



# A systems approach using OSMAC, Log P and NMR fingerprinting: An approach to novelty



Miaomiao Liu<sup>a</sup>, Tanja Grkovic<sup>a</sup>, Xueting Liu<sup>b</sup>, Jianying Han<sup>b,c</sup>, Lixin Zhang<sup>b,\*\*</sup>, Ronald J. Quinn<sup>a,\*</sup>

<sup>a</sup> Griffith Institute for Drug Discovery, Griffith University, Brisbane, QLD, 4111, Australia

<sup>b</sup> Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

<sup>c</sup> Graduate University of Chinese Academy of Sciences, Beijing, 100049, China

## ARTICLE INFO

### Article history:

Received 31 May 2017

Received in revised form

3 September 2017

Accepted 9 October 2017

### Keywords:

NMR fingerprints

OSMAC

Metabolites

## ABSTRACT

The growing number of sequenced microbial genomes has revealed a remarkably large number of secondary metabolite biosynthetic clusters for which the compounds are still unknown. The aim of the present work was to apply a strategy to detect newly induced natural products by cultivating microorganisms in different fermentation conditions. The metabolomic analysis of 4160 fractions generated from 13 actinomycetes under 32 different culture conditions was carried out by <sup>1</sup>H NMR spectroscopy and multivariate analysis. The principal component analysis (PCA) of the <sup>1</sup>H NMR spectra showed a clear discrimination between those samples within PC1 and PC2. The fractions with induced metabolites that are only produced under specific growth conditions was identified by PCA analysis. This method allows an efficient differentiation within a large dataset with only one fractionation step. This work demonstrates the potential of NMR spectroscopy in combination with metabolomic data analysis for the screening of large sets of fractions.

© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The impact of natural products on human beings has been enormous, and their study continues to influence research in the fields of chemistry, biology, and ecology [1]. Compared with synthetic and combinatorial compounds, their chemical structures are more diverse and complex, and they have often demonstrated selective activities in various biological systems [2]. Moreover, microorganisms are a potential source of structurally diverse bioactive metabolites and have yielded some of the most important drugs [3]. Among the 11 currently used nature-derived TB drugs, 7 of them were either isolated from microbes or semi-synthesized from a microbial natural product.

Historically, the increasing difficulty to find new chemical entities via high-throughput screening has led to a decline in antibiotic research, while infectious diseases associated with multidrug

resistance are spreading rapidly [4]. While many microorganisms produce secondary metabolites, the actinomycetes, in particular the bacterial genus *Streptomyces*, are an especially rich source [5]. The difficulty is that many of these gene clusters are expressed at low levels or only expressed under very specific conditions during growth in the laboratory. As a consequence of silent biosynthetic genes, the traditional methods to discover microbial natural products, which usually involve the collection and cultivation of strains, extraction, bioassay-guided isolation and structure elucidation, is often unsatisfactory. It has been demonstrated that many silent biosynthetic genes are only activated under specific conditions [6]. Several sequencing projects have also confirmed that microbes have the potential to produce a lot more compounds from a single strain than previously known [7–9]. To gain access to this untapped reservoir of potentially bioactive structures, the biosynthesis of these putative metabolites needs to be induced. Even small changes in the culture medium may not only impact the quantity of a certain compound but also the general metabolic profile of an organism [10,11]. In order to increase the number of secondary metabolites available from one microbial source, Zeek and co-workers have investigated the systematic alteration of

\* Corresponding author.

\*\* Corresponding author.

E-mail address: [r.quinn@griffith.edu.au](mailto:r.quinn@griffith.edu.au) (R.J. Quinn).

Peer review under responsibility of KeAi Communications Co., Ltd.

easily accessible cultivation parameters (for example, media composition, aeration, culture vessel, addition of enzyme inhibitors) on the production of secondary metabolites [12]. As a result, the OSMAC (One Strain - Many Compounds) approach was developed as a way of revealing nature's chemical diversity and to increase the number and yield of natural products from a single microbe.

Additionally, for microbes whose chemical diversity was not determined prior to extraction, there is a high rate of rediscovery. Therefore, we hypothesized that a chemo-informatics method based on secondary metabolite production would be more valuable and would greatly increase the value of a screening library for HTS. Metabolomics is the global measurement of the small molecule metabolites in a biological system and reflects the phenotype of (and is therefore complementary to) its underlying genomic, transcriptomic, and proteomic networks [13]. Metabolomics research typically implements analytical tools such as LC/MS and NMR to globally measure small molecule metabolites [14–17]. Combining principal component analysis (PCA) with NMR provides a visual representation of variance between NMR profiles. We hypothesized that microbes producing the same secondary metabolites would group together, whereas those producing different metabolites under different OSMAC conditions would be separated, thereby providing a method to select the best condition for a specific bacteria having distinct chemistries without having to identify each component of their corresponding extracts. A major distinction between the work presented here and other metabolomics studies is that we only focused on those induced secondary metabolites synthesized under specific culture conditions.

Here, we report the development of a practical method to survey biosynthetic potential in microorganisms, thereby identifying the most promising fractions and prioritizing them for natural product discovery. Central to our approach is to evaluate OSMAC-NMR-PCA based secondary metabolomics to more broadly investigate secondary metabolites from actinomycetes to assist with culture condition selection/dereplication to support drug discovery efforts, to discover new natural products, and to study regulation of secondary metabolite production (Fig. 1).

We used 13 representative extreme environment-associated strains from our actinomycete collection, that produced secondary metabolites or not under our standard culture condition

(unpublished data), to develop and validate the method. As a result, 13 actinomycetes were cultured under 32 specific conditions, and 832 crude extracts and 4160 fractions were generated after extraction and fractionation processes. With the use of combinational analysis tools, 37 lead fractions were finally selected and from which, several natural products with unique structure and activity were identified. Variations of this method should be applicable to other actinomycetes for the discovery of unique natural products.

## 2. Material and methods

### 2.1. General experimental procedures

NMR spectra were recorded in DMSO- $d_6$  ( $\delta_H$  2.50 and  $\delta_C$  39.5) at 25 °C on a Bruker Avance HDX 800 MHz spectrometer equipped with a TCI cryoprobe. The HPLC system included a Waters 600 pump fitted with a 996 photodiode array detector and Gilson FC204 fraction collector. All solvents used for extraction and chromatography were Lab-Scan HPLC grade, and the H<sub>2</sub>O was Millipore Milli-Q PF filtered.

### 2.2. Bacterial material

Endophytes ES120055 and ES120127 were isolated from the Traditional Chinese Medicine (TCM) plant *Saxifragaceae Astilbe* and *Cirsium shansiense* samples collected from Yunnan Province. Two desert strains LS120167 and LS120194 were collected from Taklimakan Desert. All the marine actinomycetes were isolated from sediment samples collected from South China Sea.

### 2.3. Fermentation

Thirteen strains were cultivated on an ISP2 agar plate at 28 °C for 7 days. A 250 mL Erlenmeyer flask containing 40 mL of ISP2 liquid medium was inoculated with each strain and incubated at 28 °C (220 rpm) for 48 h. Aliquots (2 mL) of the pre-culture were used to inoculate 4 × 250 mL Erlenmeyer flasks, each containing 40 mL of ISP2 liquid medium and the flasks were incubated at 28 °C (220 rpm) for 3 days. Aliquots (2 mL) of the seed cultures were aseptically transferred to 2 × 250 mL Erlenmeyer flasks, each

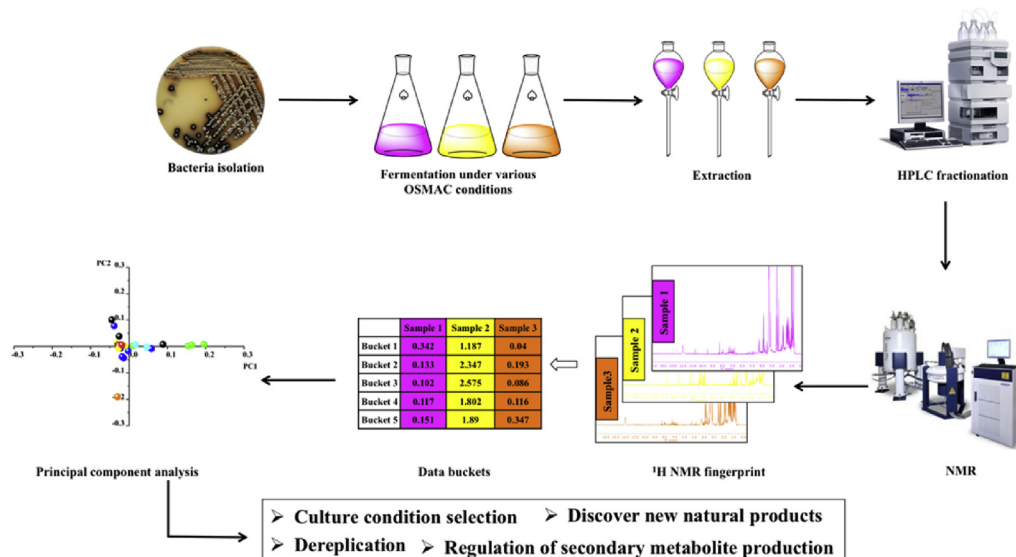


Fig. 1. Flowchart for OSMAC-NMR-PCA-based secondary metabolomics.

containing 40 mL of media (various media for strategies 1 and 2, AM2 or modified media for strategies 3 and 4), and the flasks were incubated at 28 °C (or specific temperatures for strategy 3), 220 rpm for 7 days (or 14 days for strategy 2). The broths were combined and centrifuged to yield a supernatant and a mycelial cake of each condition.

#### 2.4. Media

ISP2 (1 L): yeast extract 4.0 g, malt extract 10.0 g, dextrose 4.0 g, agar 20.0 g, pH 7.2.

R-1 (1 L): D –mannitol 20.0 g, D –glucose 20.0 g, yeast extract 5.0 g, peptone 10.0 g, KH<sub>2</sub>PO<sub>4</sub> 0.5 g, MgSO<sub>4</sub> 0.3 g, corn syrup 1.0 g.

O-1 (1 L): sucrose 3.0 g, NaNO<sub>3</sub> 0.3 g, K<sub>2</sub>HPO<sub>4</sub> 0.1 g, KCl 0.05 g, FeSO<sub>4</sub> 0.001 g, MgCl<sub>2</sub> 0.4 g.

T-1 (1 L): yeast extract 4.0 g, malt extract 10.0 g, glucose 10.0 g, pH 7.0.

T-2 (1 L): glucose 60.0 g, yeast extract 2.0 g, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 2.0 g, MgSO<sub>4</sub>·7H<sub>2</sub>O 0.1 g, K<sub>2</sub>HPO<sub>4</sub> 0.5 g, NaCl 2.0 g, FeSO<sub>4</sub>·7H<sub>2</sub>O 0.05 g, ZnSO<sub>4</sub>·7H<sub>2</sub>O 0.05 g, MnSO<sub>4</sub>·4H<sub>2</sub>O 0.05 g, CaCO<sub>3</sub> 5.0 g, pH 7.0.

T-3 (1 L): starch 10.0 g, yeast extract 4.0 g, peptone 2.0 g, CaCO<sub>3</sub> 1.0 g, Fe<sub>2</sub>(SO<sub>4</sub>)<sub>3</sub>·4H<sub>2</sub>O 0.04 g, KBr 0.1 g.

T-4 (1 L): soluble starch 24.0 g, meat extract 30.0 g, tryptone 5.0 g, yeast extract 5.0 g, glucose 1.0 g, calcium carbonate 2.0 g, pH 7.4.

T-5 (1 L): soluble starch 15.0 g, soybean meal 5.0 g, peptone 15.0 g, glycerol 15.0 g, CaCO<sub>3</sub> 2.0 g, pH 7.4.

A-1 (1 L): soluble starch 10.0 g, yeast extract 10.0 g, KH<sub>2</sub>PO<sub>4</sub> 0.5 g, corn syrup 3.0 g, glucose 20.0 g, MgSO<sub>4</sub>·7H<sub>2</sub>O 5.0 g, beef extract 3.0 g, CaCO<sub>3</sub> 2.0 g, pH 7.0.

A-2 (1 L): starch 10.0 g, glucose 10.0 g, glycerol 10.0 mL, Polypepton 5.0 g, yeast extract 2.0 g, NaCl 1.0 g, CaCO<sub>3</sub> 3.2 g, corn steep liquor 1.0 mL, pH 7.4.

PKS-1 (1 L): soluble starch 20.0 g; glucose 10.0 g; peptone 5.0 g; yeast extract 5.0 g; NaCl 4.0 g; K<sub>2</sub>HPO<sub>4</sub> 0.5 g; MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.5 g; CaCO<sub>3</sub> 2.0 g.

PKS-2 (1 L): soybean powder 20.0 g, mannitol 20.0 g.

M-1 (1 L): glucose 10.0 g, maltose 10.0 g, yeast extract 3.0 g, CaCl<sub>2</sub> 0.15 g, MgCl<sub>2</sub> 0.2 g, soybean cake meal 30.0 g, pH 7.2.

M-2 (1 L): starch 24.0 g, glucose 1.0 g, peptone 3.0 g, meat extract 3.0 g, yeast extract 5.0 g, CaCO<sub>3</sub> 4.0 g, trace metals 5.0 mL, pH 7.0.

M-3 (1L): potatoes (sliced washed unpeeled) 200.0 g, dextrose 20.0 g.

M-4 (1 L): starch 20.0 g, glycerol 8.7 g, soybean meal 10.0 g, ZnSO<sub>4</sub>·7H<sub>2</sub>O 0.01 g, CuSO<sub>4</sub>·5H<sub>2</sub>O 0.005 g, FeSO<sub>4</sub>·7H<sub>2</sub>O 0.01 g.

M-5 (1 L): Soybean flour 5.0 g, soluble starch 15.0 g, glycerol 15.0 g, peptone 15.0 g, CaCO<sub>3</sub> 2.0 g, pH 7.4.

E-1 (1 L): glucose 10.0 g, dextrin 40.0 g, Bactosoytone 25.0 g, yeast extract 1.0 g, CaCO<sub>3</sub> 3.0 g, pH 7.0.

D-1 (1 L): glucose 6.0 g, yeast extract 4.0 g, pH 7.0.

D-2 (1 L): KH<sub>2</sub>PO<sub>4</sub> 2.0 g, NH<sub>4</sub>Cl 1.5 g, MgSO<sub>4</sub>·7H<sub>2</sub>O 0.5 g, NaCl 0.5 g, glycerol 10 g, myoinositol 0.4 g, monosodium L –glutamate onohydrate 5.0 g, NaF 0.084 g, FeSO<sub>4</sub>·7H<sub>2</sub>O 0.025 g, ZnSO<sub>4</sub>·7H<sub>2</sub>O 0.01 g, CoCl<sub>2</sub>·6H<sub>2</sub>O 0.01 g, CaCO<sub>3</sub> 0.25 g, *p*-aminobenzoic acid 0.001 g, pH 7.0.

D-3 (1 L): oatmeal 20.0 g, trace salts solution 1.0 mL, pH 7.3.

B-1 (1 L): glucose 10.0 g, soluble amyllum 40.0 g, yeast extract 5.0 g, soybean powder 25.0 g, peptone 5.0 g, CaCO<sub>3</sub> 2.0 g, MgSO<sub>4</sub>·7H<sub>2</sub>O 8.0 g, FeSO<sub>4</sub>·7H<sub>2</sub>O 6.0 g, ZnSO<sub>4</sub>·7H<sub>2</sub>O 2.0 g, MnSO<sub>4</sub>·H<sub>2</sub>O 2.0 g, CoCl<sub>2</sub>·6H<sub>2</sub>O 0.5 g, Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O 2.0 g, pH 7.0.

Trace Salts Solution (100 mL): FeSO<sub>4</sub>·7H<sub>2</sub>O 0.1 g, MnCl<sub>2</sub>·4H<sub>2</sub>O 0.1 g, ZnSO<sub>4</sub>·7H<sub>2</sub>O 0.1 g.

#### 2.5. Lead-like enhanced (LLE) fraction library generation

Each LLE extract was fractionated into 5 LLE fractions which were used for <sup>1</sup>H NMR analysis. The lead-like fraction library generation has been conducted as previously described. A small amount (10 mg) of dried samples were suspended in DMSO (300 μL). The extract (100 μL) was then fractionated by HPLC using a C<sub>18</sub> Phenomenex Onyx monolithic HPLC analytical column (100 × 4.6 mm) with solvent conditions consisting of a linear gradient from 90% H<sub>2</sub>O (0.1% TFA)/10% MeOH (0.1% TFA) to 50% H<sub>2</sub>O (0.1% TFA)/50% MeOH (0.1% TFA) in 3 min at a flow rate of 4 mL/min, followed by a convex gradient to MeOH (0.1% TFA) in 3.5 min at a flow rate of 3 mL/min. This was held at 100% MeOH (0.1% TFA) for 0.5 min at a flow rate of 3 mL/min and for further 1 min at a flow rate of 4 mL/min, then a linear gradient back to 90% H<sub>2</sub>O (0.1% TFA)/10% MeOH (0.1% TFA) in 1 min at a flow rate of 4 mL/min was applied. Finally, the gradient was held at 90% H<sub>2</sub>O (0.1% TFA)/10% MeOH (0.1% TFA) for 2 min at a flow rate of 4 mL/min, ready for the next injection. Total run time for each injection was 11 min, and 5 fractions were collected between 2.0 min and 7.0 min, these include: fraction 1 (time = 2.01–3.00 min), fraction 2 (time = 3.01–4.00 min), fraction 3 (time = 4.01–5.00 min), fraction 4 (5.01–6.00 min), fraction 5 (6.01–7.00 min).

#### 2.6. Metabolic fingerprinting methodology

Three combined replicates of each of the 5 LLE fractions were analyzed by <sup>1</sup>H NMR. The samples were dissolved in 600 μL of DMSO-*d*<sub>6</sub> and run in a 5 mm NMR tube. For each sample, the following parameters were applied, pw = 30°, p1 = 9.250 μs, d2 = 0 s, d1 = 1 s, at = 2.04 s, sw = 20.03 ppm, nt = 128 scans.

#### 2.7. Anti-BCG assay

The BCG used was a *M. bovis* BCG 1173P2 strain transformed with green fluorescent protein (GFP) constitutive expression plasmid pUV3583c with direct readout of fluorescence as a measure of bacterial growth. BCG was grown at 37 °C to mid log phase in Middle brook 7H9 broth (Becton Dickinson) supplemented with 10% OADC enrichment (Becton Dickinson) 0.05% tween-80 and 0.2% glycerol, which then adjusted to OD<sub>600</sub> = 0.025 with culture medium as bacterial suspension. Aliquots (80 μL) of the bacterial suspension were added to each well of the 96-well microplates (clear flat-bottom), followed by adding compounds (2 μL in DMSO), which were serially twofold diluted. Isoniazid served as positive control and DMSO as negative control. The plate was incubated at 37 °C for 3 days, and GFP fluorescence was measured with Multi-label Plate Reader using the bottom read mode, with excitation at 485 nm and emission at 535 nm. MIC is defined as the minimum concentration of drug that inhibits more than 90% of bacterial growth reflected by fluorescence value.

#### 2.8. Data processing and multivariate analysis

The interesting regions of the NMR spectra for the analysis of the LLE fractions were identified to be from 0 to 15 ppm. The spectra were phase- and baseline-corrected manually by software Topspin. The <sup>1</sup>H NMR spectra were automatically reduced to ASCII files using AMIX (v.3.7, Bruker Biospin). Spectral intensities were scaled to the biggest peak (δ<sub>H</sub> = 2.50 ppm) and reduced to integrated regions of equal width (0.02 ppm) corresponding to the region of 0.0–15.00. The region of 2.4–3.5 was excluded from the analysis because of residual signal of DMSO and water. After a bucket table was calculated, PCA was applied by the Bruker AMIX software. Because the variables (NMR chemical shifts) in this project are comparable,

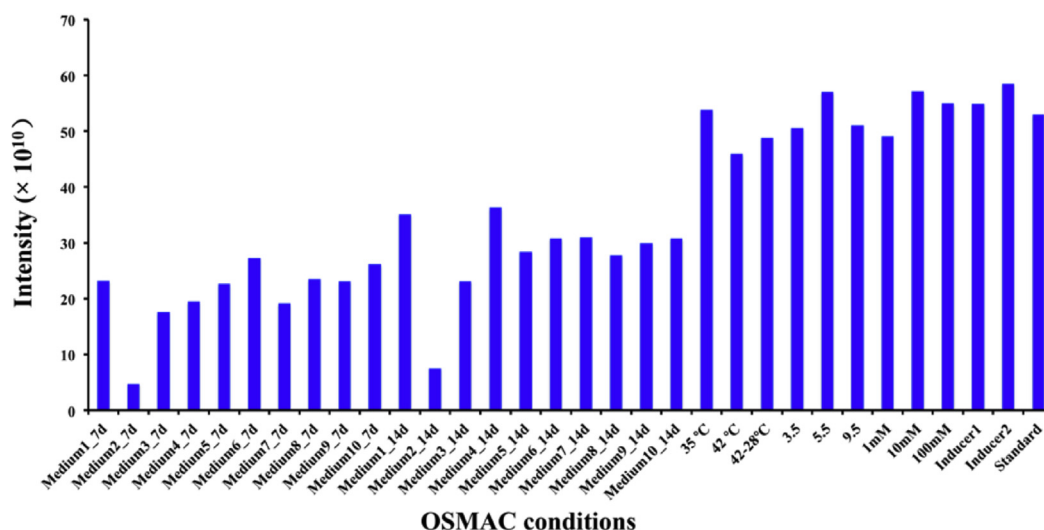


Fig. 2. Distribution of total NMR intensities among 32 OSMAC conditions as an indicator of metabolite production.

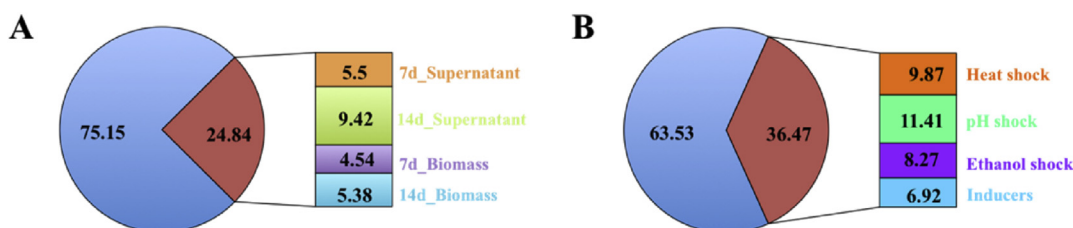


Fig. 3. Pie chart of anti-BCG activities of 1271 active LLE fractions (A. 646 active fractions from media variation, B. 625 active fractions from stress conditions).

“No scaling” for the scaling of columns in the bucket table were selected. It preserves natural differences in intensities and it highlights dominant effects. To get rid of those columns in the bucket table that contain almost constant and small values that would not produce a large variance and would not have a major influence on the PCA, the minimum variance level was set as 5%, which meant variables that produce less than 5% variance of the maximum variance were removed from the PCA calculation. Confidence level was set as 95% to give scores plot and loadings plot.

### 3. Results and discussion

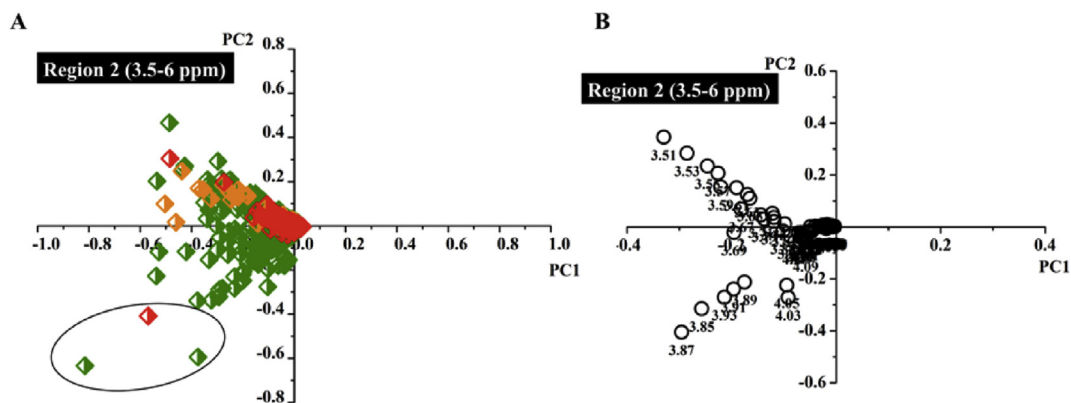
#### 3.1. Rationale and design of the method

In our effort to assemble a natural product library, actinomycetes from various extreme environments were fermented for activity screening and natural product isolation. Each strain was fermented in three media, and the crude extracts were subjected to HPLC chemical profiling and NMR fingerprinting (data not shown). From this initial effort, we found the metabolite profile of a given strain was largely medium dependent, and with the three media under the fermentation conditions examined, only around 15% of the strains showed promising metabolite profiles deemed worthy of pursuing subsequent natural product isolation. However, advances in microbial genomics have unequivocally demonstrated that we are missing ~90% of the natural product biosynthetic capacity of even the workhorse producers, the actinomycetes [18]. Thus it may not be enough to prioritize the strains for subsequent natural product discovery based on chemical profiling of the extracts under culture conditions alone. Clearly, a more systematic and effective method is needed to rapidly survey the biosynthetic

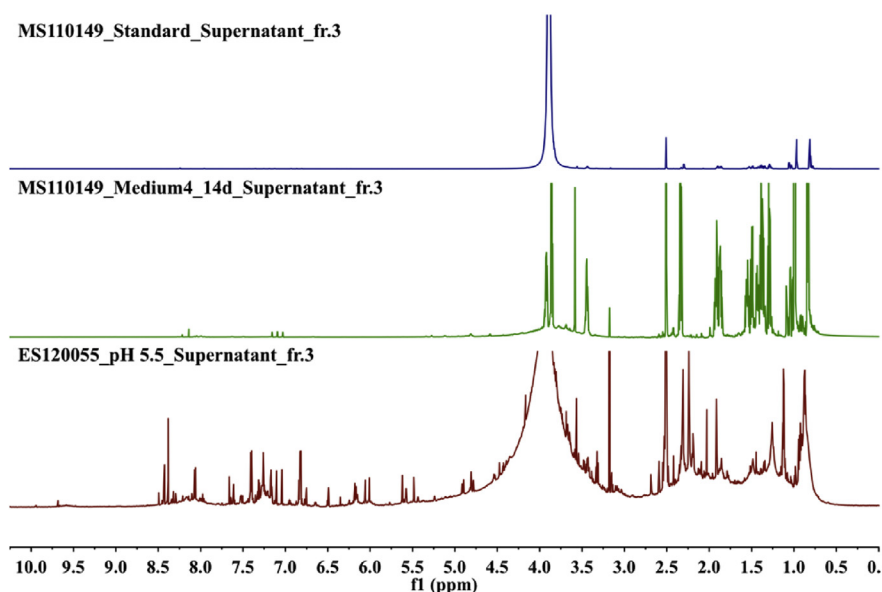
potential of a strain. The best culture condition that can induce cryptic biosynthetic pathways for novel natural product can then be identified, and prioritized with a metabolomics-based approach. This could fundamentally change the way in which microbial natural products are discovered.

The OSMAC approach has resulted from the observation that very small changes in the cultivation conditions can completely shift the metabolic profile of many microorganisms [19]. Classically, systematic variations of culture media have been used to activate the production of additional secondary metabolites from microbes [20]. There is considerable evidence that some stress responses can also be used to trigger the expression of secondary metabolic genes: two that have been applied widely are heat shock and ethanol shock [21]. It was reported both stresses act through damage to the cell envelope and by driving the accumulation of misfolded or unfolded proteins [22,23]. Moreover, varying additional cultivation parameters such as pH, oxygen supply, the addition of the enzyme inhibitors and inducers or the selection of drug-resistant bacteria mutants was also found to be effective in maximizing the diversity of microbial natural products [11,20,24–26]. Thus, four kinds of OSMAC conditions were developed in this study, namely media variation, culture period variation, stress condition, and the addition of inducer molecules.

LLE HPLC fractionation provides a second log P filter allowing any remaining high log P components to be excluded [27]. To reduce the number of fractions and decrease the labor intensity in the following process, all the extracts were collected for 5 fractions every minute from 2 to 7 min. Data acquisition for both LC fractionation and NMR determination were in batches of 96 samples run overnight using automation. While mass spectrometry is more sensitive, it does not intrinsically provide enough structural



**Fig. 4.** Scores plots and loadings plots of PC1 and PC2 of 4160 fractions, fractions from endophytes, land strains and marine strains in scores plots were shown in red, orange and green, respectively. Bucket values in loadings plots were shown as numbers below each circle. A. Scores plot of region 2 (3.5–6 ppm). B. Loadings plot of region 2 (3.5–6 ppm).



**Fig. 5.**  $^1\text{H}$  NMR spectra of 3 outliers indicated by PCA results of region 2.

information about the class of the small molecules of the fractions. Metabolite fingerprinting by NMR is a fast, convenient, and effective tool for discriminating between groups of related samples and it could identify the most important regions of the spectrum for further analysis.

PCA is an unsupervised method performed without using knowledge of sample class, which reduces the dimensionality of the data input while in a 2- or 3D map. By producing new linear combinations of the original variates, it plots data in order to indicate relationships between samples in a multidimensional space. It enables the easy comparison of microbe metabolic profiles. The NMR data from the LLE fractions of all the samples were subjected to PCA in order to highlight the differences between the LLE fractions and to identify the metabolites responsible for that distinction.

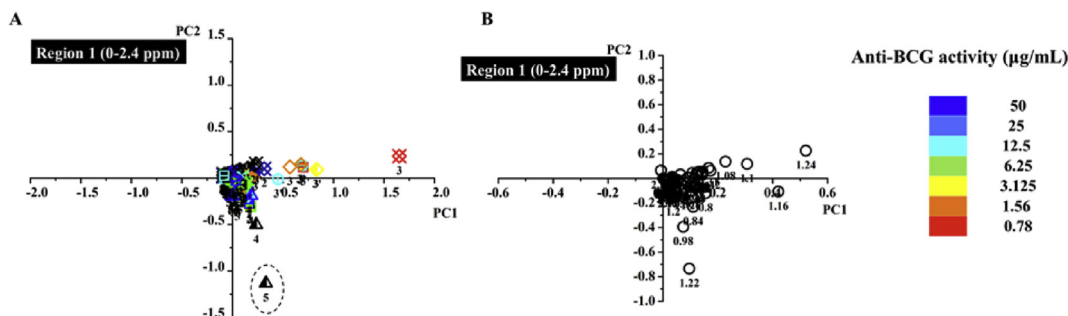
For capturing as many outliers as possible and decreasing the disturbance by differences in intensities between saturated and unsaturated areas, each spectrum was divided into 4 regions (0–2.4 ppm, 3.5–6 ppm, 6–10 ppm and 10–15 ppm) and analyzed separately. For electronic comparison of the data sets by multivariate methods it was important to ensure that there is as little

experimental variation as possible in the sample set. Each data set was automatically phased and baseline corrected, using the same processing parameters. After importation into AMIX (v.3.7, Bruker Biospin, Germany), all the negative peaks were removed. Before analysis by multivariate methods, data sets were reduced in complexity by using the “bucketing” function to generate a set number of integrated regions (0.05 ppm) of the data set. This table of ‘buckets’ data from those spectra could then be exported as a spreadsheet suitable for importation into statistical analysis software, such as AMIX or SIMCA-P.

The bucket table was then subjected to PCA, and the scores plots and the corresponding loadings plots were generated using AMIX. The scores plot presents a view of the variance in the data, and the further the groups separate, the more different the secondary metabolites. The loadings plot is geometrically related to the scores plot and describes the variance observed in the scores plot.

Besides, an anti-mycobacteria activity assay was conducted to screen the fractions and the activity profile provides another filter for fraction prioritization.





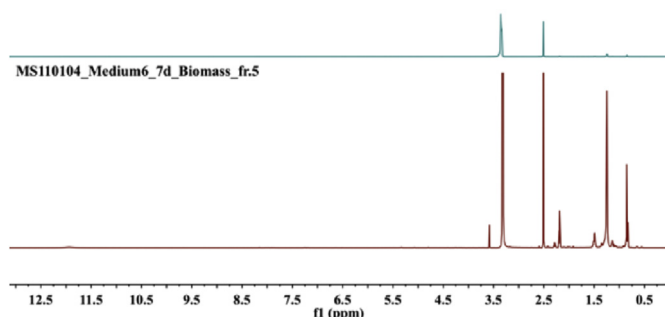
**Fig. 6.** PCA results of 320 fractions from MS110104. Anti-BCG activity was shown in different colors coded according to the legend (blue lowest and red highest activity). A. Scores plot of region 1 (0–2.4 ppm). Fraction numbers were shown below B. Loadings plot of region 1 (0–2.4 ppm). Bucket values in loadings plots were shown as numbers below each circle.

### 3.2. Validation of the method

We first selected 13 representative strains to develop the method, five that produced various classes of natural products, i.e., alkaloids, macrolides, terpenoids, peptides and macrodiolides, and 8 strains that has very low levels of secondary metabolite production (unpublished data). Each strain was cultured in 10 different media, each media for 2 culture periods (7 days and 14 days), 9 stress conditions (heat shock: 37 °C, 42 °C and 42 °C for an hour; pH shock: 3.5, 5.5 and 9.5; ethanol shock: 1 mM, 10 mM and 100 mM), 2 addition inducer molecules (L-homoserinelactone hydrochloride and N-carbobenzoxy-L-homoserinelactone), as well as the standard condition. Media was selected under the consideration of their biological origin and the reported production of secondary metabolites. We then established an effective small-scale extraction method for the HPLC-NMR-based screening and analysis. The extraction effects of solid (adsorbent DVB-NVP HLB) and liquid extraction (n-butanol or acetone) methods on supernatant and biomass samples were evaluated by HPLC and NMR. As a result, extraction solvents were determined as n-butanol or acetone for supernatant and biomass samples, respectively. After fermentation, centrifugation, extraction and HPLC fractionation, 4160 fractions were generated from 13 selected actinomycetes under 32 OSMAC conditions.

MS110104\_Standard\_Biomass\_fr.5

MS110104\_Medium6\_7d\_Biomass\_fr.5



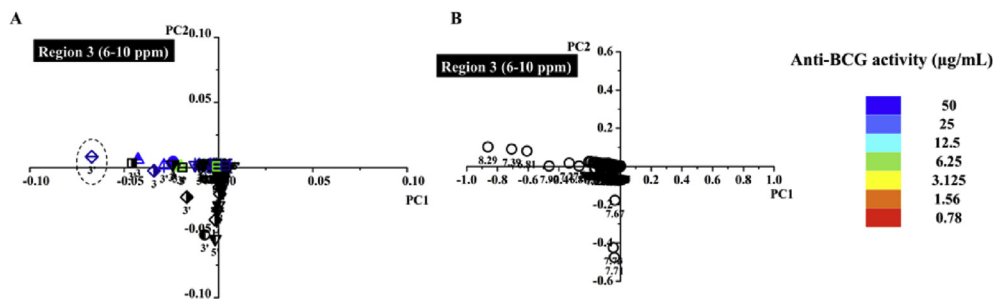
**Fig. 7.**  $^1\text{H}$  NMR spectra of the outlier MS110104\_Medium6\_7d\_Biomass\_fr.5 indicated by PCA result of region 1 and the corresponding standard fraction.

#### 3.2.1. Evaluation 1: investigation of effects from different OSMAC conditions on the total production of secondary metabolites

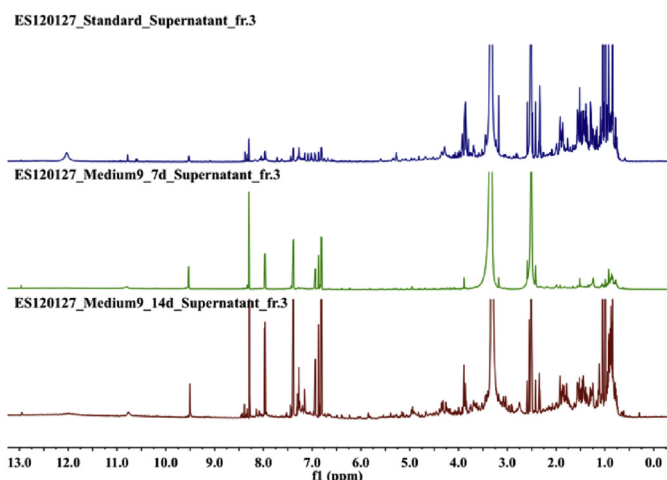
The effect of 32 OSMAC conditions on secondary metabolite production by 13 actinomycetes was studied. Total NMR intensity of each sample was calculated based on its bucket table generated from the NMR fingerprint (Fig. 2). In general, higher intensities were found in stress conditions compared to that of media variation conditions while longer cultivation time resulted in relatively higher NMR intensities, or in other words, higher secondary metabolite yields. The highest NMR intensities were found in media 1, 4 and 6, which are all nutrient rich media, while in contrast the oligotrophic medium (medium2) gave the lowest NMR intensity among all conditions, suggesting essential effects of nutritional status on the selected actinomycetes.

Temperature is an important environmental factor for cell growth and natural product biosynthesis in the fermentation process [28]. It has been reported that high temperature stress could induce reactive oxygen species (ROS) in *Taxus yunnanensis* [29]. ROS induction could enhance the production of many metabolites [30,31]. In the present study when the temperature was increased from 28 °C (standard) to 35 °C for 7 days, there was a slight enhancement of the production of the secondary metabolites (1.53% higher than standard). However, more extreme fermentation temperature (42 °C) for 7 days caused decreased secondary metabolite production (13.23% lower than standard), which may be due to the inhibition of cell growth under high temperature. In addition, keeping the culture at 42 °C for only 1 h after which the culture was reduced back to 28 °C also showed a marked decrease on the production of the secondary metabolites (7.84% lower than standard). Decreasing the pH from 7.5 to 5.5 saw a slight enhancement of secondary metabolites (1.53% higher than standard). However more acidic (3.5) or basic (9.5) pH shocks only resulted in lower productions, which could be due to the limited cell growth under those extreme conditions. Different addition concentrations of ethanol brought a change to the antibiotic production. In a range between 1 mM and 100 mM ethanol, addition of 10 mM ethanol to strain cultures enhanced the metabolites production the most (1.53% higher than standard). The effect of ethanol on metabolites biosynthesis may be due to its activity on changing the membrane structure, affecting steady-state growth and regulating related genes and carbon metabolism [32]. Both the two inducers (L-homoserinelactone hydrochloride and N-carbobenzoxy-L-homoserinelactone) lead to shifts on metabolites production, especially inducer 2 with a 10.47% higher NMR intensity than standard.

In conclusion, two selected inducer molecules, addition of a certain amount of ethanol, pH 5.5 and 35 °C culture temperature factors were observed to induce positive effects on the secondary metabolite production of selected microbes. However, extreme



**Fig. 8.** PCA results of 320 fractions from ES120127. Anti-BCG activity was shown in different colors coded according to the legend (blue lowest and red highest activity). A. Scores plot of region 3 (6–10 ppm). Fraction numbers were shown below each symbol. B. Loadings plot of region 3 (6–10 ppm). Bucket values in loadings plots were shown as numbers below each circle.

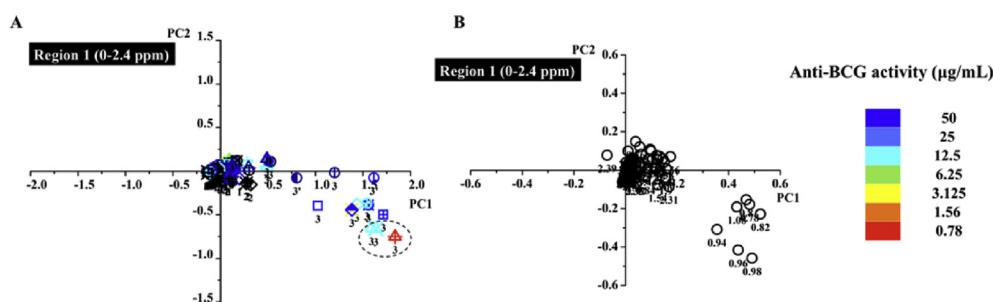


**Fig. 9.**  $^1\text{H}$  NMR spectra of the outlier ES120127\_Medium9\_14d\_Supernatant\_fr.3 indicated by PCA result of region 3 and the corresponding standard fraction and ES120127\_Medium9\_7d\_Supernatant\_fr.3.

520 crude extracts under different culture media and periods, 646 (24.84%) showed anti-BCG activity (Fig. 3A). Comparing to fractions of biomass extracts, supernatant fractions exhibited a significant higher hit rates, and longer cultivation time seems to produce more active fractions. The greatest effect was observed in the pH shock fractions, of which 11.41% displayed anti-BCG activity (Fig. 3B).

### 3.2.3. Evaluation 3: multivariate data analysis for fractions

After NMR fingerprinting and processing, 4160 NMR spectra were imported into AMIX and a data set consisted of a  $4160 \times 696$  matrix was generated, in which rows represented the samples (4160 LLE fractions), and columns represented the 696 buckets of the  $^1\text{H}$  NMR spectrum. Thus, each fraction was represented in the 696-dimensional space made of the 696 variables. Analysis of the PCA resulting scores plot (PC1 versus PC2) showed no large differences between the fractions in total region (0–15 ppm) and region 1 (0–2.4 ppm), and almost all the spots in the scores plot of region 4 (10–15 ppm) gathered in the same area. Therefore the focus was placed on the results obtained from the other regions. As seen in Fig. 4C and D (3.5–6 ppm), three fractions located in the



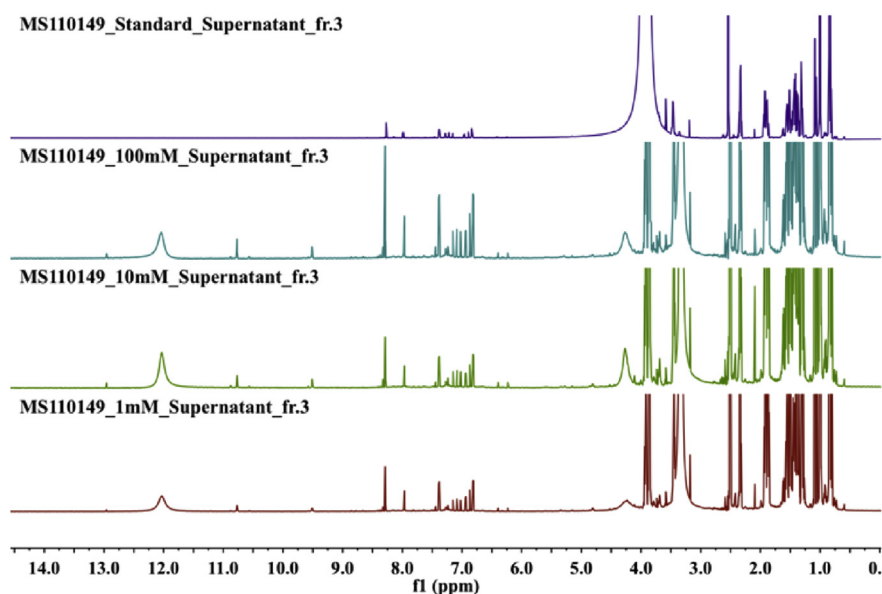
**Fig. 10.** PCA results of 320 fractions from MS110149. Anti-BCG activity was shown in different colors coded according to the legend (blue lowest and red highest activity). A. Scores plot of region 1 (0–2.4 ppm). Fraction numbers were shown below each symbol. B. Loadings plot of region 1 (0–2.4 ppm). Bucket values in loadings plots were shown as numbers below each circle.

culture conditions, such as 42 °C culture temperature, pH 3.5 or 5.5, the oligotrophic medium (medium2) all lead to significant decrease effect on secondary metabolite production.

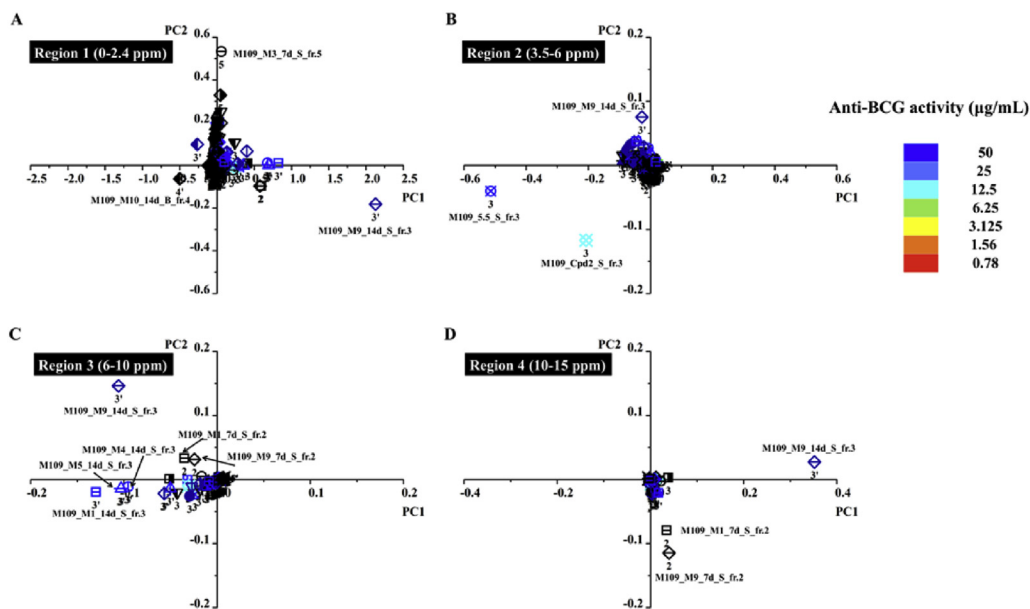
### 3.2.2. Evaluation 2: anti-BCG activity screening

Biological activity of the library including 4160 fractions was evaluated in terms of *Mycobacteria bovis* Bacillus Calmette–Guérin (BCG) inhibition. In this study, all fractions were screened with an initial concentration of 50 µg/mL. The results of the anti-TB assay show that 1271 out of the 4160 (30.55%) fractions possessed biological activity  $\leq 50$  µg/mL. Within 2600 fractions generated from

third quadrant were found not to cluster with any of the others, suggesting the possibility of existence of induced metabolites. Visual inspection of the corresponding  $^1\text{H}$  NMR spectra was carried out. However, only the fraction from the endophyte (ES120055\_pH 5.5\_Supernatant\_fr.3) displayed various signals in saturated as well as aromatic regions (Fig. 5). Not many signals could be found in the other two outliers from marine strains, especially MS110149\_Standard\_Supernatant\_fr.3. The NMR signals responsible for distinguishing them from the others are the large peaks located around 4 ppm, which may be produced during the sample preparation process, for example, the residual water in deuterated



**Fig. 11.**  $^1\text{H}$  NMR spectra of the outliers MS110149\_1mM\_Supernatant\_fr.3, MS110149\_10mM\_Supernatant\_fr.3 and MS110149\_100mM\_Supernatant\_fr.3 indicated by PCA result of region 1 and the corresponding standard fraction.



**Fig. 12.** Scores plots of 320 fractions from MS110109. Fraction numbers were shown below each symbol. Anti-BCG activity was shown in different colors coded according to the legend (blue lowest and red highest activity). A. Scores plot of region 1 (0–2.4 ppm). B. Scores plot of region 2 (3.5–6 ppm). C. Scores plot of region 3 (6–10 ppm). D. Scores plot of region 4 (10–15 ppm).

reagent could cause large signals vary from 3 to 4 ppm. This study has highlighted the importance of visual inspection of the  $^1\text{H}$  NMR data after PCA, as not all outliers identified by PCA could reliably be correlated to unique and interesting signals.

A limiting factor to using PCA was that supporting approaches were necessary to increase the scale. For example, PCA would not be suitable for analyzing more than 1000 NMR spectra at a time as larger data size would only result in identifying outliers with more extreme signals. Based on pre-analysis steps, we have found that analysis of between 200 and 500 samples was practical. For strain selection from a cultivated collection, we used a combination of

gross morphology and source organism to classify groups to be analyzed by PCA. In order to target as many outliers with unique NMR signals, the data sets generated from each strain were subjected separately to PCA analysis.

As with the previous evaluation, 13 data sets comprised of 320 fractions from each strain were reanalyzed separately using the same methodology. In total, 156 outliers were identified from 52 scores plots (13 actinomycetes  $\times$  4 regions) and 42 of them appeared more than once in different NMR spectral parts. The outliers were divided into 3 types according the visual inspection of their raw spectra. Type I outliers were highlighted mainly because



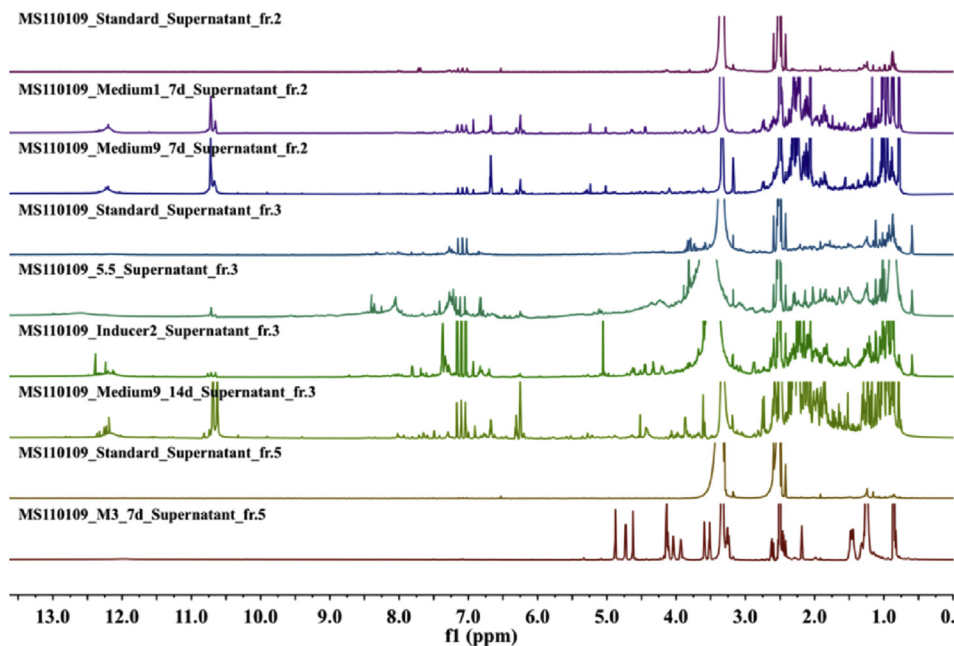


Fig. 13.  $^1\text{H}$  NMR spectra of the outliers of MS110109 and the corresponding standard fraction.

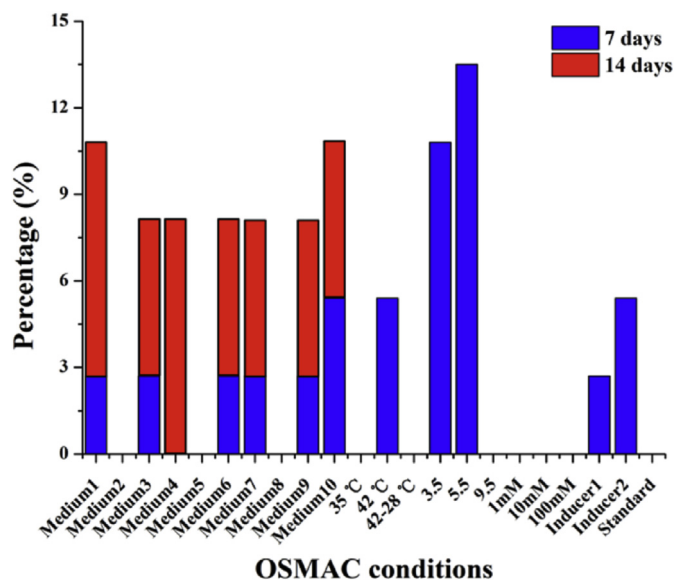


Fig. 14. The distribution of 37 prioritized outliers.

of some large signals generated in their spectra during the sample preparation or evaluation processes. As shown in Fig. 6, the fraction 5 of MS110104\_Medium6\_7d\_Biomass is well separated from the other fractions by component 2 in the scores plot. By examining the loadings plot and the respective NMR spectra (Fig. 7), it is a strong signal at 1.22 ppm was identified, with no additional peaks comparing to the standard fraction, which means type I outliers are not the target outliers in this project and therefore were excluded for further investigation.

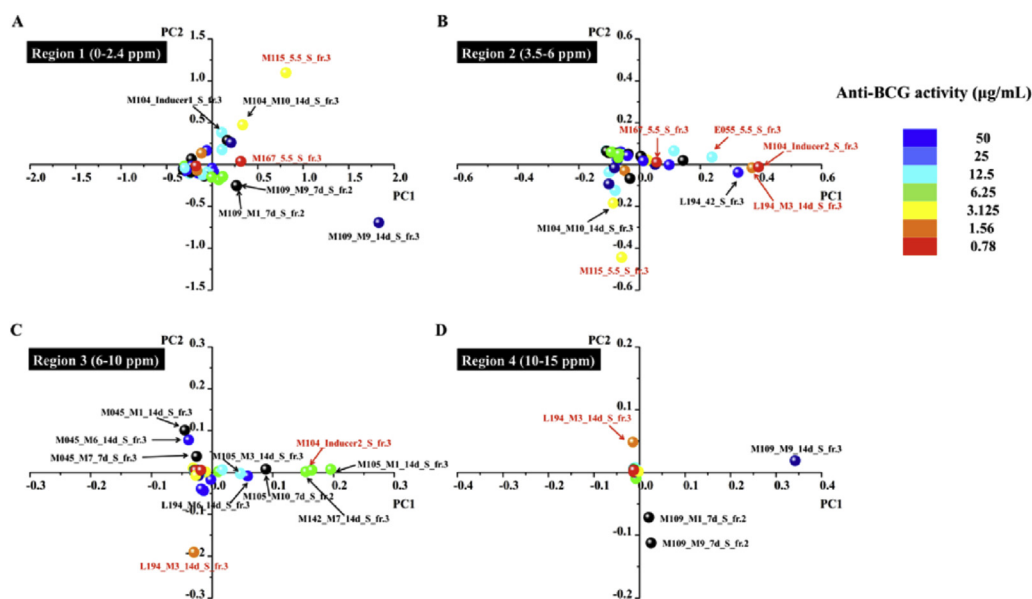
An example of outlier that displayed the effect of increasing the metabolites production was shown in Fig. 8. PCA were conducted based on bucket tables, which were intensities calculated from the selected NMR signals. Thus, PCA detects outliers according to the differences on intensities, including samples containing the same

metabolites with the others but in a higher quantity (type 2 outliers), or samples containing new generated metabolites from the others (type 3 outliers). The secondary metabolites production in endophyte ES120127 was enhanced by culturing in medium 9 for 14 days compared to the standard conditions in the same medium for 7 days (Fig. 9).

Apart from culture media variation, the addition of ethanol in standard medium was found to be beneficial to accumulation of metabolites. The third fractions of supernatant samples with the addition of different concentrations were all identified as outliers in scores plot of MS110149 dataset (Fig. 10). The NMR spectra in Fig. 11 showed there was no significant difference in signal intensities with 1 mM and 10 mM ethanol added. When ethanol level was reached to 100 mM, metabolites production was significantly enhanced. The effect of ethanol on metabolites biosynthesis has also confirmed by a series of studies and the mechanism was reported as multiple responses of intracellular signals, gene transcription and enzyme activity levels [33–35].

The third type outliers were identified because of the different metabolites produced under specific OSMAC conditions and are thus the prioritized fractions in this project, especially those outliers identified from more than one NMR region (Fig. 12). 11 outliers from marine strain MS110109 with their sample codes were shown in Fig. 12 and 3 of them appeared in more than one scores plot. After visual comparison with standard fractions, 6 fractions were determined to be the “prioritized” outliers with different NMR signals (Fig. 13), including the three repeated fractions.

Visual inspection and comparisons of 156 outliers with their standard fraction confirmed the differences indicated by the corresponding loadings plot. However, as mentioned before, some of the outliers containing strong signals from sample preparation process or the same signals with that in standard fractions were not selected as prioritized outliers in this project. As a result, 37 fractions with newly generated NMR signals under specific OSMAC conditions were identified. Fractions derived from media variations (62.16%) gave more than that from stress conditions (37.84%) (Fig. 14). Overall, cultivation for a longer time (14 days) provided more outliers than the standard culture period (7 days), containing



**Fig. 15.** Scores plots of 37 outliers fractions. Anti-BCG activity was shown in different colors coded according to the legend (blue lowest and red highest activity). A. Scores plot of region 1 (0–2.4 ppm). B. Scores plot of region 2 (3.5–6 ppm). C. Scores plot of region 3 (6–10 ppm). D. Scores plot of region 4 (10–15 ppm).

43.2% and 18.9% of outliers, respectively. Within the 10 selected media, medium 1 (10.8%) and 10 (10.8%) were the most effective media in inducing unique metabolite production, whereas, none were generated from medium 2, 5 or 8. Apart from media variation, stress shocks, including heat shock, pH shock and ethanol shock, as well as the addition of small inducer molecules also led to the production of unique metabolites. pH shock, especially acidic conditions (pH 3.5 and 5.5) provided 9 prioritized outliers, which are 64.3% of the outliers induced by all stress conditions. The second effective strategy among stress conditions was found to be the use of inducers, with 1 outlier induced by inducer 1 and 2 outliers induced by inducer 2, followed by the high temperature shock condition (42 °C), with 2 outliers.

The comparison of 37 outliers was explored by re-subjecting their NMR fingerprint data to PCA analysis. The resulting scores plots of PC1 versus PC2 flagged up a few distinct outliers in different NMR chemical shift regions (Fig. 15). Under comprehensive consideration of PCA results and their anti-BCG activity, especially those active fractions with improved activity comparing to the corresponding standard fractions, 6 fractions derived from 5 strains were considered as the most potent fractions for further chemical investigation, including 3 fractions from pH 5.5, 1 fraction from medium 3, 1 fraction from the addition of inducer 2 and 1 fraction from the high culture temperature condition. Further work will report the isolation of the identified compounds from those prioritized fractions.

#### 4. Conclusion

As a proof of concept, we have evaluated the application of OSMAC-NMR-PCA to reveal distinct candidate fractions with induced metabolites distribution in the sample set for discovery of new compounds. We have established an efficient and rapid method to prioritize 37 fractions for targeted follow-on study from 4160 microbial fractions.

In conclusion, OSMAC-NMR-PCA was effective for selecting microbial fractions with the potential to yield the most chemically diverse and novel natural products for drug discovery. Importantly,

we showed that microbes did display different secondary metabolite profiles under different culture conditions, and those differences could easily be observed using PCA. In the end, we have validated OSMAC-NMR-PCA as a rapid method to select fractions on the basis of production of secondary metabolites. In addition to fraction selection, this combined method greatly assisted with discovery of novel natural products and will be a useful tool to study regulation of natural product biosynthesis. Furthermore, even though this study had focused on the prioritization of fractions to identify new anti-TB compounds, this is a useful tool that may be applied generally to any drug discovery program to speed up identification of novel structures.

#### Acknowledgements

This research was supported by an Australian Research Council (ARC) Linkage Project DP160101429, ARC project LP120100485 and the Bill & Melinda Gates Foundation Grand Challenges Explorations Grant Phase II OPP1035218 GCE. M. Liu thanks Griffith University for the Publication Assistance Scholarship.

#### References

- [1] Cragg GM, Newman DJ. Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta* 2013;1830:3670–95.
- [2] Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 2012;75:311–35.
- [3] Leeds JA, Schmitt EK, Krastel P. Recent developments in antibacterial drug discovery: microbe-derived natural products—from collection to the clinic. *Expert Opin Investig Drugs* 2006;15:211–26.
- [4] Lew W, Pai M, Oxlade O, Martin D, Menzies D. Initial drug resistance and tuberculosis treatment outcomes: systematic review and meta-analysis. *Ann Intern Med* 2008;149:123–34.
- [5] Lam KS. Discovery of novel metabolites from marine actinomycetes. *Curr Opin Microbiol* 2006;9:245–51.
- [6] Gross H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* 2007;75:267–77.
- [7] Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, et al. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U. S. A* 2001;98:12215–20.
- [8] Wilkinson B, Micklefield J. Mining and engineering natural-product

- biosynthetic pathways. *Nat Chem Biol* 2007;3:379–86.
- [9] Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A-fumigatus* and *A-oryzae*. *Nature* 2005;438:1105–15.
- [10] Scherlach K, Hertweck C. Triggering cryptic natural product biosynthesis in microorganisms. *Org Biomol Chem* 2009;7:1753–1760.
- [11] Hofs R, Walker M, Zeeck A. Hexacyclenic acid, a polyketide from *Streptomyces* with a novel carbon skeleton. *Angew Chem Int Ed* 2000;39:3258–61.
- [12] Bode HB, Bethe B, Hofs R, Zeeck A. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBiochem* 2002;3:619–27.
- [13] Kell DB. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov Today* 2006;11:1085–92.
- [14] Yuliana ND, Jahangir M, Verpoorte R, Choi YH. Metabolomics for the rapid dereplication of bioactive compounds from natural sources. *Phytochem Rev* 2013;12:293–304.
- [15] Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, Walsh MC, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* 2001;19:45–50.
- [16] Krishnan P, Kruger NJ, Ratcliffe RG. Metabolite fingerprinting and profiling in plants using NMR. *J Exp Bot* 2005;56:255–65.
- [17] Kersten RD, Dorrestein PC. Secondary metabolomics: natural products mass spectrometry goes global. *ACS Chem Biol* 2009;4:599–601.
- [18] Nett M, Ikeda H, Moore BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Product Rep* 2009;26:1362–84.
- [19] Paranagama PA, Wijeratne EMK, Gunatilaka AAL. Uncovering biosynthetic potential of plant-associated fungi: effect of culture conditions on metabolite production by *Paraphaeosphaeria quadrisepitata* and *Chaetomium chiversii*. *J Nat Prod* 2007;70:1939–45.
- [20] Hosaka T, Ohnishi-Kameyama M, Muramatsu H, Murakami K, Tsurumi Y, Kodani S, et al. Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat Biotechnol* 2009;27:462–4.
- [21] Yoon V, Nodwell JR. Activating secondary metabolism with stress and chemicals. *J Ind Microbiol Biotechnol* 2014;41:415–24.
- [22] Puglia AM, Vohradsky J, Thompson CJ. Developmental control of the heat-shock stress regulon in *Streptomyces-coelicolor*. *Mol Microbiol* 1995;17:737–46.
- [23] Servant P, Mazodier P. Negative regulation of the heat shock response in *Streptomyces*. *Arch Microbiol* 2001;176:237–42.
- [24] Christian OE, Compton J, Christian KR, Mooberry SL, Valeriote FA, Crews P. Using jasplakinolide to turn on pathways that enable the isolation of new chaetoglobosins from *Phomopsis asparagi*. *J Nat Prod* 2005;68:1592–7.
- [25] Recio E, Colinas A, Rumbero A, Aparicio JF, Martin JF. PI factor, a novel type quorum-sensing inducer elicits pimaricin production in *Streptomyces natalensis*. *J Biol Chem* 2004;279:41586–93.
- [26] Wang FZ, Wei HJ, Zhu TJ, Li DH, Lin ZJ, Gu QQ. Three new cytochalasins from the marine-derived fungus *spicaria elegans* KLA03 by supplementing the cultures with L- and D-tryptophan. *Chem Biodivers* 2011;8:887–94.
- [27] Camp D, Davis RA, Campitelli M, Ebdon J, Quinn RJ. Drug-like properties: guiding principles for the design of natural product libraries. *J Nat Prod* 2012;75:72–81.
- [28] Boulahya KA, Guedon E, Delaunay S, Schultz C, Boudrant J, Bott M, et al. OdhI dephosphorylation kinetics during different glutamate production processes involving *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* 2010;87:1867–74.
- [29] Zhang CH, Fevereiro PS. The effect of heat shock on paclitaxel production in *Taxus yunnanensis* cell suspension cultures: role of abscisic acid pretreatment. *Biotechnol Bioeng* 2007;96:506–14.
- [30] Rao YM, Sureshkumar GK. Improvement in bioreactor productivities using free radicals: HOCl-induced overproduction of xanthan gum from *Xanthomonas campestris* and its mechanism. *Biotechnol Bioeng* 2001;72:62–8.
- [31] Dash S, Mohanty N. Response of seedlings to heat-stress in cultivars of wheat: growth temperature-dependent differential modulation of photosystem 1 and 2 activity, and foliar antioxidant defense capacity. *J Plant Physiol* 2002;159:49–59.
- [32] Chatterjee ISG, Heilmann C, Sahl HG, Maurer HH, Herrmann M. Very low ethanol concentrations affect the viability and growth recovery in post-stationary-phase *Staphylococcus aureus* populations. *Appl Environ Microbiol* 2006;72:2627–36.
- [33] Jakeman DL, Graham CL, Young W, Vining LC. Culture conditions improving the production of jadomycin B. *J Ind Microbiol Biotechnol* 2006;33:767–72.
- [34] Doull JL, Ayer SW, Singh AK, Thibault P. Production of a novel polyketide antibiotic, jadomycin B, by *Streptomyces venezuelae* following heat shock. *J Antibiot (Tokyo)* 1993;46:869–71.
- [35] Zhou WW, Ma B, Tang YJ, Zhong JJ, Zheng XD. Enhancement of validamycin A production by addition of ethanol in fermentation of *Streptomyces hygroscopicus* 5008. *Bioresour Technol* 2012;114:616–21.