Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products

Gubbens, J.; Zhu, H.; Girard, G.A.O.; Song, L.; Florea, B.I.; Aston, P.; ... ; Wezel, G.P. van

Chemistry & Biology
Article

Natural Product Proteomining, a Quantitative Proteomics Platform, Allows Rapid Discovery of Biosynthetic Gene Clusters for Different Classes of Natural Products

Jacob Gubbens,[1] Hua Zhu,[2] Geneviève Girard,[2] Lijiang Song,[3] Bogdan I. Florea,[1] Philip Aston,[3] Koji Ichinose,[4] Dmitri V. Filippov,[1] Young H. Choi,[2] Herman S. Overkleeft,[1] Gregory L. Challis,[3] and Gilles P. van Wezel[2,*]
[1]Leiden Institute of Chemistry, Leiden University, 2300RA Leiden, the Netherlands
[2]Institute of Biology Leiden, Leiden University, 2300RA Leiden, the Netherlands
[3]Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK
[4]Research Institute of Pharmaceutical Sciences, Musashino University, Shinmachi, Nishitokyo-shi, Tokyo 202-8585, Japan
*Correspondence: g.wezel@biology.leidenuniv.nl
http://dx.doi.org/10.1016/j.chembiol.2014.03.011

SUMMARY

Information on gene clusters for natural product biosynthesis is accumulating rapidly because of the current boom of available genome sequencing data. However, linking a natural product to a specific gene cluster remains challenging. Here, we present a widely applicable strategy for the identification of gene clusters for specific natural products, which we name natural product proteomining. The method is based on using fluctuating growth conditions that ensure differential biosynthesis of the bioactivity of interest. Subsequent combination of metabolomics and quantitative proteomics establishes correlations between abundance of natural products and concomitant changes in the protein pool, which allows identification of the relevant biosynthetic gene cluster. We used this approach to elucidate gene clusters for different natural products in *Bacillus* and *Streptomyces*, including a novel juglomycin-type antibiotic. Natural product proteomining does not require prior knowledge of the gene cluster or secondary metabolite and therefore represents a general strategy for identification of all types of gene clusters.

INTRODUCTION

Because of the rapid spread of drug-resistant infectious diseases, a point of no return has been reached where novel antibiotics are of absolute necessity (Payne et al., 2007). However, it has become increasingly difficult to find novel antibiotics with efficacy against multidrug-resistant pathogens such as MDR-TB (multidrug-resistant *Mycobacterium tuberculosis*), MRSA (methicillin-resistant *Staphylococcus aureus*), and the rapidly emerging multidrug-resistant Gram-negative pathogens (Arias and Murray, 2009; Rice, 2008). The actinomycetes are filamentous bacteria with a complex multicellular life cycle (Claessen et al., 2014) that produce some two-thirds of all known natural antibiotics (Hopwood, 2007). Full genome sequencing established that even the widely studied species are relatively untapped sources of natural products, such as antibiotics, anticancer agents, and fungicides (Hopwood et al., 1995). For example, over 20 gene clusters for natural products biosynthesis are present on the genome of *Streptomyces coelicolor* A3(2) alone (Bentley et al., 2002). This suggests the presence of silent biosynthetic gene clusters in all actinomycetes, sequenced and unsequenced, and new approaches are required to identify and characterize these (Challis and Hopwood, 2003; Van Lanen and Shen, 2006).

Genome mining based on next generation sequencing (NGS) (Claesen and Bibb, 2010; Cruz-Morales et al., 2013; Ikeda et al., 2014; Lautru et al., 2005; Udwary et al., 2011) and metagenomics (Owen et al., 2013) are valuable strategies to aid in the discovery and identification of natural products. With the rapidly increasing number of sequenced genomes, the amount of data on biosynthetic gene clusters is overwhelming. However, although successful examples exist where natural products were identified guided by the genome sequence alone (Knappe et al., 2008; Lautru et al., 2005), this is still a challenging process. For example, many biosynthetic gene clusters of the same class can be present on a single genome, they might be spread over multiple contigs, and each of them generally contains genes with unknown function, causing ambiguity as to the predicted structure of the products. Moreover, the products produced by "cryptic" gene clusters will likely remain undetected. Also, many strains are not easily manipulated, whereas large scale heterologous expression of metagenomic libraries is challenging (Wilson and Piel, 2013). Consequently, new methods that allow conclusive linkage between a gene cluster and a natural product of interest are needed to facilitate efficient genome mining-based drug discovery.

Expression-based analysis is a useful tool to detect active biosynthetic gene clusters (Schley et al., 2006; Udwary et al., 2011). By exploiting the essential phosphopantetheine modification of polyketide synthases (PKS) and nonribosomal peptide synthases (NRPS), two proteomics-based approaches have been designed to specifically detect these proteins, either after affinity purification (Meier et al., 2009) or by detection of reporter ions by mass spectrometry (Bumpus et al., 2009). In case no genome sequence information is available, peptides of interest are sequenced de novo to design degenerate primers for PCR amplification of parts of the gene cluster. An alternative is PCR

Table 1. Expression Level Changes of Proteins and Secondary Metabolites in *S. coelicolor*

[2]Log Ratio (Mutant/WT)[a]		
DasR	*Rok7B7*	Product[b]
Protein Levels		
		desferrioxamine
2.8	*1.0*	SCO2782 (DesA)
3.0	1.9	SCO2785 (DesD)
		undecylprodigiosin
	1.4	SCO5878 (RedX)
	1.7	SCO5879 (RedW)
	1.9	SCO5888 (FabH3)
−1.8	1.9	SCO5890
	2.1	SCO5891 (RedM)
	2.1	SCO5892
−2.2	1.9	SCO5895
	2.0	SCO5896
Metabolite Levels		
1.1	2.0	desferrioxamine mycelium
3.2	4.2	desferrioxamine excreted
−3.3	2.6	prodiginin mycelium

See also Figure S1.

[a]Changes in protein levels expressed as signal intensity in *dasR* or *rok7B7* deletion mutants versus signal intensity in the parental strain *S. coelicolor* M145. For quantitative proteomics, data are the averages of two experiments, with one experiment using opposite labeling compared with the other experiment (label swap). Italic numbers indicate that the ratio could only be determined in one of the two experiments and are included only if the same protein could be quantified (detected in both experiments with same sign) for the other deletion mutant. For metabolites, ratios were determined by LC-MS analysis and peak integration.

[b]Gene/protein annotations based on StrepDB (http://strepdb.streptomyces.org.uk).

amplification of cDNA, using degenerate primers for sequences common in biosynthetic gene clusters (Qu et al., 2011). Yet another strategy utilizes sequence tags, obtained from peptides or glycosylated natural products using tandem mass spectrometry (MS) analysis to link the natural products to gene clusters in previously sequenced organisms (Kersten et al., 2011; Kersten et al., 2013). These methods target specific types of clusters or known sequences; for example lantibiotics, which are ribosomally produced peptides subsequently modified via, among other means, lanthionine-type thioether crosslinking (Willey and van der Donk, 2007), cannot be detected in this way. Therefore, a single, straightforward method to identify the biosynthetic cluster of all types of secondary metabolites is of great interest.

The production of secondary metabolites varies greatly depending on external factors such as growth and nutritional conditions (Sánchez et al., 2010; van Wezel and McDowall, 2011). In several streptomycetes the production of poorly expressed antibiotics may be induced by the addition of *N*-acetylglucosamine, which relieves repression of biosynthetic gene clusters by the global regulator DasR (Rigali et al., 2008). Utilizing the ability to achieve strong fluctuation in secondary metabolism under different growth conditions, we developed the *natural product proteomining* approach, which allows correlating quantified levels of a secondary metabolite or bioactivity of interest to the concomitant changes in global protein expression profiles as analyzed by quantitative proteomics. In this study, we first demonstrate the positive correlation between the levels of the biosynthetic proteins and of the secondary metabolites they produce, irrespective of the nature of the metabolite. This was then applied to identify the gene clusters responsible for the biosynthesis of antibiotics in previously uncharacterized soil isolates. Because this method requires preidentification neither of the gene clusters nor of the natural product of interest, it allows identification of biosynthetic gene clusters of all types of natural products, requiring only a partial genome sequence from a single run of NGS sequencing.

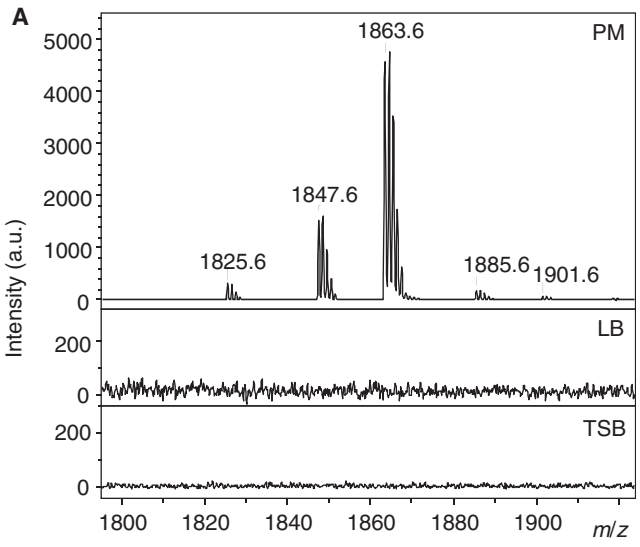
RESULTS

Correlation between Protein Expression and Secondary Metabolite Production

To investigate the correlation between the expression level of biosynthetic proteins and the amount of metabolite, we selected the *dasR*- and *rok7B7*-null mutants of the model actinomycete *S. coelicolor* A3(2). Both genes specify global regulators of metabolism in response to nutrient availability (Rigali et al., 2008; Świątek et al., 2013), and their null mutants show strong differential production of, among others, the readily detectable polyketide antibiotic undecylprodigiosin and the siderophore desferrioxamine (Craig et al., 2012). Because both products and their gene clusters have been well characterized (Barona-Gómez et al., 2004; Feitelson et al., 1985), they form ideal candidates to test the correlation between protein expression and production.

S. coelicolor M145 and its congenic *dasR*- and *rok7B7*-null mutants were grown in liquid minimal medium containing either ^{14}N or ^{15}N as the sole nitrogen source, until the late logarithmic phase when production of the red-pigmented antibiotic undecylprodigiosin became apparent. Stable isotope-labeled protein extracts ($^{14}N/^{15}N$) were mixed and separated by SDS-PAGE, followed by liquid chromatography (LC)-MS analysis of individual gel slices (Figure S1A available online). The metabolic labeling allowed for relative quantification of protein level changes between the parent and mutant strains. To filter out anomalies caused by minor differences in the isotope-labeled growth medium, only expression ratios with opposing signs in a label-swap experiment were retained. In total, expression of 346 proteins was significantly changed in either mutant, as compared with the parental strain M145 (Figure S1B). As expected, substantial expression changes were observed for proteins involved in secondary metabolite production, including production of desferrioxamines (Table 1; upregulated in both) and prodiginines (downregulated in *dasR*-null mutant and upregulated in *rok7B7*-null mutant).

In parallel, metabolite levels were determined by LC-MS for both mycelial (biomass) and spent medium (supernatant) samples from the same cultures. Both desferrioxamine B (*m/z* 614) and undecylprodigiosin/streptorubin B (*m/z* 392 and 394) were readily detected (Figure S1C) (Mo et al., 2008). LC-MS peak integration for mycelia demonstrated that desferrioxamine B levels were approximately 4-fold higher in the *rok7B7* deletion



B

Protein	normalized ^{2}log ratios		
	TSB / LB	PM / LB	PM / TSB
MrsG	0.3	4.9	4.4
MrsR2	0.4	0.9	0.4
MrsF	-0.4	5.0	5.7
MrsM	-0.1	5.1	5.6
MrsT	1.0	3.6	2.1
MrsA	1.6	3.4	1.3

Figure 1. Proteomics Analysis of Mersacidin Production

B. amyloliquefaciens HIL-Y85/54728 was grown in the indicated media: production medium (PM), LB, or TSB.

(A) After 5 days, spent media samples were subjected to MALDI-ToF MS analysis. Mersacidin production was observed when grown in PM only. The other peaks in the mass range shown corresponded to sodium and potassium adducts of these three isoforms.

(B) Protein extracts after 1 day of growth were subjected to proteomics analysis. Mersacidin was already detected after 1 day, albeit at low intensity (not shown). All six detected proteins involved in mersacidin production demonstrated elevated levels in production medium.

See also Data Set S1.

mutant and 2-fold higher in the *dasR* deletion mutant as compared with the parental strain M145 (Table 1). Undecylprodigiosin/streptorubin B were significantly reduced in the *dasR*-null mutant (<10% of wild-type levels), whereas levels were approximately 6-fold higher in the *rok7B7* mutant. These results confirm our hypothesis that protein expression levels and secondary metabolite production are strongly correlated.

We then wondered whether this observation could be extended to different types of secondary metabolites and to other microorganisms. Therefore, we analyzed *Bacillus amyloliquefaciens* HIL-Y85/54728, which produces the type B lantibiotic mersacidin, the biosynthesis of which is specified by a gene cluster consisting of ten coding sequences (CDSs) (Altena et al., 2000). Matrix-assisted laser desorption ionization-time of flight (MALDI-ToF) MS analysis of spent medium revealed that it was produced when *B. amyloliquefaciens* was grown in a synthetic production medium but not in the rich media tryptic soy broth (TSB) or Lucia broth (LB) (Figure 1A). Protein extracts were prepared from the same cultures and expression profiles correlated to the levels of mersacidin. Because the variation of growth conditions, including use of different nitrogen sources, ruled out ^{15}N metabolic labeling, dimethyl labeling of peptides (Boersema et al., 2009) was used for quantitative proteomics. Labeled peptides were first fractionated by strong cation exchange (SCX)-high-performance liquid chromatography (HPLC), followed by LC-MS analysis of each fraction. This resulted in the quantification of expression levels of six of the ten mersacidin producing proteins, including the prepeptide MrsA (Figure 1B). Expression of all proteins was upregulated in the production medium, exemplified by the immunity proteins MrsF and MrsG and the modification protein MrsM (at least 20-fold upregulated). The cluster-specific regulator MrsR2 was less strongly upregulated (less than 2-fold) than the biosynthetic proteins, which is in line with previous observations (van Wezel and McDowall, 2011). The complete proteomics data are presented in Data Set S1.

The observation that protein expression levels are directly proportional to the amount of secondary metabolite produced, regardless of the type of cluster or species, forms the basis for our natural product proteomining approach (Figure 2A): fluctuations in the proteome pool can be correlated to changes in a bioactivity of interest and/or the metabolome (as measured by MS and/or nuclear magnetic resonance [NMR]). Comparison of multiple culturing conditions and subsequent filtering of the proteomics data should allow identification of the relevant biosynthetic gene cluster in principal regardless of the exact nature of the secondary metabolite. Subsequent analysis of the cluster will aid compound identification.

Proteomining of a Soil Isolate Positively Identifies Actinomycin as a Source of Antibiotic Activity

Can natural product proteomining be applied with equal efficiency to an unknown natural isolate with a partial genome sequence obtained from a single NGS run? For this, a sufficiently accurate partial proteome prediction is required to identify the majority of the proteins, as well as significant variation in the expression of the compound of interest. To achieve the latter, many growth conditions were tested on a strain collection of some 800 actinomycetes to identify conditions that qualify best to achieve strong variation in the levels of natural products (H.Z., et al., submitted). Generally, comparison of liquid-grown minimal medium cultures supplied with growth conditions A (NaOH to pH 9), B (25 mM *N*-acetylglucosamine), C (0.8% [w/v] Bacto peptone [Difco]); D (0.5% [w/v] yeast extract), E (1%–2% [w/v] NaCl), or (–) (no additive) gave good variation in bioactivity.

A previously uncharacterized natural isolate from forest soil, *Streptomyces* sp. Che1, was analyzed, which produces a yellow/orange pigment with strong antimicrobial activity. The strain was grown for 5 days using the five additives mentioned above, and supernatants were tested for antibiotic activity using the Gram-positive bacterium *Micrococcus luteus* as an indicator strain (Figure 2B). Supernatants obtained after growth under

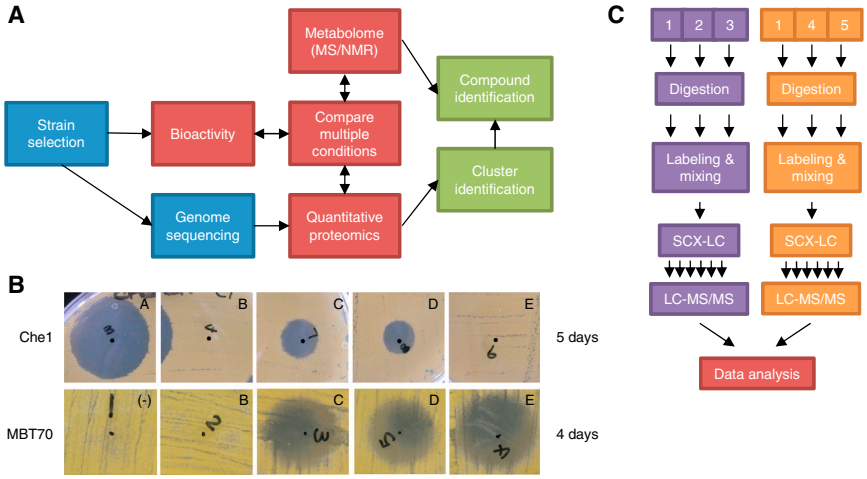


Figure 2. Natural Product Proteomining of *Streptomyces* sp. Che1 and MBT70

(A) General overview of the natural product proteomining method.

(B) *Streptomyces* sp. Che1 and MBT70 were grown in liquid NMMP media for 3–5 days using five different additives: A (NaOH to pH 9), B (25 mM *N*-acetylglucosamine), C (0.8% [w/v] Bacto peptone [Difco]), D (0.5% [w/v] yeast extract), and E (1%–2% [w/v] NaCl), or (–) (no additive). Spent medium samples were tested for antibiotics production using *M. luteus* as indicator strain.

(C) In parallel, protein expression levels in mycelia were compared using quantitative proteomics. Stable isotope labeling was performed through dimethylation of tryptic peptides. Because this method allows the comparison of three samples simultaneously, two experiments were performed, using one condition with the highest activity (condition 1) as a shared condition. The other four conditions (conditions 2–5) could thus be compared with the shared condition and one other condition (conditions 2 and 3 or conditions 4 and 5).

See also Figure S2.

condition A showed highest antimicrobial activity, and those obtained under conditions C and D effected medium growth inhibition, whereas the strain did not show detectable antimicrobial activity when grown under conditions B or E. MALDI-ToF MS analysis (Figure S2) identified multiple major peaks in the range of 1,255–1,330 Da for samples A, C, and D. The mass differences between the peaks corresponded to methylation (CH_2, +14 Da) and/or the presence of Na^+ (+22 Da) or K^+ (+38 Da) adducts, indicating that they were probably related. Signal intensities for all peaks demonstrated high correlation to antibiotic activity: high signal for condition A, low signal for conditions C and D, and no detectable signal for the other conditions (after normalization based on other signals observed in the MS spectra). Upon fractionation, the peaks were retained in the active fraction, and subsequent ^{1}H NMR spectroscopy revealed the presence of actinomycin as active compound (data not shown). Indeed, the observed masses corresponded exactly to those measured for pure actinomycin C2, C3, and D (C1; Figure S2) (Keller et al., 2010).

The partial genome sequence of *Streptomyces* sp. Che1 was obtained using a single run of paired end Illumina sequencing, and the output was assembled in 919 contigs. Identification of biosynthesis clusters using AntiSMASH (Blin et al., 2013) revealed the presence of 57 (partial) putative biosynthetic clusters. CDSs were predicted using the GeneMark algorithm, and 8,812 putative (and possibly partial) protein sequences were derived as a database for proteomics analysis. Because only three different labels are available in dimethyl labeling, the samples were compared in two independent quantitative proteomics experiments with one sample in common (samples A, B, and C and samples A, D, and E, respectively), with each experiment containing at least one sample of high activity and one sample of low activity (Figure 2C). Protein quantifications of the two experiments were combined, resulting in the identification of 2,645 proteins for Che1, with 1,863 proteins quantified in all comparisons with at least three independent events (Data Set S2). Importantly, expression of 52 predicted biosynthetic clusters (91%) could be detected. Based on the expression ratios and for each comparison independently, the proteins were divided into three quantiles, each containing an equal number of proteins: Q1 (downregulated), Q2 (unchanged), or Q3 (upregulated).

By using the NUCmer algorithm (Kurtz et al., 2004), seven contigs were positively mapped to the actinomycin gene cluster (Keller et al., 2010) (GenBank accession number HM038106, 48 kb) with 91% sequence coverage. In total, the products of 28 predicted open reading frames (ORFs) were detected and quantified in our experiment (Table 2), with the sequence of the large NRPS proteins AcmB and AcmC split over multiple CDSs and contigs, most likely because of the repetitive nature of the DNA sequence for these types of CDSs, leading to some misassembly. Two CDSs in contig 981 (CDSs 1 and 12) demonstrated expression level ratios in the correct quantile for all six comparisons out of only 34 CDSs for the complete proteome and, importantly, were the only two such hits clustered together on a single contig. Thus, the proteomics result seems to confirm that actinomycin is produced and responsible for the observed antibiotic activity.

However, because most CDSs in the actinomycin cluster did not match the correct pattern for all six comparisons, we anticipated that filtering might be more realistic by using only the four comparisons that included the proteome for the culture grown under condition A, because this culture demonstrated high activity (Figure 2B) and high actinomycin levels (Figure S2) as compared with the other conditions. Indeed, all potential CDSs in the central region of the actinomycin cluster (*acmT–acmI*, 18 CDSs) matched this filter condition (Table 2; Figure S3). Considering that the central region of the cluster contains the CDSs for the biosynthetic proteins, whereas the flanking regions contain genes coding for resistance and regulation (Keller et al., 2010), this again confirmed that expression levels of the biosynthetic proteins best match metabolite levels.

Relaxing the filter conditions to only four comparisons inevitably increased the number of hits in the total proteome considerably to 298 CDSs. Still, only seven contigs contained at least five matching CDSs clustered together (<10 nonmatching or

Table 2. Expression Level Changes of CDSs Coding for Actinomycin Biosynthesis

			Normalized Ratios (^{2}log)[a]						Quantiles[b]						Count[c]					
contig	CDS	Gene[d]	B/A	C/A	C/B	D/A	E/A	E/D	**B/A**	**C/A**	C/B	**D/A**	**E/A**	E/D	B/A	C/A	C/B	D/A	E/A	E/D
981	23	*acmrc*	−2.8	0.0	2.4	−3.5	−3.3	0.3	**Q1**	Q2	**Q3**	**Q1**	**Q1**	Q3	40	40	40	26	26	26
981	22	*acmrb*	−3.1	−0.3	2.8	−2.9	−2.0	1.9	**Q1**	Q2	**Q3**	**Q1**	**Q1**	Q3	7	7	7	9	9	9
981	21	*acmra*	−3.6	−0.3	3.2	−4.2	−3.6	0.9	**Q1**	Q2	**Q3**	**Q1**	**Q1**	Q3	31	31	31	22	22	22
981	20	*acmq*	−2.8	0.9	3.3	−3.2	−1.7	1.4	**Q1**	Q3	**Q3**	**Q1**	**Q1**	Q3	60	60	60	62	60	60
981	19	*acmq*	−2.0	1.0	2.9	−3.6	−2.6	0.9	**Q1**	Q3	**Q3**	**Q1**	**Q1**	Q3	16	16	16	9	9	9
981	18	*acmp*	−1.7	0.9	2.4	−1.3	−1.6	0.1	**Q1**	Q3	**Q3**	Q2	**Q1**	Q2	7	7	7	5	5	5
981	**12**	***acmt***	−2.8	−2.2	0.4	−2.1	−3.8	−0.8	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	**Q1**	11	11	11	18	16	16
981	**11**	***acms***	−2.6	−2.3	0.4	−3.5	−3.0	0.3	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	52	52	52	61	57	57
981	**8**	***acma***	−2.5	−1.2	1.2	−3.7	−3.0	0.9	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	39	38	38	38	33	33
981	**4**	***acmb***	−3.6	−3.1	0.5	−4.0	−4.7	−0.4	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q2	20	20	20	25	23	23
981	**1**	***acmc***	−2.1	−1.2	0.6	−4.0	−4.6	−1.0	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	**Q1**	15	15	15	9	9	9
237	**4**	***acmb/c***	−3.1	−2.6	0.6	−3.9	−3.1	1.0	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	72	72	72	84	81	81
237	**6**	***acmb/c***	−3.4	−2.4	0.9	−3.8	−3.2	0.2	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	44	44	44	43	41	41
237	**7**	***acmc***	−3.0	−2.4	0.4	−4.3	−2.5	1.9	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	12	12	12	15	15	15
1020	**1**	***acmb***	−3.7	−2.5	1.0	−3.4	−4.0	1.1	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	23	23	23	14	14	14
1020	**2**	***acmb/c***	−3.3	−2.5	0.6	−3.7	−2.3	0.5	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	11	11	11	13	13	13
414	**1**	***acmb/c***	−1.4	−2.0	−0.7	−3.6	−4.0	0.0	**Q1**	**Q1**	Q1	**Q1**	**Q1**	Q2	3	3	3	3	3	3
925	**1**	***acmb/c***	−3.8	−2.7	1.0	−1.7	−1.6	0.4	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	7	7	7	4	4	4
816	**1**	***acmc***	−3.5	−2.4	0.8	−3.7	−3.9	0.4	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	46	46	46	46	38	38
816	**2**	***acmc***	−2.9	−1.8	1.0	−3.2	−3.1	0.7	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	39	39	39	44	42	42
816	**3**	***acme***	−2.2	−2.7	−0.1	−3.2	−2.1	0.6	**Q1**	**Q1**	Q2	**Q1**	**Q1**	Q3	10	10	10	16	15	15
816	**4**	***acmf***	−2.8	−3.5	−0.5	−4.9	−3.0	1.8	**Q1**	**Q1**	Q1	**Q1**	**Q1**	Q3	3	3	3	5	5	5
816	**7**	***acmh***	−2.0	−2.6	−0.2	−2.3	−2.3	0.3	**Q1**	**Q1**	Q2	**Q1**	**Q1**	Q3	15	15	15	21	20	20
816	**8**	***acmi***	−3.2	−2.8	0.4	−4.3	−3.8	0.5	**Q1**	**Q1**	**Q3**	**Q1**	**Q1**	Q3	103	100	100	106	99	99
816	10	*acmu*	−2.7	0.9	3.4				**Q1**	Q3	**Q3**				4	4	4	*2*	*2*	*2*
816	11	*acmv*	−2.7	0.1	2.8	−3.6	−3.5	0.3	**Q1**	Q3	**Q3**	**Q1**	**Q1**	Q3	40	40	40	32	32	32
793	16	*acmw*	−3.8	−0.3	3.2	−2.6	−3.4	−0.7	**Q1**	Q2	**Q3**	**Q1**	**Q1**	**Q1**	14	14	14	10	10	10
793	14	*acmy*	−1.7	0.4	1.9				**Q1**	Q3	**Q3**				5	5	5	*2*	*2*	*2*

See also Figure S3 and Data Set S2.

[a]Changes in protein expression levels observed for the indicated CDSs when compared between growth conditions A–E.

[b]Expression ratios were divided in three equally sized quantiles for each experiment. Where the expression level change corresponded to the expected quantile, this is indicated in bold. Where all four comparisons used for filtering (containing condition A, indicated in bold) matched to the expected quantile, the CDS number is also indicated in bold.

[c]The number of quantifications events used to calculate the expression ratios. Quantifications based on less than three events (italicized) were discarded.

[d]Gene name according to GenBank.

undetected CDSs in between; Table 3), of which two contigs (816 and 981) were part of the actinomycin cluster. Closer inspection of the other contigs by annotation based on BLAST similarity searches suggested that two contained genes for the NADH-dehydrogenase complex, whereas the other three contained features related to natural product biosynthesis. When the threshold value was lowered from five to three matching CDSs per cluster, contig 237 from the actinomycin cluster (Table 2) was the only additional identified contig that clearly coded for biosynthetic activity. In conclusion, relaxing filter conditions added additional hits for the actinomycin cluster, together with some false positives. Because actinomycin was the only match when all six comparisons were considered and the expression of the genes in the cluster correlated well with the bioactivity for the four comparisons with the largest variation in activity, this validated our approach for application to link bioactivity to a gene cluster in soil isolates.

Identification of a Juglomycin

For further support of our method, we analyzed the previously undescribed soil isolate *Streptomyces* sp. MBT70. Supernatants from cultures grown for 4 days under conditions C and E contained strong antimicrobial activity, whereas those from cultures grown under condition D had slightly lower activity (Figure 2A). Hardly any or no activity was seen under conditions ([minus]) and B, respectively.

Sequencing of MBT70 yielded 396 contigs coding for 8,449 potential protein sequences and 56 (partial) putative biosynthetic clusters. Protein expression profiles of cultures grown under conditions ([minus]), C, and E (experiment 1) and conditions B,

Table 3. Candidate Clusters Demonstrating Expected Expression Level Changes

Contig[a]	Quantified[b]	Matching	Total CDSs	Anti-SMASH[c]	Blast Analysis[d]	GenBank Accession Number
Sp. Che1						
42	6	5	6	+	Streptomyces sp. W007 NRPS	AGSW0100016
412	6	6	6		NADH dehydrogenase	
419	7	6	8		NADH dehydrogenase	
814	18	17	21	+	nonactin	AF263011, AF074603
816	6	6	8	+	actinomycin	HM038106
981	5	5	12	+	actinomycin	HM038106
1256	10	8	18	+	skyllamycin	JF430460
Sp. MBT70						
Up in C and E						
561	4	3	5	+	*S. antibioticus* polyketide	Y19177
Up in D						
43	11	9	15	–	ATP synthase	
45	7	7	14	–	sugar metabolism	
443	13	13	22	+	*S. ansochromogenes* pks2	KF170322

See also Tables S1 and S2 and Data Sets S2 and S3.

[a]Changes in protein expression levels were compared between growth conditions (A–E and no additive (–); see main text and Figure 2). Expression ratios were divided in three equally sized quantiles for each comparison and filtered based on the four (three for MBT70, up in D) comparisons with the largest change in antibacterial activity (see main text). Contigs with at least five matching CDSs in a cluster (maximum gap < 10 CDSs) were selected. For *Streptomyces* sp. MBT70, only one cluster upregulated in C and E with a maximum of three matching CDSs could be identified.

[b]The region between the first matching CDS and last matching CDS was defined as a cluster as to compare the number of quantified CDSs and the number of matching CDSs to the total number of CDSs in the cluster.

[c]Clusters were analyzed with antiSMASH v2.0 (Blin et al., 2013) for the presence of secondary metabolite biosynthesis clusters. A hit is indicated with a plus sign.

[d]Sequences were compared using BLAST analysis to known streptomycetes sequences in the NCBI nr/nt and WGS (genomic shotgun sequences) databases. Hits with more than 95% identity were used for annotation.

C, and D (experiment 2) were compared by quantitative proteomics (Figure 2B), yielding 2,132 protein identifications, with 1,087 proteins quantified in all comparisons (Data Set S3). Again, expression of a high number of predicted clusters was detected (48 clusters, 86%). Similar filtering as described for *Streptomyces* sp. Che1 was applied to *Streptomyces* sp. MBT70, comparing conditions C and E with high antimicrobial activity to the other conditions with reduced or no activity. With three clustered hits, only CDSs 118–122 on contig 561 matched the criteria (Figure 3A; Table 3). However, the ratio found for the comparison between conditions C and D was striking (Table S1; Figure S3), suggesting that little or no compound was produced under condition D. Therefore, the data were reanalyzed using the three comparisons, including condition D (upregulation) and condition B (no change or downregulated). This identified three contigs with more than five matching hits (Table 3), with the top hit (contig 443, 13 matches, specifically upregulated in condition D; Table S2) encoding a PKS type I (GenBank accession number KF170322) with a yet unknown product (Zhong et al., 2013). Additional BLASTN analysis revealed that contig 549 also contained part of this cluster, giving a total of 17 matching CDSs (Table S2; Figure S3).

The candidate cluster in contig 561 is highly similar (>98% nucleotide identity) to a PKS gene cluster in *Streptomyces antibioticus* (GenBank accession number Y19177) (Colombo et al., 2001) for the first, shared steps in the biosynthesis of pyranonaphthoquinones, a class of compounds that includes actinorhodin, granaticin, and medermycin (Hopwood, 1997; Metsä-Ketelä et al., 2013). However, a gene encoding a homolog of the actinorhodin biosynthetic protein ActIV of *S. coelicolor*, catalyzing the final common step to a bicyclic intermediate, is missing from this cluster; this suggested that genes specifying subsequent steps might be located elsewhere on the MBT70 genome. TBLASTN analysis using the protein sequences from the actinorhodin cluster as input identified a gene for a homolog of ActIV (52% amino acid identity) 115 CDSs upstream of the candidate gene cluster in contig 561 (CDS3; Figures 3A and 3B). Importantly, homologs were also detected for other genes from the actinorhodin cluster and were located next to the candidate cluster and near the cyclase gene. Two additional proteomining hits were found in the genomic region near the cyclase gene (CDS11 and CDS12; Table S1), one of which was homologous to *act*VI-ORF1. Because CDS128 (final CDS), CDS1, and CDS2 specify the biotin binding, biotin carboxylase, and the carboxyl transferase domains of acetyl-CoA carboxylase, respectively, it seems likely that the last part and first part of this contig should actually be continuous and that the two identified regions on this contig therefore form one biosynthetic gene cluster (Figure 3A).

To link pyranonaphthoquinone biosynthesis under different growth conditions to the observed bioactivity, NMR-based metabolomics (Kim et al., 2010) was applied to EtOAc extracts of spent media. ^{1}H NMR spectra of five replicates of each condition were analyzed by partial least square modeling-discriminant analysis (PLS-DA; Figures 4A and 4B). Conditions C and E could be distinguished from the other conditions (Figure 4A),

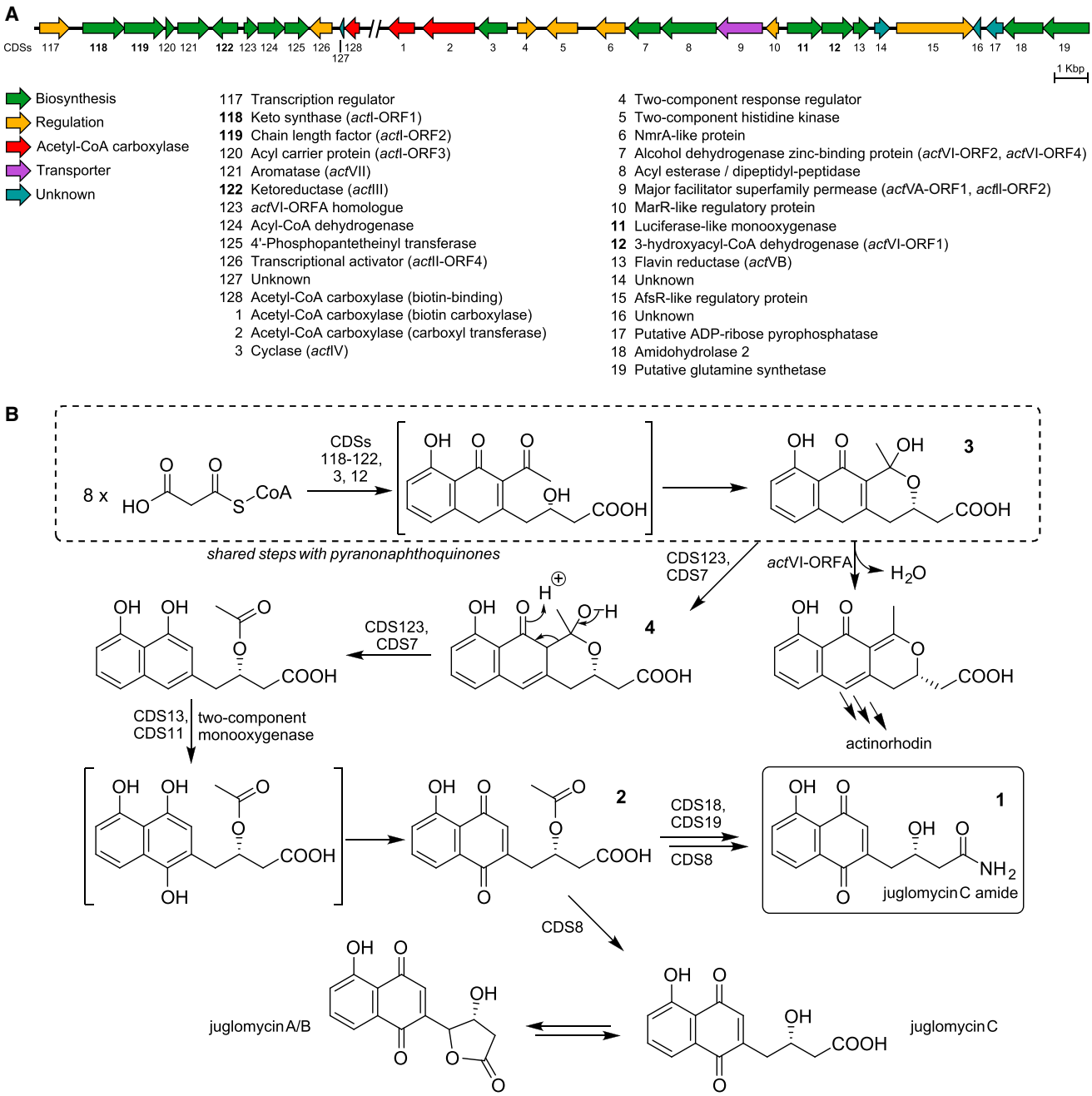


Figure 3. Gene Cluster and Biosynthesis of a Juglomycin

(A) The biosynthetic gene cluster coding for the synthesis of a juglomycin-like compound **1**. Annotations are based on BLAST homology searches. Based on the annotation of the three acetyl-CoA carboxylase genes (depicted in red), the end and beginning of contig 561 of sp. MBT70 were joined to yield one gene cluster. In case CDS products demonstrated expected expression patterns changes in proteomining, the CDSs are depicted in bold.

(B) Proposed biosynthetic route to **1**. The first, shared steps with pyranonaphthoquinone synthesis are indicated by a dashed box.

See also Figure S3, Table S1, and Table S3.

strengthening our hypothesis that a different bioactivity is produced in condition D. Main contributors to this difference were several phenolic resonances (Figure 4B). In particular, the resonance in δ7.5–δ7.6 was identified as an H-5 of naphthoquinone type compounds, which was confirmed by the correlation between H-5 and C-4 in a heteronuclear multiple bonds correlation (HMBC) spectrum (Figure 4C). Because naphthoquinones demonstrate typical absorption of >400 nm, the extract of condition E was subjected to HPLC fractionation on a C18 column, and seven fractions that absorb at this wavelength were collected, of which two demonstrated antibiotic activity against *M. luteus*. ^{1}H correlation spectroscopy, and HMBC-NMR

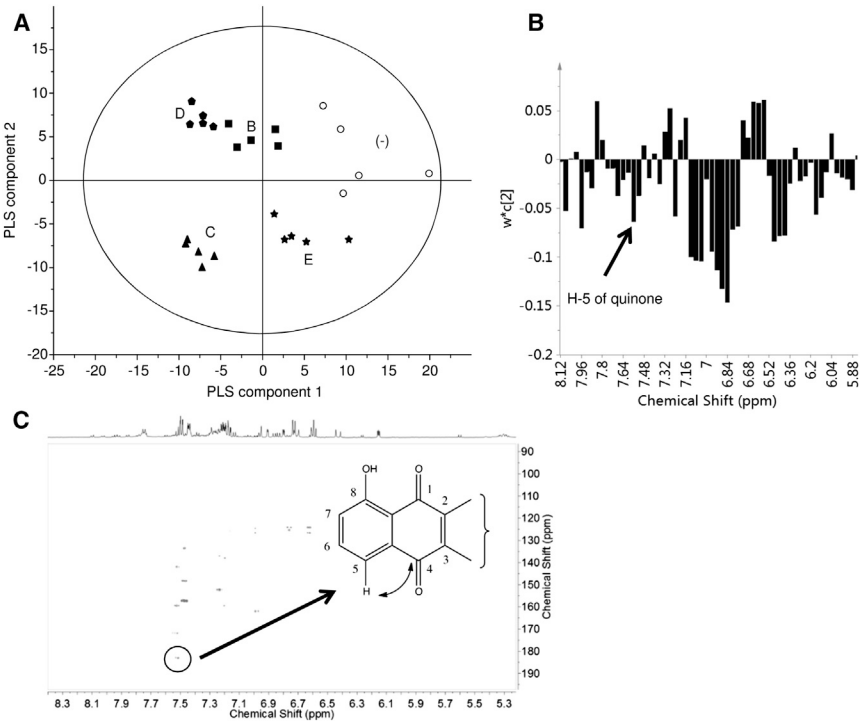


Figure 4. Metabolomics Analysis of *Streptomyces* sp. MBT70

(A and B) Five biological replicates of *Streptomyces* sp. MBT70 were grown under the conditions as described for Figure 2. ^{1}H NMR spectra of EtOAc extracts of spent medium were subjected to PLS-DA to obtain score (A) and loading (B) plots. The ellipse represents the Hotelling T^2 with 95% confidence. The arrow indicates the signal obtained for H-5 of naphthoquinone.

(C) HMBC NMR spectrum of condition C in the range of δ5.2–δ8.4 (horizontal axis for ^{1}H) and δ90–δ200 (vertical axis for ^{13}C). Again, the arrow indicates the signal obtained for H-5 of naphthoquinone.

analysis (Table S3) confirmed the presence of a naphthoquinone-type compound in one fraction. Based on these data and high resolution MS analysis ($[M+H]^+$ 276.0869 *m/z*, $C_{14}H_{13}NO_5$), we propose that the gene cluster identified by proteomining specifies a juglomycin C-type compound (Lessmann et al., 1989) with an amide functionality: β-(*S*),5-dihydroxy-1,4-naphthoquinone-2-butanamide (juglomycin C amide; Figure 3B, **1**).

Juglomycin-type compound **2** was previously identified as an actinorhodin shunt product in an *S. coelicolor act*VI-ORFA mutant (Ozawa et al., 2003), probably because of inefficient pyran ring formation (Figure 3B). Instead, the intermediate hemiketal **3** isomerizes to **4,** followed by C–C bond cleavage in a retro-Claisen type reaction and oxidation to yield naphthoquinone **2**. In our cluster, products of CDS123 (*act*VI-ORFA) and CDS7 (*act*VI-ORF2, *act*VI-ORF4) might catalyze the isomerization and subsequent C–C bond cleavage, instead of pyran ring formation. For the oxidation step, CDS11 and CDS13 (*act*VB) most probably specify a two-component flavin-dependent monooxygenase system (Taguchi et al., 2013). Although CDS11 is not an *act*VA-ORF5 homolog, its bicyclic substrate is sufficiently different from the tricyclic intermediate in actinorhodin biosynthesis to explain why an alternative luciferase-like monooxygenase is used here. Finally, to yield **1**, the products of CDS18 and/or CDS19 likely catalyze amide formation, followed by removal of acetate by the peptidase/esterase specified by CDS8. Amidation would prevent lactonization to juglomycin A/B as was observed for juglomycin C (Figure 3B) (Lessmann et al., 1989). In conclusion, natural product proteomining of *Streptomyces* sp. MBT70 yielded a juglomycin-type antibiotic that was directly linked to its biosynthetic gene cluster. Subsequent analysis of the compound and the gene cluster confirmed the proteomining result.

DISCUSSION

The genome sequencing era has recently led to the identification of unexpected large numbers of biosynthetic gene clusters in actinomycetes, often referred to as cryptic or silent clusters because of their poor expression under routine laboratory conditions. Because soil isolates can be notoriously difficult to manipulate using common molecular biological techniques, the application of eliciting conditions to induce expression provides a promising alternative (Craney et al., 2012; Zhu et al., 2014) because we could detect expression of more than 85% of all predicted clusters in soil isolates. This, in turn, calls for new technologies that allow the straightforward linkage between a potentially novel natural product and its biosynthetic gene cluster. Such a connection may help to ascertain novelty of the compound, to guide fractionation and identification attempts, and to allow molecular biological approaches, possibly in combination with heterologous expression in a suitable host (Komatsu et al., 2013). Our natural product proteomining technology provides a concept for the connection of theoretically any bioactivity—such as a secondary metabolite with antimicrobial, antifungal, or anticancer activity—to a gene cluster, using a quantitative proteomics-based approach combined with a (partial) genome sequence.

After successful initial tests in *S. coelicolor* and *B. amyloliquefaciens* demonstrating that protein expression profiles directly correlated to secondary metabolite levels, we went on to identify the gene clusters responsible for production of antimicrobial compounds observed in two soil isolates. Analysis of the actinomycin producer *Streptomyces* sp. Che1 allowed further refinement of filtering of the proteomics data, and the technology was then applied to connect a biosynthetic gene cluster to a juglomycin derivative produced by *Streptomyces* sp. MBT70. Thus, natural product proteomining allows the identification of gene clusters for different classes of natural products; specifically, we applied the method to positively identify the gene clusters specifying nonribosomal peptides and polyketides in *Streptomyces* and a lantibiotic in *Bacillus*.

We employed quantitative proteomics based on labeling with stable isotopes, which is a well-established technique that is

easy to implement. The variation of growth conditions hampers metabolic incorporation of isotopes, but chemical modification of peptides provided good results. Depending on instrument availability, isobaric tags like iTRAQ or tandem mass tags might be used instead for peptide modifications, because they allow for the simultaneous comparisons of up to eight samples in one proteomics run. The current method of dimethyl labeling is limited to three comparisons per run, necessitating the use of two runs to obtain the four comparisons that were used for data filtering. However, dimethyl labeling is more cost-effective (Kovanich et al., 2012) and, perhaps more importantly, yields more data points as compared with MS/MS-based quantification of isobaric labels.

Compared with other approaches for the identification of expressed biosynthetic clusters, based on either proteomics (Bumpus et al., 2009; Meier et al., 2009), transcriptomics (Qu et al., 2011), or peptide/glycan fingerprinting (Kersten et al., 2011; Kersten et al., 2013), the usage of expression levels should allow the detection of all types of clusters specifying natural products simultaneously. Moreover, because no specific sequences were targeted, expression levels of a majority of proteins specified by the clusters could be determined and, for the biosynthetic proteins in particular, were coregulated. For instance, the central 18 CDSs of the actinomycin biosynthetic gene cluster, containing the genes for the chain assembly proteins and subunit tailoring (Keller et al., 2010), demonstrated excellent agreement with the expected expression pattern. This allows accurate determination of cluster members, as described before for transcriptomic analysis of *Aspergillus nidulans* (Andersen et al., 2013), and piecing together gene clusters spread over multiple contigs.

To identify previously unreported biosynthetic clusters, full genome sequencing is crucial (Albright et al., 2014). Fortunately, in case of bacterial genomes, this is now a routine analysis and continues to become more efficient and cost-effective. As an alternative, RNA-Seq could be applied to determine expression level changes in the transcriptome without prior sequence information. However, in this case two technological limitations need to be overcome. First, quantification of repetitive gene sequences, as often observed in NRPS and PKS, is inherently difficult in RNA-Seq analysis as similar short reads might be misinterpreted as multiple copies of the same transcript (Martin and Wang, 2011). Second, clustering of proteomics hits on the genome was found to be a particularly important criterion for filtering of the results, and similar information on gene clustering in RNA-Seq analysis is only obtained in case the genes are transcribed as a polycistronic mRNA.

An unambiguous identifier for the bioactivity of interest, such as an MS and/or NMR signal, significantly enhances the effectiveness of natural product proteomining. The analysis becomes more complicated when a bioactivity marker can represent multiple bioactivities produced by the same species, such as growth inhibition of an indicator strain as exemplified by a second antimicrobial activity in *Streptomyces* sp. MBT70 produced under condition D (yeast extract). Subsequent NMR analysis confirmed the production of an alternative bioactive compound under this condition, the identity of which is currently under investigation. Moreover, sufficient fluctuation of production is essential for proper correlation of expression levels to metabolite levels. The lower utility of the comparison between fractions with low bioactivity (i.e., conditions D and E for *Streptomyces* sp. Che1) is a consequence of the marginal fluctuations in protein expression levels, and such comparisons are therefore less feasible. We recently analyzed the inducibility of antibiotic production in a collection of over 800 actinomycetes from mountain soils, under 40 different culturing conditions. The eliciting conditions effected strong fluctuation in antibiotic production, and many of the actinomycetes produced antibiotics against multidrug-resistant pathogens only under a few or even just one of the 40 conditions (H.Z., et al., submitted). Because "silent" antibiotic gene clusters are by definition not expressed significantly under routine growth conditions, their induction under specific conditions will lead to a major increase in bioactivity and protein expression levels, which is ideal for our approach. Indeed, the technology was designed with exploiting this property of previously unidentified natural products in mind.

In summary, we have developed an 'omics-based approach that exploits the direct correlation between bioactivity and protein expression profiles, offering an efficient and straightforward way to identify the gene cluster responsible for the production of a natural product of interest. Importantly, considering that this method requires neither prior knowledge of gene sequences nor structural information on the bioactive compound, it is uniformly applicable to all types of natural product clusters and, theoretically, even to completely new types of clusters. This will support quests for novel bioactive compounds, such as the antimicrobials that are so badly needed to combat infectious diseases associated with drug-resistant pathogens that currently threaten the health of patients in the clinic.

SIGNIFICANCE

In the genomics era, where the number of genomes sequenced increases rapidly, numerous gene clusters for natural product biosynthesis are uncovered. Still, the exact structure and production levels of a natural product can often not be deduced from the DNA sequence alone. Natural product proteomining is a method that allows linkage of a bioactive compound to a gene cluster when a (partial) genome sequence is available. This is achieved by comparison of protein expression levels to metabolite levels under a variety of growth conditions that effect differential production of the natural product of interest. Quantitative proteomics is a well-established technique and can be executed on a routine basis. Moreover, no prior knowledge is required regarding the nature of the product or its gene cluster, assuming that it is synthesized by proteins whose expression levels correlate with the levels of metabolite, which, as we demonstrated, is typically the case. For this reason, natural product proteomining is an expression-based technology that is equally applicable to all known classes of compounds and might even be applied to new compound classes. We validated the technology by identifying the biosynthetic gene clusters for antibiotics in two soil isolates of the Gram-positive bacterium *Streptomyces*. One cluster specifies the biosynthetic proteins for actinomycin, whereas the other specifies the biosynthetic proteins for juglomycin C amide. In addition, the applicability to the discovery of

lantibiotic production was demonstrated in the firmicute *Bacillus*. In principle, by scaling up the number of conditions, it should be possible to map all natural products from a single organism to their respective biosynthetic gene clusters. With that, we expect that the technology will facilitate the identification of gene clusters for compounds of high relevance, such as antibiotics for treatment of the rapidly emerging multidrug-resistant pathogens.

EXPERIMENTAL PROCEDURES

Strains and Growth Conditions

S. coelicolor A3(2) M145 was obtained from the John Innes Centre strain collection. The *dasR*-null mutant (SAF29) (Rigali et al., 2006) and *rok7B7*-null mutant (GAM33) (Świątek et al., 2013) of *S. coelicolor* were described previously. *B. amyloliquefaciens* HIL-Y85/54728 was obtained from Novacta Biosystems (Welwyn, Garden City, UK). *Streptomyces* strains Che1 and MBT70 were obtained de novo from soil samples as described (H.Z., et al., submitted). All *Streptomyces* strains were grown as indicated according to routine methods (Kieser et al., 2000).

S. coelicolor M145 and its congenic *dasR* and *rok7B7* deletion mutants were grown in adapted NMMP medium for $^{14}N/^{15}N$ labeling (Świątek et al., 2013). Samples were taken at late logarithmic phase when production of pigmented antibiotics became apparent. $^{14}N/^{15}N$-labeling experiments were performed in duplicate with a label swap to avoid that differences in media composition should affect the outcome of the proteomics experiments.

A seed culture of *B. amyloliquefaciens* was grown in TSB for 24 hr as described (Appleyard et al., 2009), before transfer (1:50, v/v) to mersacidin production medium, LB, or fresh TSB. Cultures were grown for 5 days at 30°C. Proteomics samples were taken after 24 hr because protein levels were too low after 5 days of growth.

Streptomyces strains Che1 and MBT70 were grown in liquid NMMP medium as described (H.Z., et al., submitted). Essentially, strains were grown in NMMP containing 1% (w/v) glycerol and 0.5% (w/v) mannitol as carbon sources for 4–6 days, using five different additives to create varying growth conditions: (–) (no additive), A (NaOH to pH 9), B (25 mM *N*-acetylglucosamine), C (0.8% [w/v] Bacto peptone [Difco]), D (0.5% [w/v] Bacto yeast extract [Difco]), and E (1 [MBT70] or 2% [Che1] [w/v] NaCl). For antibiotic activity assays, *M. luteus* was spread onto LB agar plates, and 20 μl of spent medium were placed on the plates. After growth at 30°C overnight, the growth inhibition zone was measured. No inhibition was observed when unused medium was spotted.

Metabolite Analysis

For *S. coelicolor*, mycelium was extracted with methanol; the extract and spent medium were analyzed directly by LC-MS for undecylprodigiosins and desferrioxamines, as detailed in Supplemental Experimental Procedures. Spent medium samples of *Streptomyces* sp. Che1 and *B. amyloliquefaciens* were mixed 1:1 (v/v) or 1:10 (v/v) with matrix solution, respectively, and analyzed by MALDI-ToF MS in the positive ion reflectron mode as detailed in Supplemental Experimental Procedures. For *Streptomyces* sp. MBT70, 20 ml of spent medium of five biological replicates was liquid-liquid partitioned using 20 ml EtOAc. This was repeated twice, after which the combined EtOAc fractions were concentrated by rotary evaporation at 40°C and subjected to NMR analysis. NMR conditions have been described previously (Kim et al., 2010) and are detailed in Supplemental Experimental Procedures.

For multivariate data analysis, ^1H NMR spectra were automatically reduced to ASCII files using AMIX (v. 3.7, Bruker Biospin). Spectral intensities were scaled to TMSP and reduced to integrated regions of equal width (0.04 ppm) corresponding to the region of δ0.3–δ10.00. The region of δ4.7–δ5.0 and δ3.28–δ3.34 were excluded from the analysis because of the residual signal of H_2O and CH_3OH-d_4, respectively. PLS-DA was performed with the SIMCA-P software (v. 13.0, Umetrics, Umeå, Sweden) with unit variance scaling methods.

For naphthoquinone purification, extracts of the five biological replicates of condition E were pooled and fractionated by reversed phase HPLC, based on absorption at 410 nm, as detailed in Supplemental Experimental Procedures. Fractions were tested for antibiotic activity as described above. Active fractions were subjected to NMR analysis and, after reconstitution in 1:1 (v/v) H_2O/acetonitrile with 0.1% (v/v) formic acid, analyzed on a Thermo LTQ-orbitrap mass spectrometer equipped with an ESI source.

Illumina Sequencing

Illumina/Solexa sequencing on Genome Analyzer IIx was outsourced (ServiceXS, Leiden, The Netherlands). One hundred nucleotide paired-end reads were obtained and analyzed as detailed in Supplemental Experimental Procedures.

Quantitative Proteomics Analysis

Mycelia or *Bacillus* cells were harvested by centrifugation, washed, and sonicated for 5 min at 12 W output power using 5 s on/5 s off intervals in 100 mM Tris/HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM dithiothreitol. Debris was removed by centrifugation at 16,000 × *g* for 10 min at 4°C. Protein concentration of the extracts was determined using a Bradford protein assay, using BSA as standard. ^{14}N-labeled and ^{15}N-labeled mycelial extracts were mixed 1:1 for protein content, and proteins were separated by SDS-PAGE on mini-Protean TGX Any kD gels (Bio-Rad) using a broad range protein ladder (Bio-Rad), followed by in gel digestion, all as described (Świątek et al., 2013). In-solution digestion and dimethyl labeling of Che1, MBT70, and *Bacillus* extracts were performed as described (Gubbens et al., 2012), using 0.167 mg of total protein per sample. Labeled peptides were mixed 1:1:1 to yield mixtures containing 0.5 mg of protein each. Acetonitrile was removed using a vacuum concentrator, and peptides were subjected to SCX fractionation as detailed in Supplemental Experimental Procedures. LC-MS/MS analysis on an LTQ-Orbitrap (Thermo, Waltham, MA) for both gel-extracted peptides ($^{14}N/^{15}N$ labeling) (Florea et al., 2010) and SCX fractions (dimethyl labeling) (Gubbens et al., 2012) was performed as described, respectively.

Data analysis of $^{14}N/^{15}N$ labeled samples using MSQuant (Mortensen et al., 2010) has been described elsewhere (Świątek et al., 2013). Data analysis of dimethyl-labeled samples was performed using MaxQuant 1.2.2.5 (Cox and Mann, 2008) as described (Gubbens et al., 2012). For *B. amyloliquefaciens* HIL-Y85/54728, the *B. amyloliquefaciens* FZB42 complete proteome set (Uniprot 2012_10) with 98.5% sequence identity (Herzner et al., 2011) was appended with the ten mersacidin-producing proteins annotated for *B. amyloliquefaciens* HIL-Y85/54728 (Uniprot). For the *Streptomyces* strains Che1 and MBT70, identified CDSs were translated to obtain a protein database, and the two mixtures obtained for each strain were analyzed in one MaxQuant run. Normalized protein expression ratios were split in three equally sized quantiles (up, unchanged, or down). Expression ratio filtering was based on selection of the expected quantile for each comparison.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, three tables, and three data sets and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2014.03.011.

ACKNOWLEDGMENTS

We are grateful to Hans van den Elst for product purification and Mervyn Bibb for discussions. *Bacillus amyloliquefaciens* HIL-Y85/54728 was a kind gift from Novacta Biosystems (Welwyn, Garden City, UK). This work was supported by VENI and VICI grants from the Netherlands Technology Foundation STW to J.G. and G.P.v.W., respectively, and a China Scholarship Council grant to H.Z.

Received: August 31, 2013
Revised: March 2, 2014
Accepted: March 24, 2014
Published: May 8, 2014

REFERENCES

Albright, J.C., Goering, A.W., Doroghazi, J.R., Metcalf, W.W., and Kelleher, N.L. (2014). Strain-specific proteogenomics accelerates the discovery of

natural products via their biosynthetic pathways. J. Ind. Microbiol. Biotechnol. *41*, 451–459.

Altena, K., Guder, A., Cramer, C., and Bierbaum, G. (2000). Biosynthesis of the lantibiotic mersacidin: Organization of a type B lantibiotic gene cluster. Appl. Environ. Microbiol. *66*, 2565–2571.

Andersen, M.R., Nielsen, J.B., Klitgaard, A., Petersen, L.M., Zachariasen, M., Hansen, T.J., Blicher, L.H., Gotfredsen, C.H., Larsen, T.O., Nielsen, K.F., and Mortensen, U.H. (2013). Accurate prediction of secondary metabolite gene clusters in filamentous fungi. Proc. Natl. Acad. Sci. USA *110*, E99–E107.

Appleyard, A.N., Choi, S., Read, D.M., Lightfoot, A., Boakes, S., Hoffmann, A., Chopra, I., Bierbaum, G., Rudd, B.A., Dawson, M.J., and Cortes, J. (2009). Dissecting structural and functional diversity of the lantibiotic mersacidin. Chem. Biol. *16*, 490–498.

Arias, C.A., and Murray, B.E. (2009). Antibiotic-resistant bugs in the 21st century: A clinical super-challenge. N. Engl. J. Med. *360*, 439–443.

Barona-Gómez, F., Wong, U., Giannakopulos, A.E., Derrick, P.J., and Challis, G.L. (2004). Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. J. Am. Chem. Soc. *126*, 16282–16283.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature *417*, 141–147.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., and Weber, T. (2013). antiSMASH 2.0: A versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res. *41* (Web Server issue), W204–W212.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat. Protoc. *4*, 484–494.

Bumpus, S.B., Evans, B.S., Thomas, P.M., Ntai, I., and Kelleher, N.L. (2009). A proteomics approach to discovering natural products and their biosynthetic pathways. Nat. Biotechnol. *27*, 951–956.

Challis, G.L., and Hopwood, D.A. (2003). Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. Proc. Natl. Acad. Sci. USA *100* (*Suppl 2*), 14555–14561.

Claesen, J., and Bibb, M. (2010). Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. Proc. Natl. Acad. Sci. USA *107*, 16297–16302.

Claessen, D., Rozen, D.E., Kuipers, O.P., Søgaard-Andersen, L., and van Wezel, G.P. (2014). Bacterial solutions to multicellularity: A tale of biofilms, filaments and fruiting bodies. Nat. Rev. Microbiol. *12*, 115–124.

Colombo, V., Fernández-de-Heredia, M., and Malpartida, F. (2001). A polyketide biosynthetic gene cluster from *Streptomyces antibioticus* includes a LysR-type transcriptional regulator. Microbiology *147*, 3083–3092.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Craig, M., Lambert, S., Jourdan, S., Tenconi, E., Colson, S., Maciejewska, M., Ongena, M., Martin, J.F., van Wezel, G., and Rigali, S. (2012). Unsuspected control of siderophore production by N-acetylglucosamine in streptomycetes. Environ Microbiol Rep *4*, 512–521.

Craney, A., Ozimok, C., Pimentel-Elardo, S.M., Capretta, A., and Nodwell, J.R. (2012). Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. Chem. Biol. *19*, 1020–1027.

Cruz-Morales, P., Vijgenboom, E., Iruegas-Bocardo, F., Girard, G., Yáñez-Guerra, L.A., Ramos-Aboites, H.E., Pernodet, J.L., Anné, J., van Wezel, G.P., and Barona-Gómez, F. (2013). The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. Genome Biol. Evol. *5*, 1165–1175.

Feitelson, J.S., Malpartida, F., and Hopwood, D.A. (1985). Genetic and biochemical characterization of the *red* gene cluster of *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. *131*, 2431–2441.

Florea, B.I., Verdoes, M., Li, N., van der Linden, W.A., Geurink, P.P., van den Elst, H., Hofmann, T., de Ru, A., van Veelen, P.A., Tanaka, K., et al. (2010). Activity-based profiling reveals reactivity of the murine thymoproteasome-specific subunit beta5t. Chem. Biol. *17*, 795–801.

Gubbens, J., Janus, M., Florea, B.I., Overkleeft, H.S., and van Wezel, G.P. (2012). Identification of glucose kinase-dependent and -independent pathways for carbon control of primary metabolism, development and antibiotic production in *Streptomyces coelicolor* by quantitative proteomics. Mol. Microbiol. *86*, 1490–1507.

Herzner, A.M., Dischinger, J., Szekat, C., Josten, M., Schmitz, S., Yakéléba, A., Reinartz, R., Jansen, A., Sahl, H.G., Piel, J., and Bierbaum, G. (2011). Expression of the lantibiotic mersacidin in *Bacillus amyloliquefaciens* FZB42. PLoS ONE *6*, e22389.

Hopwood, D.A. (1997). Genetic Contributions to Understanding Polyketide Synthases. Chem. Rev. *97*, 2465–2498.

Hopwood, D.A. (2007). *Streptomyces* in Nature and Medicine: The Antibiotic Makers. (New York: Oxford University Press).

Hopwood, D.A., Chater, K.F., and Bibb, M.J. (1995). Genetics of antibiotic production in *Streptomyces coelicolor* A3(2), a model streptomycete. Biotechnology *28*, 65–102.

Ikeda, H., Kazuo, S.Y., and Omura, S. (2014). Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. J. Ind. Microbiol. Biotechnol. *41*, 233–250.

Keller, U., Lang, M., Crnovcic, I., Pfennig, F., and Schauwecker, F. (2010). The actinomycin biosynthetic gene cluster of *Streptomyces chrysomallus*: A genetic hall of mirrors for synthesis of a molecule with mirror symmetry. J. Bacteriol. *192*, 2583–2595.

Kersten, R.D., Yang, Y.L., Xu, Y., Cimermancic, P., Nam, S.J., Fenical, W., Fischbach, M.A., Moore, B.S., and Dorrestein, P.C. (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. Nat. Chem. Biol. *7*, 794–802.

Kersten, R.D., Ziemert, N., Gonzalez, D.J., Duggan, B.M., Nizet, V., Dorrestein, P.C., and Moore, B.S. (2013). Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. Proc. Natl. Acad. Sci. USA *110*, E4407–E4416.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000). Practical *Streptomyce*s Genetics. (Norwich: John Innes Foundation).

Kim, H.K., Choi, Y.H., and Verpoorte, R. (2010). NMR-based metabolomic analysis of plants. Nat. Protoc. *5*, 536–549.

Knappe, T.A., Linne, U., Zirah, S., Rebuffat, S., Xie, X., and Marahiel, M.A. (2008). Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of *Burkholderia thailandensis* E264. J. Am. Chem. Soc. *130*, 11446–11454.

Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozone, I., Izumikawa, M., Hashimoto, J., Takagi, M., Omura, S., Shin-ya, K., et al. (2013). Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth. Biol. *2*, 384–396.

Kovanich, D., Cappadona, S., Raijmakers, R., Mohammed, S., Scholten, A., and Heck, A.J. (2012). Applications of stable isotope dimethyl labeling in quantitative proteomics. Anal. Bioanal. Chem. *404*, 991–1009.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12.

Lautru, S., Deeth, R.J., Bailey, L.M., and Challis, G.L. (2005). Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nat. Chem. Biol. *1*, 265–269.

Lessmann, H., Krupa, J., Lackner, H., and Jones, P.G. (1989). New juglomycins. Z. Naturforsch. B *44*, 353–363.

Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. Nat. Rev. Genet. *12*, 671–682.

Meier, J.L., Niessen, S., Hoover, H.S., Foley, T.L., Cravatt, B.F., and Burkart, M.D. (2009). An orthogonal active site identification system (OASIS) for proteomic profiling of natural product biosynthesis. ACS Chem. Biol. *4*, 948–957.

Metsä-Ketelä, M., Oja, T., Taguchi, T., Okamoto, S., and Ichinose, K. (2013). Biosynthesis of pyranonaphthoquinone polyketides reveals diverse strategies for enzymatic carbon-carbon bond formation. Curr. Opin. Chem. Biol. *17*, 562–570.

Mo, S., Sydor, P.K., Corre, C., Alhamadsheh, M.M., Stanley, A.E., Haynes, S.W., Song, L., Reynolds, K.A., and Challis, G.L. (2008). Elucidation of the *Streptomyces coelicolor* pathway to 2-undecylpyrrole, a key intermediate in undecylprodiginine and streptorubin B biosynthesis. Chem. Biol. *15*, 137–148.

Mortensen, P., Gouw, J.W., Olsen, J.V., Ong, S.E., Rigbolt, K.T., Bunkenborg, J., Cox, J., Foster, L.J., Heck, A.J., Blagoev, B., et al. (2010). MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. J. Proteome Res. *9*, 393–403.

Owen, J.G., Reddy, B.V., Ternei, M.A., Charlop-Powers, Z., Calle, P.Y., Kim, J.H., and Brady, S.F. (2013). Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. Proc. Natl. Acad. Sci. USA *110*, 11797–11802.

Ozawa, M., Taguchi, T., Itoh, T., Ebizuka, Y., Booker-Milburn, K.I., Stephenson, G.R., and Ichinose, K. (2003). Structure and biosynthetic implication of (S)-NHAB, a novel shunt product, from a disruptant of the *actVI*-ORFA gene for actinorhodin biosynthesis in *Streptomyces coelicolor* A3(2). Tetrahedron *59*, 8793–8798.

Payne, D.J., Gwynn, M.N., Holmes, D.J., and Pompliano, D.L. (2007). Drugs for bad bugs: Confronting the challenges of antibacterial discovery. Nat. Rev. Drug Discov. *6*, 29–40.

Qu, X.D., Lei, C., and Liu, W. (2011). Transcriptome mining of active biosynthetic pathways and their associated products in *Streptomyces flaveolus*. Angew. Chem. Int. Ed. Engl. *50*, 9651–9654.

Rice, L.B. (2008). Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE. J. Infect. Dis. *197*, 1079–1081.

Rigali, S., Nothaft, H., Noens, E.E., Schlicht, M., Colson, S., Müller, M., Joris, B., Koerten, H.K., Hopwood, D.A., Titgemeyer, F., and van Wezel, G.P. (2006). The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. Mol. Microbiol. *61*, 1237–1251.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A.W., Hopwood, D.A., and van Wezel, G.P. (2008). Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. EMBO Rep. *9*, 670–675.

Sánchez, S., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., Rocha, D., Sánchez, B., Avalos, M., Guzmán-Trampe, S., et al. (2010). Carbon source regulation of antibiotic production. J. Antibiot. *63*, 442–459.

Schley, C., Altmeyer, M.O., Swart, R., Müller, R., and Huber, C.G. (2006). Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. J. Proteome Res. *5*, 2760–2768.

Świątek, M.A., Gubbens, J., Bucca, G., Song, E., Yang, Y.H., Laing, E., Kim, B.G., Smith, C.P., and van Wezel, G.P. (2013). The ROK family regulator Rok7B7 pleiotropically affects xylose utilization, carbon catabolite repression, and antibiotic production in *streptomyces coelicolor*. J. Bacteriol. *195*, 1236–1248.

Taguchi, T., Yabe, M., Odaki, H., Shinozaki, M., Metsä-Ketelä, M., Arai, T., Okamoto, S., and Ichinose, K. (2013). Biosynthetic conclusions from the functional dissection of oxygenases for biosynthesis of actinorhodin and related *Streptomyces* antibiotics. Chem. Biol. *20*, 510–520.

Udwary, D.W., Gontang, E.A., Jones, A.C., Jones, C.S., Schultz, A.W., Winter, J.M., Yang, J.Y., Beauchemin, N., Capson, T.L., Clark, B.R., et al. (2011). Significant natural product biosynthetic potential of actinorhizal symbionts of the genus *frankia*, as revealed by comparative genomic and proteomic analyses. Appl. Environ. Microbiol. *77*, 3617–3625.

Van Lanen, S.G., and Shen, B. (2006). Microbial genomics for the improvement of natural product discovery. Curr. Opin. Microbiol. *9*, 252–260.

van Wezel, G.P., and McDowall, K.J. (2011). The regulation of the secondary metabolism of *Streptomyces*: New links and experimental advances. Nat. Prod. Rep. *28*, 1311–1333.

Willey, J.M., and van der Donk, W.A. (2007). Lantibiotics: Peptides of diverse structure and function. Annu. Rev. Microbiol. *61*, 477–501.

Wilson, M.C., and Piel, J. (2013). Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chem. Biol. *20*, 636–647.

Zhong, X., Tian, Y., Niu, G., and Tan, H. (2013). Assembly and features of secondary metabolite biosynthetic gene clusters in *Streptomyces ansochromogenes*. Sci. China Life Sci. *56*, 609–618.

Zhu, H., Sandiford, S.K., and van Wezel, G.P. (2014). Triggers and cues that activate antibiotic production by actinomycetes. J. Ind. Microbiol. Biotechnol. *41*, 371–386.

