341076 | 322,333,221,232,222,222,223,312 | | |
| | | YP_003338143 | 222,222,223,222,222,002,112,222 | | |
| | | YP_003342290 | 221,222,233,222,222,002,112,223 | | |
| | | YP_003337926 | 211,222,223,122,222,002,112,122 | | |
| *Amycolatopsis mediterranei* | 10.2 | YP_003767198 | 332,233,222,332,322,222,223,222 | | [87] |
| | | YP_003768373 | 222,222,223,122,222,022,122,223 | | |
| | | YP_003766961 | 222,222,223,222,222,002,112,223 | | |
| | | YP_003766215 | 232,222,212,222,222,222,223,311 | | |
| | | YP_003765296 | 222,222,232,222,222,112,212,222 | | |
| | | YP_003766936 | 222,222,223,222,222,002,222,222 | | |
| *Saccharothrix espanaensis* | 9.36 | YP_007037427 | 332,333,311,333,322,222,223,311 | | [69] |
| | | YP_007036874 | 333,333,222,333,322,222,223,312 | | |
| | | YP_007037734 | 221,222,223,222,222,002,112,232 | | |
| | | YP_007038712 | 222,222,201,222,210,121,122,200 | | |
| | | YP_007037432 | 222,222,222,222,222,002,012,222 | | |
| | | YP_007038943 | 211,222,122,111,122,002,112,122 | | |
| *Streptomyces roseosporus* | 7.8 | ZP_04706746 | 332,333,322,333,322,222,223,322 | Daptomycin | [47] |
| | | ZP_04707179 | 333,333,322,333,322,222,223,312 | | |
| | | ZP_04713195 | 333,333,322,333,322,222,223,322 | | |
| | | ZP_04712095 | 333,333,222,233,322,222,223,312 | Arylomycin | [41] |
| | | ZP_04709435 | 222,222,222,122,233,002,112,222 | Pacidamycin | [85] |
| | | ZP_04709625 | 211,222,101,122,200,111,001,100 | | |

produced by an NRPS mechanism that employs an MbtH homolog (DptG).

As genome mining of actinomycetes progresses in the future, the MbtH multiprobe concept can be used to screen individual stains, pools of strains, environmental samples, or pools of environmental samples for the presence of gifted strains by using low pass DNA sequencing. More intense DNA sequencing, annotation, and expression studies can be focused on individual gifted strains or strain populations enriched for gifted strains. The multiprobe concept

**Table 5** Bleomycin ORF13 (large MbtH homolog) and BlmX (NRPS) homologs in actinomycetes

| Microorganism | ORF13 homologs | Amino acids (% identity) | BlmX homologs | Amino acids (% identity) |
|---|---|---|---|---|
| *Streptomyces verticillus* | AAG02368 (ORF13) | 187 (100) | AAG02355 (BlmX) | 2140 (100) |
| *Streptomyces mobaraensis* | EME99234 | 187 (99) | EMF00638 | 2120 (99) |
| *Streptoalloteichus hindustanus* | ABL74949 | 189 (65) | ABL74936 (TlmX) | 2132 (65) |
| *Mycobacterium abscessus* | EIC67554 | 192 (61) | EIC67541 | 2122 (59) |
| *Actinosynnema mirum* | ACU37556 | 203 (55) | ACU37568 | 2109 (59) |
| *Streptomyces flavoviridis* | ACG60748 | 195 (54) | ACG60782 (ZbmX) | 2140 (53) |

might also be extended to PKS pathways if appropriate, relatively small pathway predictive peptide sequences can be identified.

**Acknowledgments** This work was supported by CognoGen Biotechnology Consulting.

## References


1. Albright JC, Goering AW, Doroghazi JR, Metcalf WW, Kelleher NL (2013) Strain-specific proteogenomics accelerates discovery of natural products via their biosynthetic pathways. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1373-4
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
3. Baltz RH (2005) Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? SIM News 55:186–196
4. Baltz RH (2007) Antimicrobials from actinomycetes: back to the future. Microbe 2:125–131
5. Baltz RH (2008) Renaissance in antibacterial discovery from actinomycetes. Curr Opin Pharmacol 8:557–563
6. Baltz RH (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol 37:759–772
7. Baltz RH (2010) Genomics and the ancient origins of the daptomycin biosynthetic gene cluster. J Antibiot 63:506–511
8. Baltz RH (2011) Strain improvement in actinomycetes in the postgenomic era. J Ind Microbiol Biotechnol 38:657–666
9. Baltz RH (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. J Ind Microbiol Biotechnol 38:1747–1760
10. Bentley SD, Chater KF, Cerdeño-Tárraga AM et al (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147
11. Blodgett JA, Zhang JK, Metcalf WW (2005) Molecular cloning, sequence analysis, and heterologous expression of the phosphinothricin tripeptide biosynthetic gene cluster from *Streptomyces viridochromogenes* DSM 40736. Antimicrob Agents Chemother 49:230–240
12. Blodgett JAV, Oh D-C, Cao S, Currie CR, Kolter R, Clardy J (2010) Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. Proc Natl Acad Sci USA 107:11692–11697
13. Boddy CN (2013) Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1368-1
14. Boll B, Heide L (2013) A domain of RubC1 biosynthesis that can functionally replace MbtH-like proteins in tyrosine adenylation. ChemBioChem 14:43–44
15. Challis GL (2008) Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology 154:1555–1569
16. Chen H, Hubbard BK, O'Connor SE, Walsh CT (2002) Formation of β-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. Chem Biol 9:103–112
17. Choo SW, Wong YL, Yusoff AM, Leong ML, Wong GJ, Ong CS, Ng KP, Ngeow YF (2012) Genome sequence of the *Mycobacterium abscessus* strain 93. J Bacteriol 194:3278
18. Corre C, Challis GL (2009) New natural product biosynthetic chemistry discovered by genome mining. Nat Prod Rep 26:977–986
19. Davidsen JM, Bartley DM, Townsend CA (2013) Non-ribosomal propeptide precursor of nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein NocI. J Am Chem Soc 135:1749–1759
20. Demain AL (2013) Importance of microbial natural products and the need to revitalize their discovery. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1325-z
21. Donadio S, Monciardini P, Sosio M (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Nat Prod Rep 24:1073–1109
22. Du L, Sánchez C, Chen M, Edwards DJ, Shen B (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chem Biol 7:623–642
23. Felnagle EA, Rondon MR, Berti AD, Crosby HA, Thomas MG (2007) Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. Appl Environ Microbiol 73:4162–4170
24. Felnagle EA, Podevels AM, Barkei JJ, Thomas MG (2011) Mechanistically distinct nonribosomal peptide synthetases assemble the structurally related viomycin and capreomycin antibiotics. ChemBioChem 12:1859–1867
25. Galm U, Schimana J, Fiedler H-P, Schmidt J, Li S-M, Heide L (2002) Cloning and analysis of the simocyclinone biosynthetic gene cluster of *Streptomyces antibioticus* Tu 6040. Arch Microbiol 178:102–114
26. Galm U, Wendt-Pienkowski E, Wang L, Oh T-J, Yi F, Tao M, Coughlin JM, Shen B (2009) The biosynthetic gene cluster of zorbamycin, a member of the bleomycin family of antitumor antibiotics, from *Streptomyces flavoviridis* ATTC 21892. Mol BioSyst 5:77–90
27. Galm U, Wendt-Pienkowski E, Wang L, Huang S-X, Unsin C, Tao M, Coughlin JM, Shen B (2011) Comparative analysis of the biosynthetic gene clusters and pathways for three structurally related antitumor antibiotics: bleomycin, tallysomycin, and zorbamycin. J Nat Prod 74:526–536
28. Giddings LA, Newman DJ (2013) Microbial natural products: molecular blueprints for antitumor drugs. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1331-1

29. Gomez-Escribano JP, Bibb MJ (2013) Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1348-5
30. Herbst DA, Boll B, Zocher G, Stehle T, Heide L (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. J Biol Chem 288:1991–2003
31. Higgs RE, Zahn JA, Gygi FD, Hilton MD (2001) Rapid method to estimate the presence of secondary metabolites in microbial extracts. Appl Environ Microbiol 67:371–376
32. Hojati Z, Milne C, Harvey B, Gordon L, Borg M, Flett F, Wilkinson B, Sidebottom PJ, Rudd BA, Hayes MA, Smith CP, Micklefield J (2002) Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. Chem Biol 9:1175–1187
33. Ikeda H, Shinya K, Ōmura S (2013) Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1327-x
34. Kaysser L, Wemakor E, Sedding K, Hennig S, Siedenberg S, Gust B (2011) Identification of a napsamycin biosynthetic gene cluster by genome mining. ChemBioChem 12:477–478
35. Keller U, Lang M, Crnovcic I, Pfennig F, Schauwecker F (2010) The actinomycin biosynthetic gene cluster of *Streptomyces chrysomallus*: a genetic hall of mirrors for the synthesis of a molecule with mirror symmetry. J Bacteriol 192:2583–2595
36. Land M, Lapidus A, Mayilraj S et al (2009) Complete genome sequence of *Actinosynnema mirum* type strain (101). Stand Genomic Sci 1:46–53
37. Lauer B, Russwurm R, Bormann C (2000) Molecular characterization of two genes from *Streptomyces tendae* Tü901 required for the formation of the 4-formyl-4-imidazolin-2-one-containing nucleoside moiety of the peptidyl nucleoside antibiotic nikkomycin. Eur J Biochem 267:1698–1706
38. Lauer B, Russwurm R, Schwarz W, Kálmánczhelyi A, Bruntner C, Rosemeier A, Bormann C (2001) Molecular characterization of co-transcribed genes from *Streptomyces tendae* Tü901 involved in the biosynthesis of the peptidyl moiety and assembly of the peptidyl nucleoside antibiotic nikkomycin. Mol Gen Genet 264:662–673
39. Lautru S, Oves-Costales D, Pernodet J-L, Challis GL (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. Microbiology 153:1405–1412
40. Li L, Deng W, Song J, Ding W, Zhao Q-F, Peng C, Song W-W, Tang T-L, Liu W (2008) Characterization of the saframycin A gene cluster from *Streptomyces lavendulae* NRRL 110002 revealing a nonribosomal peptide synthetase system for assembling the unusual tetrapeptidyl skeleton in an iterative manner. J Bacteriol 190:251–263
41. Liu WT, Kerten RD, Yang YL, Moore BS, Dorrestein PC (2011) Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in *Streptomyces roseosporus*. J Am Chem Soc 133:18010–18013
42. Lombó F, Velasco A, Castro A, de la Calle F, Braña AF, Sánchez-Puelles JM, Méndez C, Salas JA (2006) Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *Streptomyces* species. ChemBioChem 7:366–376
43. Luzhetskyy A, Rebets Y, Brötz E, Tokovenko B (2013) Actinomycetes biosynthetic potential: how to bridge in silico and in vivo. J Ind Microbiol Biotechnol. doi:10-1007/s10295-013-1352-9
44. Magarvey NA, Haltli B, He M, Greenstein M, Hucul JA (2006) Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant Gram-positive pathogens. Antimicrob Agents Chemother 50:2167–2177
45. Mast Y, Weber T, Golz M, Ort-Winklbauer R, Gondran A, Wohlleben W, Schinko E (2011) Characterization of the 'pristinamycin supercluster' of *Streptomyces pristinaespiralis*. Microb Biotechnol 4:192–206
46. Mast YJ, Wohlleben W, Schinko E (2011) Identification and functional characterization of phenylglycine biosynthetic genes involved in pristinamycin biosynthesis in *Streptomyces pristinaespiralis*. J Biotechnol 155:63–67
47. Miao V, Coëffet-LeGal M-F, Brian P, Brost R, Penn J, Whiting A, Martin S, Ford R, Parr I, Bouchard M, Silva CJ, Wrigley SK, Baltz RH (2005) Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. Microbiology 151:1507–1523
48. Miao V, Brost R, Chapple J, She K, Coëffet-Le Gal M-F, Baltz RH (2006) The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. J Ind Microbiol Biotechnol 33:129–140
49. Müller C, Nolden S, Gebhardt P, Heinzelmann E, Lange C, Puk O, Welzel K, Wohlleben W, Schwartz D (2007) Sequencing and analysis of the biosynthetic gene cluster of the lipopeptide antibiotic friulimicin in *Actinoplanes friuliensis*. Antimicrob Agents Chemother 51:1028–1037
50. Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. Nat Prod Rep 26:1362–1384
51. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981–2010. J Nat Prod 75:311–335
52. Ningsih F, Kitani S, Fukushima E, Nihira T (2011) VisG is essential for biosynthesis of virginiamycin S, a streptogramin type B antibiotic, as a provider of the nonproteogenic amino acid phenylglycine. Microbiology 157:3213–3220
53. Nolan M, Sikorski J, Janko M et al (2010) Complete genome sequence of *Streptosporangium roseum* type strain (NI 9100). Stand Genome Sci 2:29–37
54. Ochi K, Tanaka Y, Tojo S (2013) Activating the expression of cryptic genes by *rpoB* mutations in RNA polymerase or by rare earth elements. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1349-4
55. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. J Bacteriol 190:4050–4060
56. Olano C, Gómez C, Pérez M, Palomino M, Pineda-Lucena A, Carbajo RJ, Braña AF, Méndez C, Salas JA (2009) Deciphering the biosynthesis of the RNA polymerase inhibitor streptolydigin and generation of glycosylated derivatives. Chem Biol 16:1031–1044
57. Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. Nat Biotechnol 25:447–453
58. Patzer SI, Braun V (2010) Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of *Streptomyces* sp. ATCC 700974. J Bacteriol 192:426–435
59. Pojer F, Li S-M, Heide L (2002) Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics. Microbiology 148:3901–3911
60. Praseuth AP, Wang CC, Watanabe K, Hotta K, Oguri H, Oikawa H (2008) Complete sequence of biosynthetic gene cluster responsible for producing triostin A and evaluation of quinomycin-type antibiotics from *Streptomyces trinostinicus*. Biotechnol Prog 24:1226–1231

61. Qin S, Zhang H, Li F, Zhu B, Zheng H (2012) Draft genome sequence of marine *Streptomyces* sp. Strain W007, which produces angucyclinone antibiotics with a benz[a]anthracene skeleton. J Bacteriol 194:1628–1629
62. Qu X, Jiang N, Xu F, Shao L, Tang G, Wilkensen B, Liu W (2011) Cloning, sequencing and characterization of the biosynthetic gene cluster of sanglifehrin A, a potent cyclophilin inhibitor. Mol BioSyst 7:852–861
63. Quadri LE, Sello J, Keating TA, Weinreb PH, Walsh CT (1998) Identification of *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem Biol 5:631–645
64. Rackham EJ, Gruschow S, Ragab AE, Dickens S, Goss RJM (2010) Pacidamycin biosynthesis: identification and heterologous expression of the first uridyl peptide antibiotic gene cluster. ChemBioChem 11:1700–1709
65. Schweintek P, Szczepanowski R, Ruckert C et al (2012) The complete genome sequence of the acarbose producer *Actinoplanes* sp. SE50/110. BMC Genomics 13:112
66. Sosio M, Stinchi S, Beltrametti F, Lazzarini A, Donadio S (2003) The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by *Nonomuraea* species. Chem Biol 10:541–549
67. Sosio M, Kloosterman H, Bianchi A, de Vreugd P, Dijkhuizen L, Donodio S (2004) Organization of the teicoplanin gene cluster in *Actinoplanes teichomyceticus*. Microbiology 150:95–102
68. Stegmann E, Rausch C, Stockert S, Burkert D, Wohlleben W (2006) The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production. FEMS Microbiol Lett 262:85–92
69. Strobel T, Al-Dilaimi A, Blom J, Gessner A, Kalinowski J, Luzhetska M, Pühler A, Szczepanowski R, Bechtold A, Rücker C (2012) Complete genome sequence of *Saccharothrix espanaensis* DSM 44229 and comparison to the other completely sequenced *Pseudonocardiaceae*. BMC Genomics 13:465
70. Tao M, Wang L, Wendt-Pienkowski E, George NP, Galm U, Zhang G, Coughlin JM, Shen B (2007) The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-94 ATTC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. Mol BioSyst 3:60–74
71. Thomas MG, Chan YA, Ozanick SG (2003) Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. Antimicrob Agents Chemother 47:2823–2830
72. Wang G, Nie L, Tan H (2003) Cloning and characterization of *sanO*, a gene involved in nikkomycin biosynthesis in *Streptomyces ansochromogenes*. Lett Appl Microbiol 37:452–457
73. Wang L, Wang S, He Q, Yu T, Li Q, Hong B (2012) Draft genome sequence of *Streptomyces globisporus* C-1027, which produces an antitumor antibiotic consisting of a nine-membered enediyne with a chromoprotein. J Bacteriol 194:4144
74. Wang L, Xie Y, Li Q, He N, Yao E, Xu H, Yu Y, Chen R, Hong B (2012) Draft genome sequence of *Streptomyces* sp. Strain SS, which produces a series of uridyl peptide antibiotic sansanmycins. J Bacteriol 194:6988–6989
75. Wang Y, Chen Y, Shen Q, Yin X (2011) Molecular cloning and identification of the laspartomycin biosynthetic gene cluster from *Streptomyces viridochromogenes*. Gene 483:11–21
76. Wang Z-X, Li S-M, Heide L (2000) Identification of the coumermycin $A_1$ biosynthetic gene cluster of *Streptomyces rishiriensis* DSM 40489. Antimicrob Agents Chemother 44:3040–3048
77. Weber T, Laiple KJ, Pross EK, Textor A, Grond S, Welzel K, Pelzer S, Vente A, Wohlleben W (2008) Molecular analysis of the kirromycin biosynthetic gene cluster revealed beta-alanine as precursor of the pyridone moiety. Chem Biol 15:175–188
78. Yang H, He T, Zhu W, Lu B, Sun W (2013) Whole-genome shotgun assembly and analysis of the genome of *Streptomyces mobaraensis* DSM 40847, a strain for the industrial production of microbial transglutaminase. Genome Announc 1:e0014313
79. Yarbrough GG, Taylor DP, Rowlands RT, Crawford MS, Lasure LL (1992) Screening microbial metabolites for new drugs—theoretical and practical issues. J Antibiot 46:535–544
80. Yin X, Zabriskie M (2006) The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. Microbiology 152:2969–2983
81. Zahn JA, Higgs RE, Hilton MD (2001) Use of direct-infusion electrospray mass spectrometry to guide empirical development of improved conditions for expression of secondary metabolites from actinomycetes. Appl Environ Microbiol 67:377–386
82. Zerbe K, Pylypenko O, Vitali F, Zhang W, Rouset S, Heck M, Vrijbloed JW, Biscoff D, Bister B, Sussmuth RD, Pelzer S, Wohlleben W, Robinson JA, Schlichting I (2002) Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. J Biol Chem 277:47476–47485
83. Zerikly M, Challis GL (2009) Strategies for the discovery of new natural products by genome mining. ChemBioChem 10:625–633
84. Zhang C, Kong L, Liu Q, Lei X, Zhu T, Yin J, Lin B, Deng Z, You D (2013) *In vitro* characterization of echinomycin biosynthesis: formation and hydroxylation of L-tryptophanyl-*S*-enzyme and oxidation of (*2S*,*3S*) β-hydroxytryptophan. PLOS ONE 8:e56772
85. Zhang W, Ostach B, Walsh CT (2010) Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. Proc Natl Acad Sci USA 107:16828–16833
86. Zhang W, Heemstra JR, Walsh CT, Imker HJ (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. Biochemistry 49:9946–9947
87. Zhao W, Zhong Y, Yuan H et al (2010) Complete genome sequence of the rifamycin SV-producing *Amycolatopsis mediterranei* U32 revealed its genetic characteristics in phylogeny and metabolism. Cell Res 20:1096–1108
88. Zhu F, Qin C, Tao L et al (2011) Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. Proc Natl Acad Sci USA 31:12943–12948
89. Zhu H, Sandiford SK, van Wezel GP (2013) Triggers and cues that activate antibiotic production by actinomycetes. J Ind Microbiol Biotechnol. doi:10.1007/s10295-013-1309-z