

## Genome Mining of Endophytic *Streptomyces wadayamensis* Reveals High Antibiotic Production Capability

Célio F. F. Angolini,<sup>\*,a</sup> Ana B. Gonçalves,<sup>a</sup> Renata Sigrist,<sup>a</sup> Bruno S. Paulo,<sup>a</sup> Markiyam Samborsky,<sup>b</sup> Pedro L. R. Cruz,<sup>a</sup> Adriana F. Vivian,<sup>c</sup> Eduardo M. Schmidt,<sup>a</sup> Marcos N. Eberlin,<sup>a</sup> Welington L. Araújo<sup>d</sup> and Luciana G. de Oliveira<sup>\*,a</sup>

<sup>a</sup>Departamento de Química Orgânica, Instituto de Química, Universidade Estadual de Campinas, 13083-970 Campinas-SP, Brazil

<sup>b</sup>Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, Cambridgeshire, CB2 1GA, England

<sup>c</sup>Faculdade de Ciências Farmacêuticas, Universidade Estadual de Campinas, 13083-970 Campinas-SP, Brazil

<sup>d</sup>Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, 05508-900 São Paulo-SP, Brazil

The actinobacteria *Streptomyces wadayamensis* A23, an endophytic strain, was recently sequenced and previous work showed qualitatively that the strain inhibits the growth of some pathogens. Herein we report the genome analysis of *S. wadayamensis* which reveals several antibiotic biosynthetic pathways. Using mass spectrometry, we were able to identify desferoxamines, several antimycins and candicidin, as predicted. Additionally, it was possible to confirm that the biosynthetic machinery of the strain when compared to identified known metabolites is far underestimated. As suggested by biochemical qualitative tests, genome encoded information reveals that the strain A23 has high capability to produce antibiotics.

**Keywords:** genome sequencing, genome mining, molecular networking, *Streptomyces*, antibiotic activity

### Introduction

*Streptomyces* genera carry an extremely versatile group of biosynthetic machineries capable to produce complex molecules of medical, agricultural and economical significance. They are recognized as the most resourceful producers of molecules that present antibiotic activities.<sup>1</sup> The mutualism existing among several actinobacteria and plants, ants and other organisms also seemed to be a key element to the capability to produce such small molecules.<sup>2,3</sup>

Since the pioneers work in partial genome sequencing to a more recent and wide investment on whole genome sequencing, it became clear that the ability of *Streptomyces* and other actinobacteria to biosynthesize promising metabolites is underestimated.<sup>1,4</sup> Recent developments on next generation sequencing revealed a diverse plethora of

gene clusters encoding for cryptic targets. It became clear that by accessing sequences genome encoded it is possible to predict enzymatic functions and the hole of an entire gene cluster involved in secondary metabolites biosynthesis.

Genome-mining tools have therefore opened a new and innovative avenue for natural products research.<sup>5</sup> When combined with bioinformatic tools, it contributes to envisage a global map of biosynthetic gene clusters (BGC). The possibility to connect genes to natural products enables to direct the discovery of novel molecules, for instance, metabolites as coelichelin,<sup>6</sup> orfamide A<sup>7</sup> and salinilactam<sup>8</sup> were found by searching programs guided by the prediction of physical-chemical properties, anticipation of precursors incorporation, manipulation of higher levels and pathway specific regulators, inactivation of biosynthetic genes and analytical evaluation, and by *in vitro* reconstitution.<sup>9-18</sup>

Driven by the current status of the genome to natural products programs we invested in the Whole Genome

\*e-mail: luciana@iqm.unicamp.br, celio@iqm.unicamp.br

Shotgun project of *Streptomyces wadayamensis* A23, an endophytic strain isolated from *Citrus reticulata* (tangerine).<sup>19</sup> Biochemical qualitative tests showed that strain A23 inhibits the growth of *Candida albicans* pathogen, *Bacillus megaterium*, *Neisseria meningitidis* strains, and multiresistant *Staphylococcus aureus*.<sup>20</sup> Herein we report our preliminary results generated after genome annotation and mining of the WGS project. Annotation was critical to predict that *S. wadayamensis* has high capability to biosynthesize antibiotics. Experimental evaluation combined to powerful analytical tools revealed the production of several bioactive metabolites as anticipated by mining.

## Experimental

### Chemical and samples

Water was purified using a Milli-Q water purification system (Millipore). Beef extract, yeast extract, peptone and agar were purchased from Oxoid (Hampshire, UK). HPLC grade methanol was purchased from Tedia Brazil (Rio de Janeiro, Brazil).  $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ , HCl, chromeazurol S, hexadecyltrimethylammonium bromide (HTMS), PIPES,  $\text{KH}_2\text{PO}_4$ , NaCl,  $\text{NH}_4\text{Cl}$ , KOH, glucose, D-mannitol, casein hydrolysate,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ ,  $\text{CaCl}_2$ ,  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ ,  $\text{H}_3\text{BO}_3$ ,  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ ,  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  and  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  were purchased from Sigma-Aldrich. The actinobacterium strain A23, identified as *Streptomyces wadayamensis*, was isolated from a population existing in the plant tissue of *Citrus reticulata* (tangerine).

### Culture media

The media used in this work were vegetative N (peptone 5 g L<sup>-1</sup>, soluble starch 20 g L<sup>-1</sup>, meat extract 2 g L<sup>-1</sup>, yeast extract 3 g L<sup>-1</sup>, soy bean meal 2 g L<sup>-1</sup>,  $\text{CaCO}_3$  1 g L<sup>-1</sup>, pH 7.0); A medium (peptone 4 g L<sup>-1</sup>, meat extract 4 g L<sup>-1</sup>, yeast extract 2 g L<sup>-1</sup>, soybean meal 2 g L<sup>-1</sup>, maltose 20 g L<sup>-1</sup>, dextrine 10 g L<sup>-1</sup>, pH 7.0); INA (soy bean meal 15 g L<sup>-1</sup>,  $\text{CaCO}_3$  5 g L<sup>-1</sup>, glycerol 30 mL L<sup>-1</sup>, NaCl 2 g L<sup>-1</sup>, pH 7.0); RA3 (peptone 2 g L<sup>-1</sup>, yeast extract 4 g L<sup>-1</sup>, malt extract 10 g L<sup>-1</sup>, glucose 10 g L<sup>-1</sup>, glycerol 5 mL L<sup>-1</sup>,  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  2 g L<sup>-1</sup>, pH 7); GPMY (potato starch 20 g L<sup>-1</sup>, yeast extract 5 g L<sup>-1</sup>, malt extract 5 g L<sup>-1</sup>, glycerol 20 mL L<sup>-1</sup>, pH 7.0) AFMS (dextrose 20 g L<sup>-1</sup>, yeast extract 2 g L<sup>-1</sup>, soybean meal 8 g L<sup>-1</sup>,  $\text{CaCO}_3$  4 g L<sup>-1</sup>, NaCl 1 g L<sup>-1</sup>, pH 7.3); GYM (yeast extract 4 g L<sup>-1</sup>, malt extract 10 g L<sup>-1</sup>, glucose 4 g L<sup>-1</sup>, pH 7.0) and M8 (soluble starch 20 g L<sup>-1</sup>, meat extract 2 g L<sup>-1</sup>, yeast extract 2 g L<sup>-1</sup>, glucose 10 g L<sup>-1</sup>,  $\text{CaCO}_3$  3 g L<sup>-1</sup>, pH 7.0).

### Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) sample preparation

A single colony of *Streptomyces wadayamensis* was selected and added to 1 mL of vegetative N medium (seed). The cell was grown for 24 h at 30 °C and 250 rpm. After this time, the seed medium was added in 100 mL of the culture medium described above. *Streptomyces wadayamensis* was cultivated for 7 days in 500 mL Erlenmeyer flasks on reciprocal shaker at the same conditions. The medium was centrifuged and the mycelium was separated. The broth was extracted three times with ethyl acetate. The organic layer was evaporated. 1 mg of the culture extract was dissolved in 1 mL of methanol and centrifuged for 10 min at 13,000 rpm to remove any precipitate. This solution was used as stock solution from which aliquots of 100 µL were removed and diluted in 1 mL of methanol.

### FT-ICR-MS analysis

The equipment was an LTQ FT Ultra 7T (ThermoScientific) equipped with a nano-electrospray ionization (ESI) source (TriVersa NanoMate 100 system). Ionization was carried out both on the negative and positive ion mode on an *m/z* between 100 and 2000. The spectra generated were analyzed with the Xcalibur software.

### Desorption electrospray ionization imaging mass spectrometry (DESI-IMS) sample preparation

For microbial IMS sample preparation, a 0.25 µL sample from a *Streptomyces* spore solution was inoculated on thin-layered agar. The antibiotic assay medium 2 (1.5 g L<sup>-1</sup> beef extract, 3.0 g L<sup>-1</sup> yeast extract, 6 g L<sup>-1</sup> peptone and 1.5% agar) was prepared by first placing a sterile microscope slides in a Petri dish, followed by the pouring of agar medium. The inoculated medium was incubated at 30 °C, and the Petri dish was sealed with parafilm to minimized premature dehydration. After the incubation period (48 h), the microscope slides were removed from the Petri dish, photographed and put in a vacuum desiccator for complete agar dehydration at room temperature.

### DESI-IMS analysis

IMS was performed using a Prosolia DESI source (Model OS-3201) connected to a Q Exactive Hybrid Quadrupole-Orbitrap. The DESI configuration was set with an emitter height of 2.5 mm, mass spectrometer inlet height of 0.1 mm, inlet to emitter distance of 3.8 mm, 58° sprayed angle, 5.0 kV spray voltage, inlet capillary

temperature of 320 °C, 100 V S-lens, 160 psi ultra-pure nitrogen nebulizing gas pressure and a sprayed solvent of methanol in a 3.0  $\mu\text{L min}^{-1}$  flow rate. Images were collected scanning from  $m/z$  133-1500 with a step size of 200 mm, a scan rate of 741  $\mu\text{m sec}^{-1}$  and a pixel size of 200  $\times$  200  $\mu\text{m}$ .

#### MS and tandem mass spectrometry (MS/MS) for network analyses

For analysis of *Streptomyces wadayamensis* A23 metabolite profiles, cells cultures and medium were removed from microscope slides and extract with 0.5 mL methanol after DESI-IMS analysis. The extract were centrifuged for solid parts removal, and the remaining solution was analyzed by direct infusion ESI in a Thermo Scientific Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer. The MS/MS experiments were all conducted with 3.5 kV spray voltage, inlet capillary temperature of 320 °C, 100 V S-lens in a 3.0  $\mu\text{L min}^{-1}$  syringe flow rate.

#### Data analysis

The DESI-IMS data was converted into imaged files using the Firefly data conversion software (version 2.1.05)<sup>21</sup> and viewed using the BioMAP software (version 3.8.04).<sup>22</sup> For molecular network creation, the tandem mass data (.raw) were converted in mzXML using msconvert from ProteoWizard and the GNPS automated platform<sup>23</sup> in

default settings.<sup>24</sup> In addition, Cytoscape 2.8<sup>25</sup> was used for molecular network visualization.

## Results and Discussion

As previously described, the genome draft of *S. wadayamensis* A23 has near to 7 million base pairs, encoding 6006 protein-coding sequences with 73.5% GC content.<sup>19</sup> The Illumina shotgun library produced 7.8 million reads in a total of 2.4 GB data assembled to generate 180 contigs. Genome annotation was carried out using a customized pipeline and revealed several putative secondary metabolites encoded by *S. wadayamensis*.

The input of the genome data into the antibiotics & Secondary Metabolite Analysis SHell (antiSMASH) server (version 3.0.5)<sup>26</sup> resulted in an output predicting the presence of 32 gene clusters codifying for secondary metabolites biosynthesis. Biosynthetic systems as terpene, nrps, bacteriocin-terpene, t1-pks, t3-pks, bacteriocin, nrps-t1pks, tiopeptide-lantipeptide, lantipeptide-nrps-t1-pks, siderophore, ectoine, lassopeptide, lantipeptide and other clusters are apparently encoded in *S. wadayamensis* genome (Table S1). Crossing the output results from antiSMASH<sup>27</sup> within the MIBiG (Minimum Information about a Biosynthetic Gene cluster) pipeline,<sup>28</sup> it was possible to predict more accurately certain metabolites and/or the most likely class that the gene cluster fits as presented in Table 1.

**Table 1.** Most significant gene clusters encoded in *Streptomyces wadayamensis* genome (antiSMASH-MIBiG output)<sup>a</sup>

Cluster overview	Coordinates (nt location in the contig)	Gene cluster type	Predicted metabolite
Cluster 1	174778 - 225031	nrps-t1pks	SGR PTMs BGC (100% of genes show similarity) frontalamides BGC (100% of genes show similarity)
Cluster 2	258744 - 285332	terpene	hopene BGC (76% of genes show similarity)
Cluster 4 and 16	1 - 80422 1 - 69344	t1pks	FR-008 BGC (50% of genes show similarity) FR-008 BGC (95% of genes show similarity)
Cluster 5	85376 - 126473	t3pks	herboxidiene BGC (12% of genes show similarity)
Cluster 6	208192 - 240719	terpene-bacteriocin	carotenoid BGC (54% of genes show similarity) isorenieratene BGC (54% of genes show similarity)
Cluster 7	255844 - 300115	other	indigoidine BGC (80% of genes show similarity) auricin deoxysugar moieties BGC (80% of genes show similarity)
Cluster 8	174427 - 195716	terpene	albaflavenone BGC (100% of genes show similarity)
Cluster 15	75862 - 86260	ectoine	ectoine BGC
Cluster 17	65097 - 139158	nrps-t1pks	antimycin BGC (93% of genes show similarity)
Cluster 21	35751 - 96211	nrps	mannopeptimycin BGC (48% of genes show similarity)
Cluster 22	18428 - 30248	siderophore	desferoxamine B BGC (100% of genes show similarity)
Cluster 25	6543 - 28843	terpene	germacradienol / geosmine synthase
Cluster 26	14854 - 29885	siderophore	aerobactin IucA / IucC family protein

<sup>a</sup>Not all annotated gene clusters are represented in this chart.

However, it is important to emphasize that, although the antiSMASH platform predicts 32 possible BGC, this actual number could be underestimated as the genome is organized in 180 contigs and some clusters are truncated (for example, both clusters 4 and 16 are complementary and correspond to the BGC of FR-008).

After *in silico* analysis, the metabolic profile of the ethyl acetate extracts from cultivation in several mediums was evaluated by FT-ICR-MS and Orbitrap in both the positive and negative ion modes using a  $m/z$  range between 100 and 2000. To identify the most common ions present in the extracts, we performed a manual search in the dictionary of natural products (DNP),<sup>29</sup> with near to 226000 structures from wide-ranging sources.

Although there is a difference in intensity for the signals in the chromatographic profile (Figure S1), almost the same set of ions were found in the conditions tested for fermentation (data not shown).

A careful evaluation of A medium pointed to an ion of  $m/z$  601.3552 with attributed formula  $C_{27}H_{48}N_6O_9$ . Among the structures generated in the DNP database, we identified nocardamine, also known as desferoxamine E (Figure 1), a siderophore with high affinity for iron, which forms a complex that is reabsorbed by the cell via the ATP-dependent transport system.<sup>30</sup> We had already observed the production of siderophores by qualitative tests on agar plates containing chrome azurol sulfonate.<sup>20</sup> Additionally, we performed direct DESI-IMS on an agar culture and observed another siderophore of same family, that is, the desferoxamine B of  $m/z$  561.3640 (Figure 1).<sup>31</sup>

Both siderophores predicted by the gene cluster 22 were therefore characterized by chemical/biological tests (chrome azurol) and MS/MS fragmentation patterns. The fragmentation of protonated desferoxamine E was mainly characterized by the neutral loss of a molecule of succinylcadaverine (200 Da, Figure S2).<sup>32</sup>

Desferoxamine (MIBiG BGC 0000941) is a highly conserved gene cluster in *Streptomyces* genera (Figure 2). The antiSMASH output for *Streptomyces wadayamensis* A23 resulted in 10 different sequences deposited in databases that possesses high similarity within our query sequence (*Streptomyces albus* J107, *Streptomyces* sp. SM8, *Streptomyces* sp. PVA 94-07, *Streptomyces* sp. GBA 94-10, *Streptomyces mutabilis* TRM4554, *Streptomyces wadayamensis* A23 LGO A23 AS7, *Streptomyces* sp. JS01, *Streptomyces griseus* subsp. *griseus* NBRC 1335, *Streptomyces griseus* XylebKG-1). The gene cluster comprises a set of 5 main enzymes (5 to 9) with essential functions for the siderophore biosynthesis: siderophore biosynthetic enzyme (5), acetyltransferase (6), monooxygenase (7), pyridoxal-dependent decarboxylase (8) and siderophore-interacting protein (9).

A total of 6 product ions related to antimycins were also identified, which are part of a family of polyketides-nonribosomal peptides already isolated from *Streptomyces* sp. exhibiting confirmed antifungal activity and predicted on gene cluster 17.<sup>33,34</sup> The antimycins are biosynthesized by a mixed cluster nrps-t1pks and the main differences among them are located in the chains R<sup>1</sup> and R<sup>2</sup> positions, which corresponds to an aliphatic and fat acid chain, respectively,

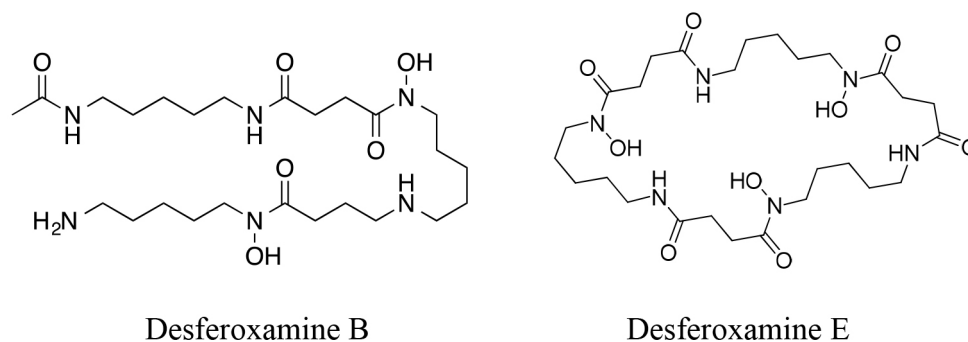


Figure 1. Siderophore desferoxamines B and E chemical structures.

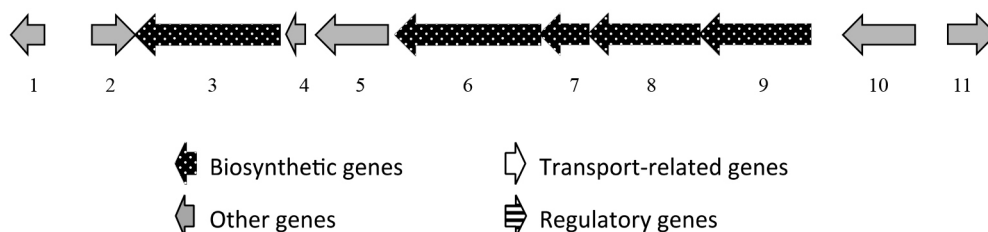


Figure 2. Desferoxamine biosynthetic gene cluster from *Streptomyces wadayamensis* A23.

and characterize the main losses observed by MS/MS (Table 2).<sup>33,34</sup> Most of the product ions from antimycins fragmentation are common among all isoforms, e.g.,  $m/z$  263, 245, 219, 191 and 161, however and fortunately, there is often one fragment ( $R^1$ , Figure 3A) which is characteristic for each antimycins variant. Therefore, based on these product ions, the production of antimycin variants such as A11, A13, A14, A16 and A20 could be excluded from our samples. As some variants have the same precursor ions and the same characteristic product ions, the fragmentation pattern may correspond, for instance, to A1a or A1b, A12 or A19, A8 or A17 (where  $R^1$  corresponds to distinct acyl chains in the structural isomers), therefore it is not possible to distinguish between them by MS spectrometry and additional

analysis, such as NMR, would be necessary to discriminate the structural isomers.

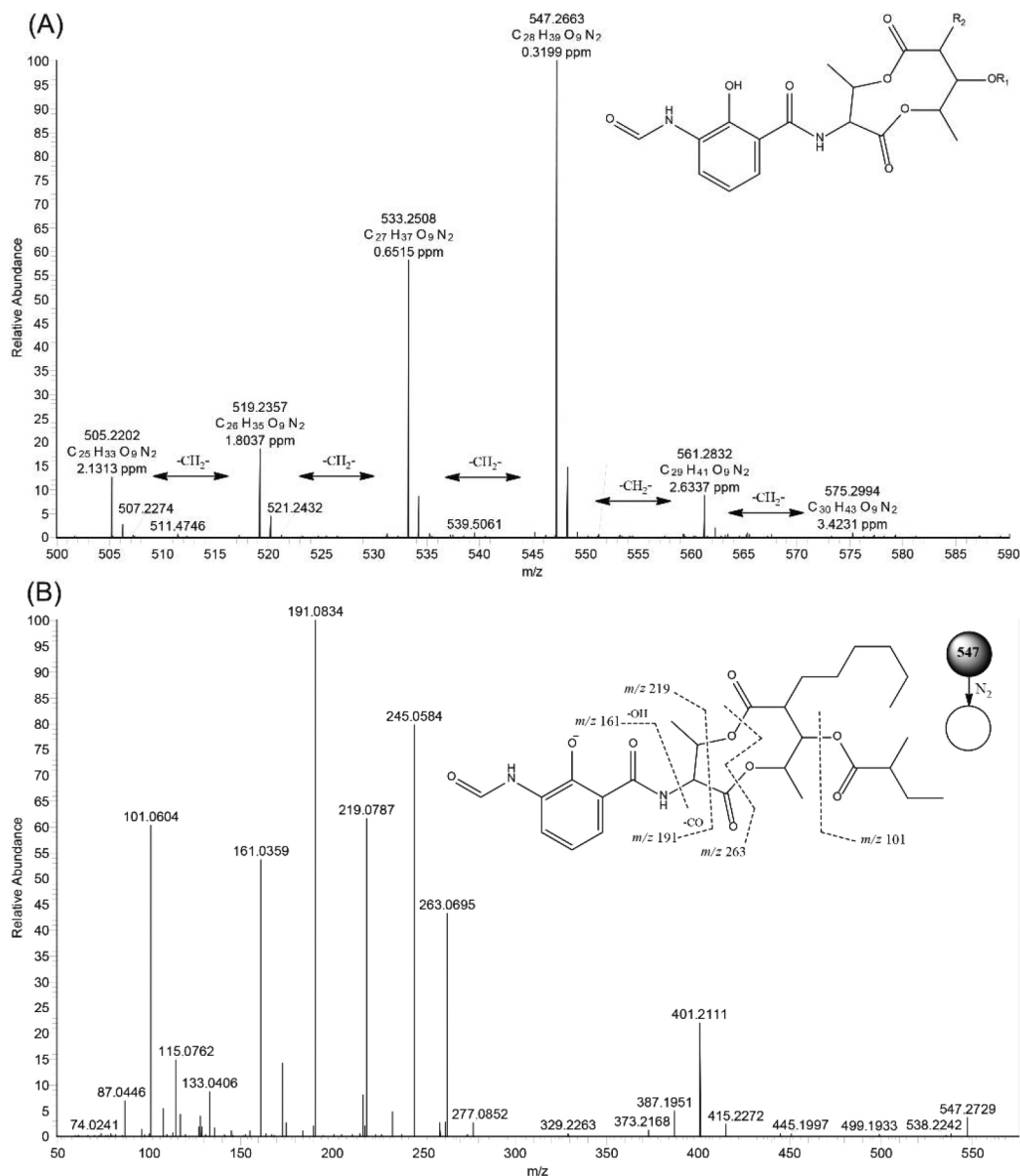
The production of antimycins A1, A12 and/or A19 ( $C_{28}H_{40}N_2O_9$ ) identified by the  $[M - H]^-$  anion of  $m/z$  547.2663; antimycin A4 ( $C_{25}H_{34}N_2O_9$ ) of  $m/z$  505.2202; antimycin A3 and/or A7 ( $C_{26}H_{36}N_2O_9$ ) of  $m/z$  519.2357; antimycin A2, A8 and/or A17 ( $C_{27}H_{38}N_2O_9$ ) of  $m/z$  533.2508; antimycin A10 and/or A15 ( $C_{29}H_{42}N_2O_9$ ) of  $m/z$  561.2832 and antimycin A16 ( $C_{30}H_{44}N_2O_9$ ) of  $m/z$  575.2994 were accordingly confirmed by A23 strain in the cultivation conditions (Figure 3 and Figure S3).

The gene cluster codifying for antimycins in *S. wadayamensis* is 100% identical to the one described by Seipke *et al.*<sup>35</sup> and fed in MIBiG platform (BGC0000958),

**Table 2.** The antimycins PKS-NRPS hybrids variants and their ESI(-)-MS/MS characteristics

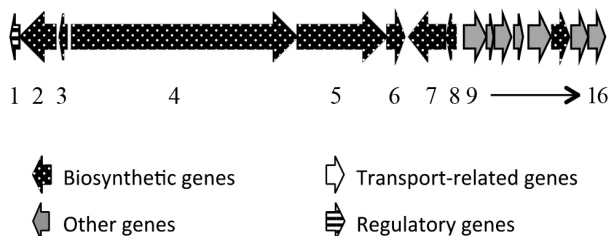
$m/z$ (error)	Antimycin		$R^1$	Characteristic product ion	$R^2$
505.2202 (1.07 ppm)	A4	a	-COCH(CH <sub>3</sub> ) <sub>2</sub>	87.0446	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
		b	-COCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	87.0446	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
	A20		-COCH <sub>3</sub>	43.0183	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
519.2357 (0.94 ppm)	A3	a	-COCH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	101.0602	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
		b	-COCH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	101.0602	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
	A7	a	-COCH(CH <sub>3</sub> ) <sub>2</sub>	87.0446	-(CH <sub>2</sub> ) <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>
		b	-COCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	87.0446	-(CH <sub>2</sub> ) <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>
533.2508 (0.34 ppm)	A2	a	-COCH(CH <sub>3</sub> ) <sub>2</sub>	87.0446	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
		b	-COCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	87.0446	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
	A8	a	-COCH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	101.0602	-(CH <sub>2</sub> ) <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>
		b	-COCH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	101.0602	-(CH <sub>2</sub> ) <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>
	A11		-COCH <sub>2</sub> CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	115.0759	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
	A17		-COCH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	101.0602	-(CH <sub>2</sub> ) <sub>4</sub> CH <sub>3</sub>
547.2663 (0.18 ppm)	A1	a	-COCH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	101.0602	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
		b	-COCH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	101.0602	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
	A12		-COCH <sub>2</sub> CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	115.0759	-(CH <sub>2</sub> ) <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>
	A13		-COCH <sub>2</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	129.0915	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
	A19		-COCH(CH <sub>3</sub> )CH(CH <sub>3</sub> ) <sub>2</sub>	115.0759	-(CH <sub>2</sub> ) <sub>4</sub> CH <sub>3</sub>
561.2832 (1.47 ppm)	A10	a	-COCH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	101.0602	-(CH <sub>2</sub> ) <sub>3</sub> CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>
		b	-COCH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	101.0602	-(CH <sub>2</sub> ) <sub>3</sub> CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>
	A14		-COCH <sub>2</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	129.0915	-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>
	A15		-COCH <sub>2</sub> CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	115.0759	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
575.2994 (1.96 ppm)	A16		-COCH <sub>2</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	129.0915	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>





**Figure 3.** (A) ESI(-) FT-ICR-MS of *Streptomyces* A23 crude extract (AFMS medium) showing deprotonated antimycin analogues; (B) MS/MS fragmentation pattern of the deprotonated antimycin A1a (note that this fragmentation pattern could be expected for A1b, A12 and A19 as well).

comprising the genes 1 to 16 in Figure 4. The modular polyketide synthase gene *antD* (gene number 5, module 3, Figure 4) carries a KS, AT, ACP and TE domain. According

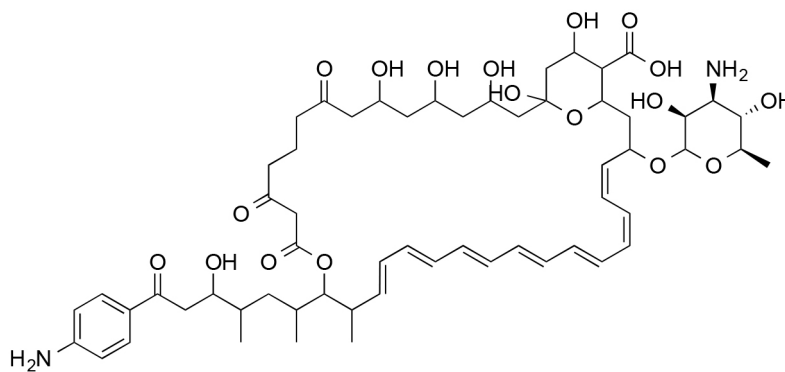


**Figure 4.** The organization of antimycins gene cluster in *Streptomyces wadayamensis* A23.

to the differences observed in R1 chains, it is possible to infer that the AT specificity is highly atypical with the incorporation of several acyl-CoAs subunits.

The mining also indicated a possible metabolite described by the clusters 4 and 16 belonging to the FR-008 BGC, a complex polyene with high similarity to the candicidin gene cluster found in *Streptomyces griseus* IMRU3570 (Figure 5).<sup>36</sup> Chen *et al.*<sup>36</sup> originally described this metabolite and its gene cluster in 2003 and throughout this study it was possible to conclude that candicidin and FR-008 are identical compounds.

A meticulous search led us to find candicidin ( $m/z$  1109.5784,  $C_{59}H_{84}N_2O_{18}$ , Figure 6) in the crude extract,



### Candicidin

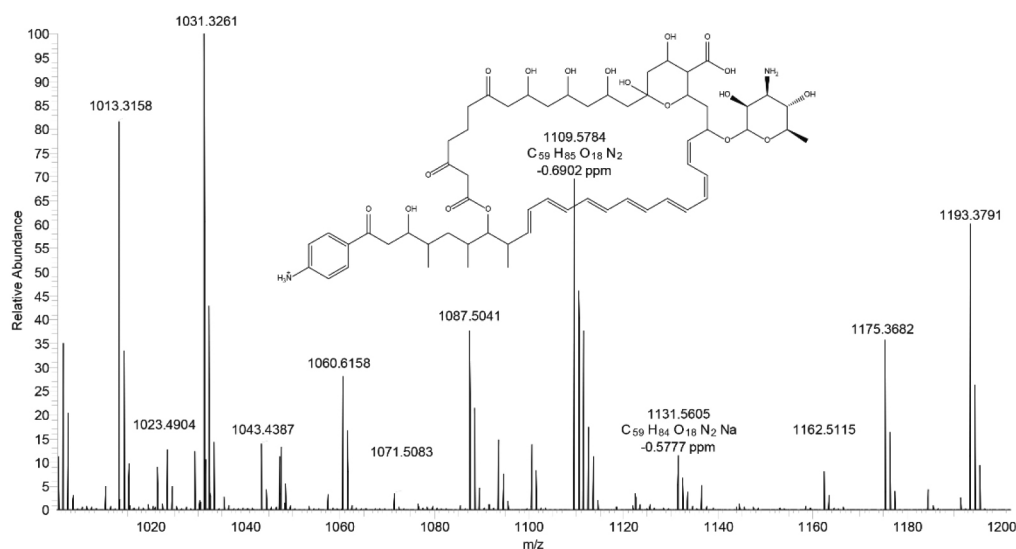
**Figure 5.** Chemical structure of the complex polyene candicidin.

which could be most likely, the compound responsible for *C. albicans* growth inhibition.<sup>36</sup>

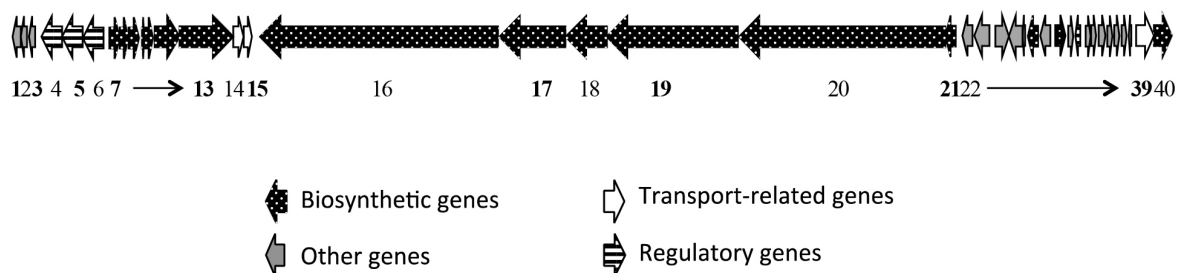
An indication for the production of candicidin in *S. wadayamensis* was already anticipated in previous studies<sup>20</sup> by the use of metabolic probes to promote dereplication. The KS-AT domains from FscB and FscD (gene numbers 17 and 20, respectively, Figure 7) were identified using this methodology. The candicidin gene

cluster organization in the endophytic strain is 100% identical to the gene cluster description for FR-008 (BGC0000061 in MIBiG repository).

To enhance the A23 dereplication, we performed the molecular networking analysis of A23 cultivation extracts, but no further known molecules were found. Interestingly, we were able to cross several information found by FT-ICR-MS and DESI-IMS. As molecular networking also



**Figure 6.** ESI(+)-Orbitrap MS of *Streptomyces* A23 extract (A medium) showing the observed ions in the range  $m/z$  1000-1200.



**Figure 7.** Candicidin gene cluster organization in the endophytic *Streptomyces wadayamensis*.

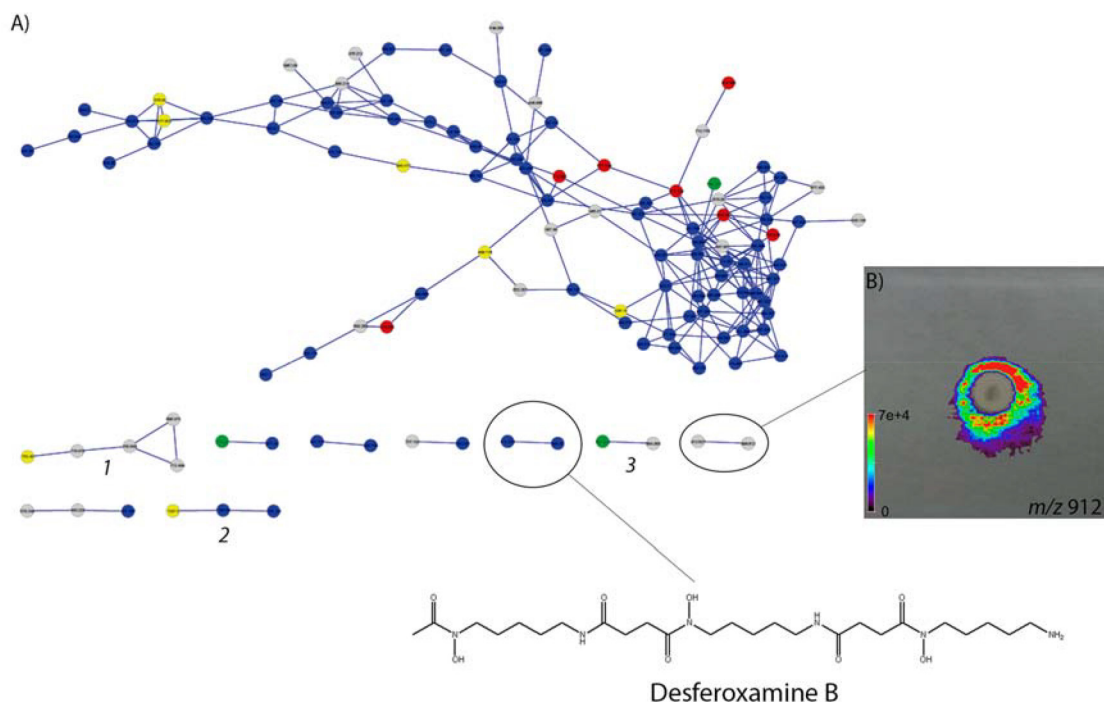
enables clustering families of structurally related natural products by MS/MS fragmentation pattern analysis, and DESI-IMS is able to correlate molecules of same spatial distribution during the microorganism growth, it is possible therefore to establish a correlation within a related set of compounds that could share chemical similarities.<sup>24,37</sup> Using molecular networking dereplication, we were able to find desferoxamine B (described in cluster 22) and their iron complex form (Figure 8),<sup>31</sup> and we also tagged three other clusters of potentially unknown natural products (cluster 1:  $m/z$  700.645; cluster 2:  $m/z$  1193.38; cluster 3:  $m/z$  912.627).

IMS experiments revealed that metabolites possibly described by the gene clusters from Table 1 and Table S1 could be visualized over the agar media around the colony. A preliminary MS/MS analysis showed that the ion  $m/z$  912 seems to be a nonribosomal linear peptide (NRP) not yet described (Figure S4). We failed to find a similar pattern in networking or in dictionary of natural products. In total, we found three different compounds ( $m/z$  884, 898, and 912) differing in 14 u ( $\text{CH}_2$ ) between each other from which is possible to observe a similar arrangement distribution over the agar media in IMS data.

Finally, mining the genome revealed gene cluster 21, related to a NRPS responsible to mannopeptimycin biosynthesis (48% of the genes show similarity) produced by several strains of *Streptomyces hygroscopicus*. Mannopeptimycins are glycosylated cyclic hexapeptides

(glycopeptides) first described as a complex mixture of compounds named AC98 and showed effective activity against Gram-positive bacteria (Figure 9).<sup>38</sup> Its activity was also confirmed against methicillin resistant staphylococci and vancomycin resistance enterococci.<sup>39</sup>

The biosynthetic pathway for mannopeptimycins was described by Margavey *et al.*<sup>40</sup> in 2006. Looking closely and performing blast and protein function comparison, it is possible to verify that the main genes *mppA* and *mppB* responsible for the hexapeptide core are 100% identical in *S. wadayamensis* A23 (genes 23 and 24 in Figure 10A) and *S. hygroscopicus* NRRL 30439, revealing similar operon type organization (Figure 10B). The most part of the right hand outside genes (except for *mppM*, *mppN*, *mppX*, *mppY*, *mppZ*, and *mppZI*) are also identical in both strains. *mppM* and *mppN* correspond to isovaleryl transferases, responsible for the isovaleryl incorporation in the terminal *O*-linked mannose in mannopeptimycins  $\gamma$ ,  $\delta$ , and  $\epsilon$ . The absence of these genes in the BGC codified by *S. wadayamensis* suggests that this strain could produce  $\alpha$  or  $\beta$ -mannopeptimycins. The genes *mppX*, *mppY* and *mppZ* have unknown function and *mppZI* is a transcriptional regulator. The left hand side genes in *Streptomyces wadayamensis* codifies for a single major facilitator transporter (gene 22), while in the mannopeptimycin gene cluster used for comparison there are two genes (E and F) codifying for this function. In the MIBiG entry *mppI* and



**Figure 8.** (A) Annotated molecular networking of orbitrap MS/MS fragmentation data acquisition from crude extract of *Streptomyces* A23 in 3 different growth times: 40 h (red nodes), 60 h (green nodes) and 90 h (yellow nodes). Compounds present in all experimental conditions are represented with blue nodes and other combinations in gray nodes; (B) superposed image showing the spatial distribution for  $m/z$  912 over the medium around the microbial colony.



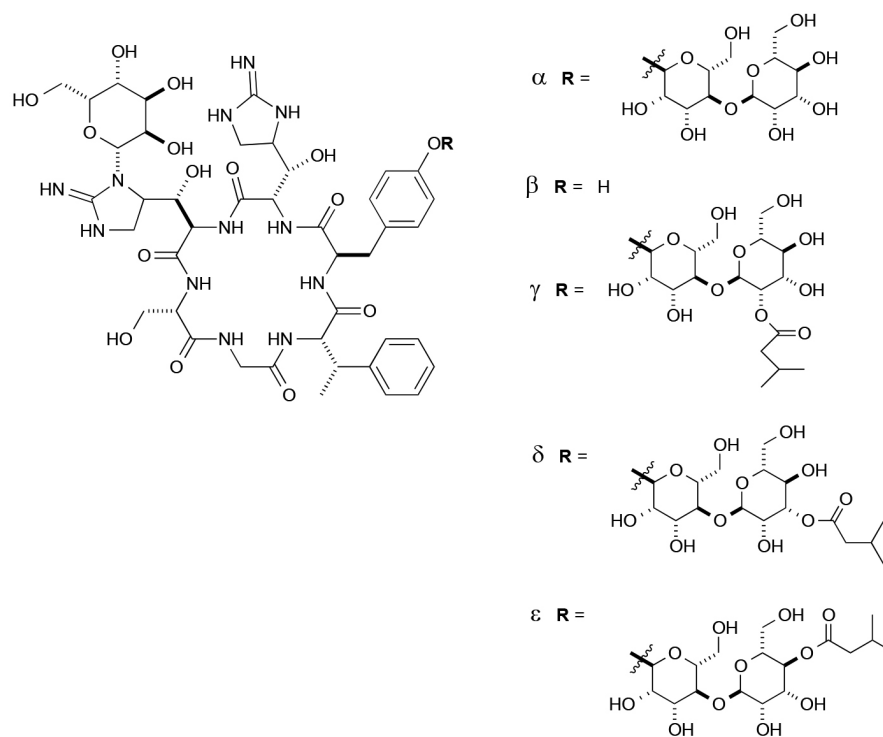


Figure 9. Mannopeptimycins chemical structures.

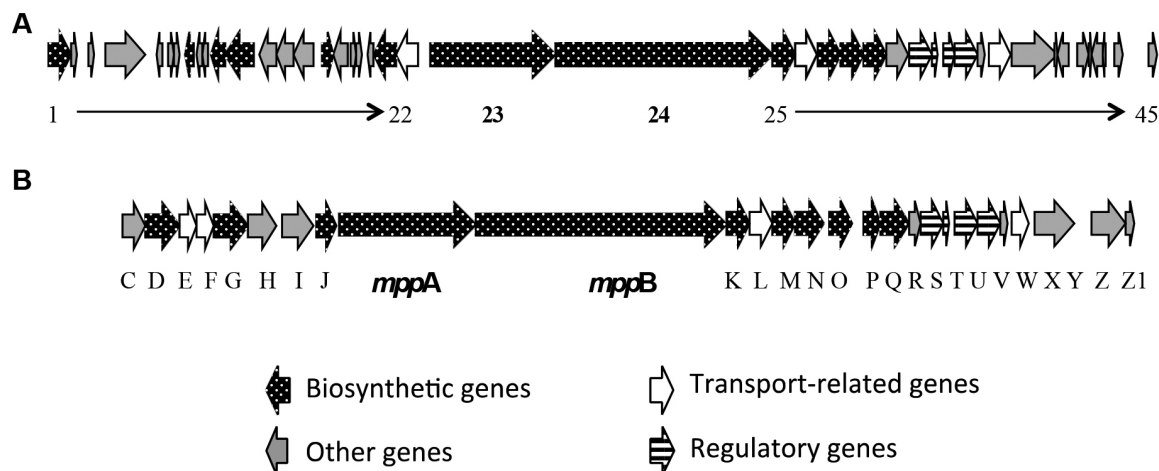


Figure 10. (A) Mannopeptimycin gene cluster organization in *Streptomyces wadayamensis* A23 (adapted from antiSMASH output); (B) mannopeptimycin gene cluster organization in *Streptomyces hygroscopicus* NRRL 30439 (adapted from reference 37).

*mppH* codifies for polyprenyl fosfo-mannosyl transferases, *mppJ* for a methyltransferase, *mppG* for a polyprenyl mannose synthase, *mppD* for a sugar binding protein and *mppC* for an acetyltransferase. For mannopeptimycin BGC from *S. wadayamensis* it is possible to describe several unknown function genes (labelled in grey) and two glucose specific phosphotransferase (genes 12 and 21) among others. Under the fermentation conditions we couldn't identify the production of mannopeptimycin, but still the information is genome encoded and is under investigation.

All the molecules identified so far shows that *Streptomyces*

*wadayamensis* A23 has a high capability to produce antibiotics. Still, we need to correlate the chemical entities with the biochemical activity prompted to inhibit the growth of pathogens as observed. Studies related to the unknowns metabolites produced by A23 strain are in progress.

## Conclusions

The genome sequencing of *Streptomyces wadayamensis* A23 combined with gene annotation and mass spectrometry pattern evaluation revealed the strain is a promising producer

of several secondary metabolites. As such tool improves the dereplication approach it was possible to observe the production of the siderophores deferroxamines B and E, several antimycins antibiotics and the polyene candicidin. These results open a wide field of research encompassing “from genome to natural products” as a promising strategy to find new biological and pharmacological targets.

## Supplementary Information

Supplementary Information (LC, LC-MS and MS/MS data) is available free of charge at <http://jbc.sqb.org.br>.

## Acknowledgments

We gratefully acknowledge FAPESP (project grant 2014/12727-5 to L. G. O. and 2010/51677-2 to M. N. E.), PETROBRAS (grant 4712-0), and the University of Campinas. C. F. F. A. and B. S. P. acknowledges CNPq (studentships 162191/2015-4 and 130933/2015-5). A. B. G. is grateful for the PIBIC-SAE studentship. R. S. acknowledges FAPESP (studentships 2013/12598-8 and 2015/01013-4) and L. G. O. is also grateful for the grant support provided by IFS (Grant Support No. F/4735-1) and by the program For Women in Science (2008, Brazilian Edition).

## References

- 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.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabinowitz, E.; Rajandream, M.-A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A.; *Nature* **2002**, *417*, 141.
- Crawford, J. M.; Clardy, J.; *Chem. Commun.* **2011**, *47*, 7559.
- Schmidt, E. W.; *Nat. Chem. Biol.* **2008**, *4*, 466.
- Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S.; *Nat. Biotechnol.* **2003**, *21*, 526.
- Corre, C.; Challis, G. L.; *Chem. Biol.* **2007**, *14*, 7.
- Challis, G. L.; Ravel, J.; *FEMS Microbiol. Lett.* **2000**, *187*, 111.
- Gross, H.; Stockwell, V. O.; Henkels, M. D.; Nowak-Thompson, B.; Loper, J. E.; Gerwick, W. H.; *Chem. Biol.* **2007**, *14*, 53.
- Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S.; *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10376.
- Tohyama, S.; Eguchi, T.; Dhakal, R. P.; Akashi, T.; Otsuka, M.; Kakinuma, K.; *Tetrahedron* **2004**, *60*, 3999.
- McAlpine, J. B.; Bachmann, B. O.; Pirae, M.; Tremblay, S.; Alarco, A.-M.; Zazopoulos, E.; Farnet, C. M.; *J. Nat. Prod.* **2005**, *68*, 493.
- Fischbach, M. A.; Walsh, C. T.; *Chem. Rev.* **2006**, *106*, 3468.
- Banskota, A. H.; McAlpine, J. B.; Sørensen, D.; Aouidate, M.; Pirae, M.; Alarco, A.-M.; Omura, S.; Shiomi, K.; Farnet, C. M.; Zazopoulos, E.; *J. Antibiot.* **2006**, *59*, 168.
- Banskota, A. H.; McAlpine, J. B.; Sørensen, D.; Ibrahim, A.; Aouidate, M.; Pirae, M.; Alarco, A.-M.; Farnet, C. M.; Zazopoulos, E.; *J. Antibiot.* **2006**, *59*, 533.
- Lin, X.; Hopson, R.; Cane, D. E.; *J. Am. Chem. Soc.* **2006**, *128*, 6022.
- Scherlach, K.; Hertweck, C.; *Org. Biomol. Chem.* **2006**, *4*, 3517.
- Wenzel, S. C.; Meiser, P.; Binz, T. M.; Mahmud, T.; Müller, R.; *Angew. Chemie Int. Ed.* **2006**, *45*, 2296.
- Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udway, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L.; *J. Am. Chem. Soc.* **2006**, *128*, 14754.
- Liu, C.; Zhu, J.; Li, Y.; Zhang, J.; Lu, C.; Wang, H.; Shen, Y.; *ChemBioChem* **2015**, *16*, 998.
- de Oliveira, L. G.; Tormet Gonzalez, G. D.; Samborsky, M.; Marcon, J.; Araujo, W. L.; de Azevedo, J. L.; *Genome Announc.* **2014**, *2*, 1.
- da Cruz, P. L. R.; Giarola, L. R.; Moraes, S. S.; da Silva, D. E. S. G.; Marcon, J.; Azevedo, J. L.; Araújo, W. L.; de Oliveira, L. G.; *Quim. Nova* **2015**, *38*, 333.
- <http://www.prosolia.com/products/firefly>, accessed in June 2016.
- <http://www.imzml.org/>, accessed in June 2016.
- <http://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>, accessed in June 2016.
- Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C.; *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1743.
- Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T.; *Genome Res.* **2003**, *13*, 2498.
- Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. a.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H.; *Nucleic Acids Res.* **2015**, *43*, W237.
- Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. a.; Weber, T.; Takano, E.; Breitling, R.; *Nucleic Acids Res.* **2011**, *39*, W339.
- <http://mibig.secondarymetabolites.org/index.html>, accessed in June 2016.
- <http://dnp.chemnetbase.com>, accessed in June 2016.
- Köster, W.; *Res. Microbiol.* **2001**, *152*, 291.

31. Angolini, C. F. F.; Vendramini, P. H.; Araújo, F. D. S.; Araújo, W. L.; Augusti, R.; Eberlin, M. N.; de Oliveira, L. G.; *Anal. Chem.* **2015**, *87*, 6925.
32. Ueki, M.; Suzuki, R.; Takamatsu, S.; Takagi, H.; Uramoto, M.; Ikeda, H.; Osada, H.; *Actinomycetologica* **2009**, *23*, 34.
33. Kim, H.; Esser, L.; Hossain, M. B.; Xia, D.; Yu, C.-A.; Rizo, J.; van der Helm, D.; Deisenhofer, J.; *J. Am. Chem. Soc.* **1999**, *121*, 4902.
34. Atta, H. M.; Ahmad, M. S.; *Aust. J. Basic Appl. Sci.* **2009**, *3*, 126.
35. Seipke, R. F.; Barke, J.; Brearley, C.; Hill, L.; Yu, D. W.; Goss, R. J. M.; Hutchings, M. I.; *PLoS One* **2011**, *6*, e22028.
36. Chen, S.; Huang, X.; Zhou, X.; Bai, L.; He, J.; Jeong, K. J.; Lee, S. Y.; Deng, Z.; *Chem. Biol.* **2003**, *10*, 1065.
37. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonisi, H. M.; Gerwick, W. H.; Dorrestein, P. C.; *J. Nat. Prod.* **2013**, *76*, 1686.
38. De Voe, S. E.; Kunstmann, M. P.; *US pat. US3495004DA* **1970**.
39. He, H.; Williamson, R. T.; Shen, B.; Graziani, E. I.; Yang, H. Y.; Sakya, S. M.; Petersen, P. J.; Carter, G. T.; *J. Am. Chem. Soc.* **2002**, *124*, 9729.
40. Magarvey, N. A.; Haltli, B.; He, M.; Greenstein, M.; Hucul, J. A.; *Antimicrob. Agents Chemother.* **2006**, *50*, 2167.

Submitted: March 22, 2016

Published online: June 10, 2016

**FAPESP has sponsored the publication of this article.**