CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Omics based approaches for the identification of novel bioactive secondary metabolites from marine sponge derived bacterial isolates
Authors	Patry, Sloane
Publication date	2021-02
Original Citation	Patry, S. A. D. 2021. Omics based approaches for the identification of novel bioactive secondary metabolites from marine sponge derived bacterial isolates. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2021, Sloane Anaïs Deborah Patry. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-13 05:01:56
Item downloaded from	https://hdl.handle.net/10468/11872



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Omics based approaches for the identification of novel bioactive secondary metabolites from marine sponge derived bacterial isolates.

Sloane Anaïs Deborah Patry

[ORCID: 0000-0001-8300-5760]

A thesis submitted for the degree of Doctor of Philosophy

National University of Ireland, Cork, School of Microbiology, February 2021

Head of School: Prof. Paul O'Toole

Supervisor: Prof. Alan Dobson

Table of contents

Table of contents 2
Declaration 6
Dedication 7
List of figures 8
List of tables 11
Acknowledgments 13
Abstract 20
1 Chapter 1: Introduction 22
1.1 The antimicrobial resistance crisis 22
1.1.1 The AMR world threat and the need for new bioactive antimicrobial compounds. 22
1.1.2 List of AMR Priorities 26
1.1.3 Counter methods including new drug discovery 29
1.2 Marine Genetic Resources (MGR) 34
1.2.1 Marine biodiscovery pipeline 34
1.2.2 Marine natural products (MNPs) as a source of new bioactive secondary metabolites 36
1.2.3 Marine sponges 39
1.2.4* Marine *Streptomyces 40
1.3 "Omics" and bioinformatics methods for biodiscovery of marine microbial natural products. 44
1.3.1 Genomics 48
1.3.1.1 Secondary metabolites Biosynthetic Gene Clusters sm(BGC). 49
1.3.2 Metabolomics 53
1.3.2.1 (1) Chemical extraction 54
1.3.2.2 (2) MS- based techniques 55
1.3.2.3 (3) Manual dereplication strategies 58
1.3.2.4 (4) Nuclear magnetic resonance (NMR). 59
1.3.2.5 (5) Molecular networks 60
1.3.2.6 (6) Bioactivity mappings 61
References 65

2 Chapter 2: Utilization of OSMAC, Genomic mining, LC-MS and Metabolomic analysis to detect and identify potential new bioactive metabolites in marine *Streptomyces* strains, from deep-sea sponge isolate B226SN104 and shallow water sponge isolate SM3. 90

Abstract 90

Introduction 92

Results and Discussion 95

2.1 Genomics and microbiological activity of the *Streptomyces spp* isolates 95

2.1.1 Secondary metabolite gene clusters of the *Streptomyces spp* isolates. 95

2.1.2 Biological activities of B226SN104 and SM3 isolates 97

2.2 Metabolomics and chemical analysis 98

2.2.1 Temperatures involvement in the production of secondary metabolites from both B226SN104 and SM3 isolates. 98

2.2.2 Nutrient source media involvement in the production of secondary metabolites from both B226SN104 and SM3 isolates 103

2.3 Potential new metabolites detected from SM3 and B226SN104 crude extracts 109

Conclusion 126

Material and Methods 127

References 133

3 Chapter 3: Genome mining, together with metabolomic and molecular networking approaches to elucidate metabolite production in the marine sponge isolate *Streptomyces* SM9. 147

Abstract 147

Introduction 148

Results and Discussion 149

3.1 Bioactivity and detection of metabolites 150

3.1.1 Bioactivity screening 150

3.1.2 Detection of metabolites – chemical results 151

3.1.3 Detection of metabolites – metabolomic results 157

3.2 Identification of the secondary metabolite regions responsible of the biosynthesis of the Antimycin A_1, Desferrioxamine B, Desferrioxamine E and Surugamide A detected metabolites from SM9 extracts. 170

3.2.1 Genomic results – Identification of the secondary metabolite regions responsible of the biosynthesis of the Antimycin A_1 173

3.2.2 Genomic results – Identification of the secondary metabolite regions responsible of the biosynthesis of the Desferrioxamine B and Desferrioxamine E 177

3.2.3 Genomic results – Identification of the secondary metabolite regions responsible of the biosynthesis of the Surugamide A 181
Conclusions 183
Material and Methods 187
References 193
4 Chapter 4: Detection of bioactive secondary metabolites, from the marine *Streptomyces* SM9 crude extracts. 205
Abstract 205
Introduction 207
Results and Discussion 211
4.1 Bioactive molecular network approach 211
4.1.1 Bioactivities of crude extracts 211
4.1.2 Bioactivity mapping of the anti-cancer, anti-inflammatory and anti-fungal activities 216
4.1.2.1 Anti-cancer activity mapping 216
4.1.3 Bioactive molecules of interest 226
4.2 Additional anti-oxidant and anti-microbial screening of SM9-M19 crude extract against human bacterial pathogens. 230
4.2.1 Antimicrobial screening 230
4.2.2 Antioxidant activity 232
Conclusions 234
Material and Methods 237
References 247
5 General Discussion 251
Discussion 251
References 262
Appendix 265
Publication by the author 265
Other activities 265
Poster presentation communication 265
Marketing communications 266
PhD progress meeting communications 266
Short technical or scientific courses 267
Demonstration activities 268
International collaborations: 268
International secondments: 268

Supplementary figures.....272
Supplementary tables 283

Declaration

I declare that the research presented in this thesis is my work and has not been submitted to any other degree, either at University College Cork or elsewhere.

When contributions of other are involved, this were indicated clearly by references to the literature, material and methods and acknowledgment of collaborative research.

This work has been conducted and completed under the supervision of Prof. Alan Dobson in the School of Microbiology based at University College Cork, in Ireland.

In addition, part of this work was conducted in "host" universities during secondments, such as in the Marine Biodiscovery Centre (MBC) at the University of Aberdeen, in Aberdeen, Scotland (UK), supervised by Dr. Rainer Ebel ; and at the "Instituto Biochimica Proteine" (IBP) at "Consiligo Nazionale delle Ricerche" , at Napoli (Italy), supervised by Dr. Donatella de Pascale.

Dedication

Cette thèse est dédiée à mes parents, Bruno Patry et Laurence Patry née Cagniard, qui ont toujours cru en moi (même quand moi, je n'y croyais plus), m'ont apporté de l'espoir et m'ont enseigné comment ne jamais abandonner. Je ne serai jamais arrivée là sans tous vos sacrifices, votre amour et votre soutien sans faille pendant toutes ces années.

Je vous en serai éternellement reconnaissante, je vous aime fort.

Sloane.

This thesis is dedicated to my parents, Bruno Patry and Laurence Patry nee Cagniard, who always believed in me (even when I wasn't myself), supported me, bring me hope in the darkness and showed me how to never give up. I will never have achieve this without all your sacrifices, love and caring, during all those years.

I will always love you and be thankful.

Sloane.

List of figures

Figure 1: Schematic illustration of the number of deaths attributable to Antimicrobial Resistance (AMR) every year compared to major causes of death by 2050 worldwide (Source: O'Neill, 2014). 23
Figure 2: Schematic illustration of financial Antimicrobial Resistance's impact on World GDP in trillion of USD, by 2050 (Source: O'Neill, 2014). 25
Figure 3: A typical Marine biodiscovery pipeline as envisaged by EU funded project Pharmasea. (Source: http://www.ibp.cnr.it/research/donatella-de-pascale). The collaboration project brought researchers to some of the deepest and coldest places on the planet. Scientists from all around the globe worked together to collect and screen samples of mud and sediment from huge previously untapped oceanic trenches. The large-scale, four-year Pharmasea project was backed by more than €9.5 millions of EU funding and brought together 24 partners from 13 countries from industry, academia and non-profit organisations (Accessed on : http://www.ibp.cnr.it/research/donatella-de-pascale). 35
Figure 4: Origin of samples......36
Figure 5: Scientific classification and illustrations of Streptomyces species 41
Figure 6: Integrated Omics approaches general descriptions (a) and applications (b). .46
Figure 7: Representative biosynthetic pathways of (A) PK: 6-deoxyerythronolide B (6-dEB), (B) NRP: gramicidin S, and (C) PK/NRP hybrid: yersiniabactin (Source: Li et al., 2018). 50
Figure 8: Domain organization of the different type of PKSs. (Source: Lim et al,. 2016). 52
Figure 9: Design of OSMAC based approach developed in this study. (Adapted from Antoraz et al., 2015).....64
Figure 10: The OSMAC approach. Figure adapted from (Antoraz et al., 2015). 93
Figure 11: Heat map of the predicted secondary metabolite gene clusters in the genomes of Streptomyces isolates using AntiSMASH and heatmap function in R. 95
Figure 12: Molecular network of SM3 fermented at 28°C and 30°C (a). Venn diagram of the specific and shared detected compounds, including potential new compounds, for SM3 (b) and B226SN104 strain (c) fermented at 28°C and 30°C respectively. 99
Figure 13: Molecular network of SM3 fermented in six media (a). Enlarged sub-network view of the compounds Sarmentoside B in the cluster (b) and of the compound Cyclo (L-leucyl-L-PROPYL) isolated in the network (c). Chemical structure of manually dereplicated compounds (d). 107
Figure 14: Venn diagram of the specific and shared compounds (a), including potential new compounds (b), produced by SM3, detected in the four most promising media in both aqueous and organic phase......110
Figure 15: Venn diagram of the specific and shared compounds (a), including potential new compounds (b), produced by B226SN104 strain, detected in the four most promising media in organic phase......111
Figure 16: Venn diagram of the specific and shared metabolites, including potential new metabolites, produced by SM9 strain, fermented in 7 media at 28°C, detected chemically in crude extracts from both organic and aqueous phase, by LC-MS and manual de-replication. 152

Figure 17: Venn diagram of the specific and shared metabolites produced by SM9 strain in six fermentation media from the organic phase crude extract samples (a), including their potential new metabolites (b). Venn diagram of the specific and shared potential new metabolites produced by SM9 strain, detected in the four most promising media extracts from the organic phase crude extracts (c). 154
Figure 18: Molecular Network of SM9 fermented in 6 media (a). Enlarged view of the isolated metabolites Concanamycin B and Maculosin in the network (b). Enlarged sub-network view of the metabolites Antimycin A1 and Surugamide A in their respective clusters (c) and (d) respectively. Chemical structure of manually de-replicated compounds (e)......159
Figure 19: Enlarged sub-network view of the Desferrioxamine B, Desferrioxamine E, glu-ala-pro-pro and pro-thr-ala-ile metabolites in the metabolite cluster, all together with their relative chromatograms from the LC-MS and manual de-replication analysis. 169
Figure 20: Detail of NRPS region 1193.1 identified by AntiSMASH version 5.0.0 (a). KnownClusterBlast of the region cluster showing similarities and differences with the referenced BGC0000958 antimycin A_1 gene cluster (b). 174
Figure 21: Gene legend (a) from the NRPS region 1193.1 identified by AntiSMASH version 5.0.0, with the Blastp identifications obtained from NCBI database (b)......176
Figure 22: Detail of the Siderophore region 759.1 identified by AntiSMASH version 5.0.0 (a). KnownClusterBlast of the region cluster showing similarities with referenced desferrioxamine B and desferrioxamine E gene clusters (b). Details of the similarities and differences with the following referenced gene clusters: BGC0001453 desferrioxamine B (c) and BGC0001572 desferrioxamine E (c)......179
Figure 23: Detail of the NRPS region 1026.1 identified by AntiSMASH version 5.0.0 (a) KnownClusterBlast of the region cluster showing similarities (b) and differences (c) with the referenced BGC0001792 Surugamide A gene cluster......182
Figure 24: Bioactive natural discovery workflow, involving a bioassay-guided dereplication (A) and bioactive molecular networking (B)......208
Figure 25: Bioactivity mapping of the anti-cancer activity against the human melanoma cell line A2058, from SM9 crude extracts, with legend of the bioactive molecular network (a). View of the bioactive clusters containing the bioactive metabolites with significant bioactivity score and/or potential new metabolites (b and c). 218
Figure 26: Bioactivity mapping of the anti-inflammatory activity of SM9 extracts against the LPS induced TNF-α production in Thp-1 cells, with legend of bioactive molecular network (a). View of the bioactive clusters containing the bioactive molecules with significant bioactive score (b) as well as isolated bioactive molecules in the network (c). 223
Figure 27: Bioactivity mapping of the anti-fungal activity of SM9 extracts against C. glabrata CBS138, with legend of bioactive molecular network (a). View of the bioactive cluster containing the bioactive molecule with significant bioactive score as well as isolated bioactive molecule in the network (b). 225
Figure 28: Anti-bacterial assays against: (a) S.aureus 6538p ; (b) B.metallica LMG 24068, (c) A.baumannii AB13 and (d) P.aeruginosa PA01. 231
Figure 29: Anti-oxidant activity of SM9-M19 crude extract, at T30min, 490nm, visible in percentage of activity (%)......233

Supplementary Figure 1: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from SM3 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......272
Supplementary Figure 2: Chromatograms from LC-MS analysis of the organic crude extract samples from SM3 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......273
Supplementary Figure 3: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from SM3 when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......274
Supplementary Figure 4: Chromatograms from LC-MS analysis of the organic phase crude extract samples from SM3 when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......275
Supplementary Figure 5: Chromatograms from LC-MS analysis of the organic phase crude extract samples from B226SN104 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......276
Supplementary Figure 6: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from B226SN104 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......277
Supplementary Figure 7: Chromatograms from LC-MS analysis of the organic phase crude extract samples from B226SN104 when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......278
Supplementary Figure 8: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from B226SN104 – when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP. 279
Supplementary Figure 9: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from SM9 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......280
Supplementary Figure 10: Chromatograms from LC-MS analysis of the organic phase crude extract samples from SM9 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP)......281
Supplementary Figure 11: Gene details from the desferrioxamine gene cluster identified by AntiSMASH, with the related Blastp results from NCBI. 282

List of tables

Table 1: Emergent microbial threats. Adapted from Fair & Tor, 2014 including data from 2019 AR Threats Report; Asokan et al,. 2019; WHO antifungal report; WHO 2020; WHO 2017.....29
Table 2: Composition of the different media used for the fermentation of the Streptomyces sp SM3 and Streptomyces sp B226SN104 strains......97
Table 3: bioactivity table of the deep sea isolate B226SN104 and the shallow water isolate SM3 both tested by well diffusion assay against the relevant pathogenic species: Staphylococcus aureus (NCIMB 9518), Pseudomonas aeruginosa (PAO1), Bacillus subtilis (IE32), Escherichia coli (NCIMB 12210), Candida glabrata (CBS138), and Candida albicans (Sc5314).....97
Table 4: Some compounds produced by SM3 and B226SN104 in different conditions following the OSMAC approach......105
Table 5: Potential new compounds detected from LC-MS analysis of the crude extracts samples from SM3.....124
Table 6: Bioactivity of the shallow water isolate SM9 tested by well diffusion assays against the relevant pathogenic species: Staphylococcus aureus (NCIMB 9518), Pseudomonas aeruginosa (PAO1), Bacillus subtilis (IE32), Escherichia coli (NCIMB 12210), Candida glabrata (CBS138), and Candida albicans (Sc5314)......151
Table 7: Some interesting identified metabolites produced by SM9 in different media conditions following the OSMAC approach.....166
Table 8: Secondary metabolite region clusters of SM9, identified by AntiSMASH version 5.0.0. Region: the region number; Type: the product types as detected by antiSMASH; From, To: the location of the region (in nucleotides); Most similar known cluster: the closest compound from the MiBIG database, along with its type (e.g. t1pks+t3pks); Similarity: a percentage of genes within the closest known compound that have a significant BLAST hit to genes within the current region (Blin et al., 2019).....173
Table 9: Table of the NCCLS assay (National Committee for Clinical Laboratory Standards, broth microdilution method) against C. glabrata CBS138......211
Table 10: Assay category: anti-cancers, assay name: A2058_MTT......212
Table 11: Assay category: anti-inflammatory, Assay name: TNF_THPAIF......213
Table 12: Composite table of SM9 crude extracts bioactivities results, together with the number of potential new metabolites detected by previous LC-MS/MS analysis from chapter 2......215

Supplementary Table 1: MS data for SM3.1 aqueous phase crude extracts......284
Supplementary Table 2: MS data for SM3.1 organic phase crude extracts......287
Supplementary Table 3: MS data for SM3.2 aqueous phase crude extracts......290
Supplementary Table 4: MS data for SM3.2 organic phase crude extracts.....292
Supplementary Table 5: MS data for B226SN104.1 organic phase crude extracts.....300
Supplementary Table 6: MS data for B226SN104.1 aqueous phase crude extracts.....304
Supplementary Table 7: MS data for B226SN104.2 organic phase crude extracts.....311

Supplementary Table 8: MS data for B226SN104.2 aqueous phase crude extracts.....314
Supplementary Table 9: MS data for SM9 aqueous phase crude extracts 319
Supplementary Table 10: MS data for SM9 aqueous phase crude extracts 325

Acknowledgments

First of all, I would like to thank my supervisor, Prof. Alan Dobson, who chose me and trusted me for this project. During this big journey of learning to become a researcher, Alan, you guided me, supported me and gave me the independence to lead this project in the direction of my own interests and curiosities, and never blocked me while watching me grow and learn. You supported and guided me though this long journey and helped me to overcome so many obstacles. In addition, I am very grateful for your help regarding my English grammar, which was a lot of work! Thank you so much – I am not sure there is enough 'thank you' I can say to make it even with you and the amount you have done for me.

Secondly, I would like to thank the EU commission, MSCA Action, for funding and supporting this great ITN *Marpipe* project (*MarPipe - ID:* 721421). Your investment funded our research, gave us the opportunity to expand our knowledge and our training nationally and abroad, at international secondments and conferences. You supported our lives and finally succeeded to raise the *"next generation of marine biodiscovery scientists"* that I am very proud to be part of. Thank you.

Thirdly to Dr. David Clarke, I would like to thank you for giving me my first chance in academia, for believing in me and for introducing me to the international world of microbiology. I would never forget how you helped me when I first came to Ireland.

Fourthly, I would like to thank Dr. Stephen Jackson, or as we call him, Steve, who guided me, supported me, and showed me how to become a great scientist, and how to stand up for myself. You taught me science, supported me during obstacles, and always had a nice smile to bring me hope and reassure me when I needed it. You are a great scientist and an excellent human being and I really hope that one day I may reach your level. You are my Scientist role model! I also have not given up on making you love cats, maybe someday…

To my loving partner, Barry, who encouraged me to do a PhD in the first place, who believed in me, who was always there for me, supported me and reminded me that I was capable of it each time I doubted myself. You made me see the strength and abilities that were in me and that I did not even know were there. I would have not come to Ireland if it were not for you.

To past and present UCC office/lab colleagues, namely, Bruno, Arno, Lynn, Bea, Eduardo, Lekha, Erik, Jamie, Roisin, Halimah, Maureen, and Clodagh. You shared your smiles with me and morally supported me though our lovely coffee breaks! We shared good days and bad together, successes and failures in our experiments at the lab, and we also shared our thoughts on every subject possible just for the fun. Thank you so much for this.

To my Marpipe family, namely, Anky, Jane, Grant, Arianna, Kevin, Asimenia, Yannik, Daniel, Maria, Alejandro and Florent. We started this beautiful and crazy journey together, grew up together in this consortium and we learned so much from each other. We shared our smiles, our success, our failures, our ideas, our motivation and also our humour! All of you became my friends and I am really glad that I had this opportunity to meet each one of you. We are all so different, coming from different cultures and at the same time we are all so similar, we

always reach to each other with the same heart. I really hope that we will keep meeting together as a group post-Marpipe! During my PhD journey, you were my little light in the darkness – there is something very comforting in knowing that no matter where you are in the globe or in your PhD life, there is someone else in the exact same spot as you, despite the geography and despite the science field itself. I guess in a funny way, we all became PhD blood brothers and sisters ;)

With an additional special thanks to my closest Marpipe Friends Anky and Jane, I will never forget our secondment at Aberdeen with just the three of us: so many good memories.

To my secondment supervisors, Rainer and Donatella, who welcomed me in their labs in the friendliest way possible, who taught me so much about work, science but also about your cultures and your lovely cities. Thank you so much for your welcome, kindness, support and trust.

To my dear friend Chiara Papulino, for your friendship, your optimism and for sharing your smiles with me. You have been my little bright light during my secondments at both Aberdeen and Napoli. I am very glad that MarPipe gave me the opportunity to meet you and hope we will meet again.

To Barry's family, namely Neil, Noreen, Patrick and Catherine, who welcomed me in their family and became my adopted family in Ireland. Thank you for your very warm welcome.

To my dear, sweet boy, Calin (his name means Cuddle in French), who made my PhD life so much lighter and cuddly. His purr and cuddle in harsh times as well as in good times really helped me to cope with everything. Thanks to him, I am creating my own cattery of fluffy friends.

To Paddy, who since day one in UCC always helped me when I needed it and hasn't stopped since. I remember one day that I was locked in UCC because it was past 7pm and at this stage, my English was not great, so I did not know how to express that I was locked in the department and I needed help to open the doors and leave! Despite my bad English, you understood me, smiled at me and showed me the little button to click on the door to leave. I also remember when the autoclave broke and I had so many media bottles to sterilize for my big experiment that I panicked a bit! But you reassured me by sharing your own autoclave with me for my media. I really appreciated it. I have so many memories of you saving my day that I cannot recall all of them there but I am really glad that I had you near me at every occasion. Thank you for all your help during these years, thank you for your optimistic smile and thank you to always for talking to me and give me a boost. You really brighten up my days!

To Hilda Bohane, from the microbiology secretary's office, for helping me so many times with administrative paperwork. Thank you for your extraordinary kindness.

Finally to the bacteria around us, which are amazing, beautiful and without whom, we would not be here today, nor could I have done this great project.

Acknowledgments to my family and loved ones - in French

Je voudrais tout d’abord remercier mes parents, Bruno et Laurence Patry, qui ont toujours cru en moi et qui m’ont toujours soutenu. Vous avez fait de moi, qui je suis aujourd'hui et je n'en serais pas arrivée là sans tous vos sacrifices et sans tout votre amour pendant toutes ses années et surtout pendant cette thèse ; je vous enserai éternellement reconnaissante.

Ensuite, j'aimerais remercier mes frères, sans eux non plus, je ne serais pas arrivée en thèse. Merci Maxence pour ton aide pendant ma scolarité, de tes précieux conseils et je n’oublie pas nos appels téléphoniques, parfois jusqu’à deux heures du matin pour m’aider à finir mes exercices de maths. Merci Willem, pour m’avoir montré qu’il ne fallait pas hésiter à saisir des opportunités et aller les chercher soi-même, même jusqu'au bout du monde ! Tu m’as appris à m'aventurer en dehors des sentiers battus.

Merci à ma chère cousine Ines et à ma trendre marraine Annie, pour votre amour, votre présence malgré la distance géographique et malgré mes absences, votre soutien sans faille pendant toutes ces années et surtout pendant mon doctorat.

Merci aussi à mes cousins, Enrick et Grégory pour vos conseils pleins d’empathie et surtout sans aucun jugement. Nous sommes différents mais nous partageons des similarités dans notre vécu qui nous unissent. Merci d’avoir été là pour moi lorsque j'en ai eu besoin.

Je voudrais également remercier mes meileur.e.s ami.e.s en France, qui malgré la distance et malgré mes absences, n’ont pas douté une seule seconde de mon amitié, ont continué à me

soutenir et à m’aimer de la même façon. Merci à Charlène, Élodie, Rachid, Jodie, Stephanie, Alexandra, Justine et Fanny, pour votre soutien sans faille et pour votre amour sans limite temporelle et géographique.

Je voudrais également remercier mes meileur.e.s ami.e.s en France, qui malgré la distance et malgré mes absences, n’ont pas douté une seule seconde de mon amitié, ont continué à me soutenir et a m’aimer de la même façon. Merci à Charlène, Élodie, Rachid, Jodie, Stephanie, Alexandra, Justine et Fanny, pour votre soutien sans failles et pour votre amour sans limite temporelle et géographique.

Finalement, un grand merci à ma chère et tendre grand-mère, Mireille Patry, qui a toujours été présente pour moi et à mon défunt grand-père, Yves Patry, qui m’a enseigné la curiosité et à résoudre des problèmes à travers mon premier puzzle. Des années plus tard, je suis toujours aussi curieuse et j’essaie toujours de résoudre des problèmes, juste plus vastes et plus scientifiques. Je vous remercie tous les deux pour tout votre amour en provence d’ici et de l’au-delà.

"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale." Marie Curie.

Abstract

Antibiotic resistance is a major threat to public health worldwide which, urgently requires the discovery and development of novel antimicrobial drugs with new modes of action. The principal goal of this project is the discovery of new bioactive secondary metabolites that may lead to the discovery of new antibiotics to help fight against the antibiotic resistance crisis. The marine environment is a unique habitat where deep-sea microorganisms, because of their ability to adapt to this extreme environment, have the potential to produce novel secondary metabolites with potent biological activities. Furthermore, microbes associated with marine sponges are exposed to highly competitive environments both physiologically and nutritionally which is likely to promote the production of novel secondary metabolites. The genus *Streptomyces* produces secondary metabolic compounds that are rich in biological activity and marine *Streptomyces* strains host abundant small molecule biosynthetic gene clusters (smBGCs) which encode polyketides, Non ribosomal peptide synthases (NRPS), siderophore, bacteriocin and lantipeptides. In this respect, by focusing on *Streptomyces* samples taken from sponges in marine environments off the Irish coast, we aimed to characterize new marine derived bacterial strains and to explore their bioactive potential. In order to maximize the number of potential novel bioactive metabolites that could be be identified from each *Streptomyces* strain, and to generate the most detailed information underpinning their associated metabolic pathways, the One Strain MAny Compounds (OSMAC) approach, combined with a multi-omics approach, was employed in this project.

In total, the *Streptomyces* isolates (B226SN104, SM3 and SM9), produced 612 metabolites including but not limited to, Bisucaberin B, Maculosin, Desferrioxamine B, Desferrioxamine E, Concanamycin B, Lipoamicoumacin B, Probestin, Surugamide A, Fiscalin B, Gibbestatin B and Antimycin A1. The *Streptomyces* isolates produced metabolites active against the growth

of antimicrobial-resistant microorganisms categorized as an “Urgent” or “Serious” threat by the World Health Organization (WHO), including *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Candida* spp. In addition, other activities of biomedical interest were identified, namely anti-cancer, anti-inflammatory and antioxidant activities.

Furthermore, the omics approach allowed us to predict 19 potential bioactive and potential novel promising molecules from the *Streptomyces* SM9 isolate.

Taken together, this project demonstrated the significant potential the modified OSMAC/multi-omics approach could have in expanding the number of novel secondary and bioactive metabolites that could be generated from existing environmental microbial isolates. Furthermore, it has contributed to the fight against antimicrobial resistance by identifying as of yet potential undiscovered molecules with antimicrobial potential.

1 Chapter 1: Introduction

1.1 The antimicrobial resistance crisis

1.1.1 The AMR world threat and the need for new bioactive antimicrobial compounds.

Antimicrobial resistance (AMR) has already been recognized as a serious global health threat to public health systems in the world. This has resulted in a large number of high-level policy initiatives, such as The World Health Organisation (WHO), the United Nations (UN), Transatlantic Taskforce on Antimicrobial Resistance (TATFAR), the Global Antibiotic Resistance Partnership (GARP), the Global Health Security Agenda (GHSA), the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR), the European Antimicrobial Resistance Surveillance Network (EARS-Net) and the Central Asian and Eastern European Surveillance of Antimicrobial Resistance (CAESAR). *(Mushtaq, 2016; WHO, 2015a; WHO, 2011; Nahrgang et al., 2018*; IACG.2019; *WHO, O'Neill.2014).* Indeed worldwide reports and studies have shown the alarming signs of the drug-resistance crisis and the potential disastrous consequences of AMR, including the ground breaking report from the *Interagency Coordination Group on Antimicrobial Resistance* (IAGC) to the secretary-general of the United Nations on April 2019; which stipulates that if no action is taken, that drug resistant diseases could cause around 10 million deaths worldwide each year by 2050, force up to 24 million people into extreme poverty by 2030 and damage the economy to an extent that would be as catastrophic as the 2008-2009 global financial crisis (*UN Ad hoc Interagency Coordinating Group on Antimicrobial Resistance; WHO, O'Neill. 2014*). Furthermore, Dadgostar. 2019 and O'Neill. 2014, have both detailed the morbidity and mortality consequences of AMR affecting patients, predicted by 2050. Among them, the mortality rates

due to AMR by 2050 will be 10 million, (Figure 1), which confirm the 10 Million deaths due to AMR predicted by IACG. 2019. They predicted approximately 4.7 million deaths in Asia; 4.1 million in Africa; 390,000 in Europe; 509,000 in America and 22,000 in Oceania (Dadgostar. 2019; O'Neill. 2014) (Figure 1).

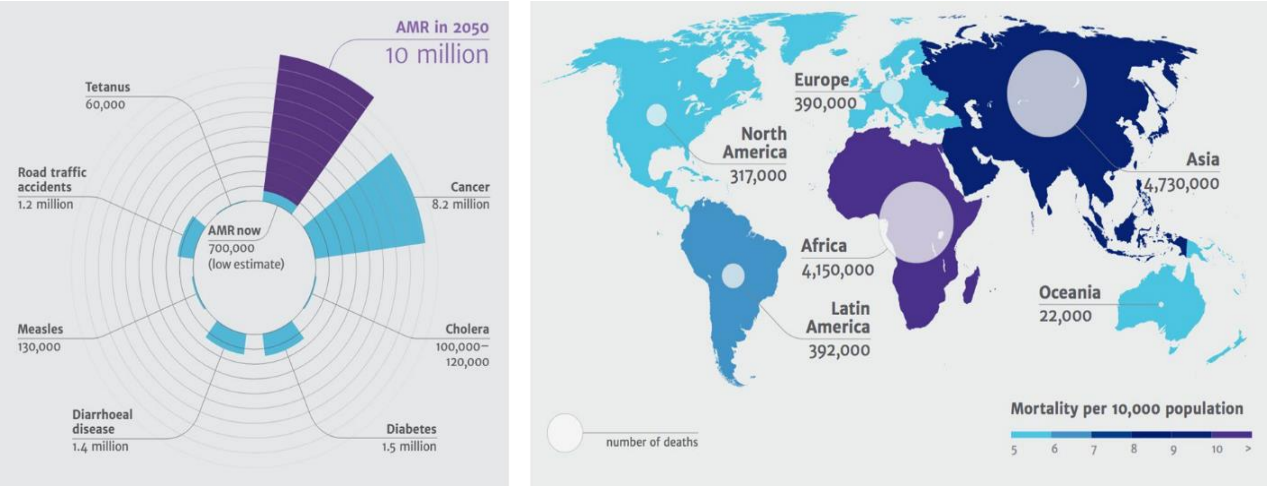


Figure 1: Schematic illustration of the number of deaths attributable to Antimicrobial Resistance (AMR) every year compared to major causes of death by 2050 worldwide (Source: O'Neill, 2014).

Currently across the world, at least 700,000 individuals lose their lives because of drug-resistant infections each year, including approximately 230,000 people who die from multidrug-resistant tuberculosis (IACG. 2019; O'Neill 2016). In addition, alarming levels of resistance have been reported in countries of all income levels, with the result that common diseases, including respiratory tract infections, sexually transmitted infections and urinary tract infections, are becoming untreatable and lifesaving medical procedures riskier to perform. (IACG. 2019; Fadela.2019). Indeed, many surgery procedures, including medical lifesaving procedures and routine procedures, such as hip operations, caesarean sections, bowel surgery, appendectomy, bone fracture repair, amputation and others, rely completely on the use of antibiotics to prevent and lower the risk of post-surgery infections (Dadgostar. 2019, Santoro-

Lopez. 2014; O'Neill.2014). This is even worse when considering organ transplantation, cancer treatment and childbirth. Indeed, organ transplantation exposes the patients to different infections, thus it will increase the likelihood of transplant failure and death (Dadgostar.2019, Santoro-Lopez.2014). Modern cancer treatments also suppress the patients' immune systems, making them more vulnerable to infections. Therefore without appropriate and effective antibiotics to prevent or treat infection, chemotherapy will also become a much riskier proposition and may lead to an increase in deaths due to cancer (O'Neill. 2014). Regarding the safety of childbirth, including caesarean sections, the consequences of increased antibiotic resistance in clinical isolates will also lead to increases in both maternal and infant mortality (O'Neill. 2014). Also, in a world without effective antibiotics and increased AMR, the spreading of nosocomial infections in hospitals may even result in the closure of hospital wings (Dadgostar. 2019).

In addition to the mortality rates, the IACG report also highlights the high financial cost of delaying investments and action against AMR by stipulating that the world will have to pay far more in the future to cope with the disastrous impact of uncontrolled antimicrobial resistance (IACG. 2019). This was also mentioned in both the O'Neill 2014 review and by Dadgostar.2019: which stipulate that if no action is being taken to tackle AMR, a total of $100.2 GPD trillion will be lost by 2050 (Figure 2). Multi drug resistant Tuberculosis alone could cost $16.7 trillion by 2050 (Dadgostar. 2019; The Economist 2019; TB Alliance 2019).

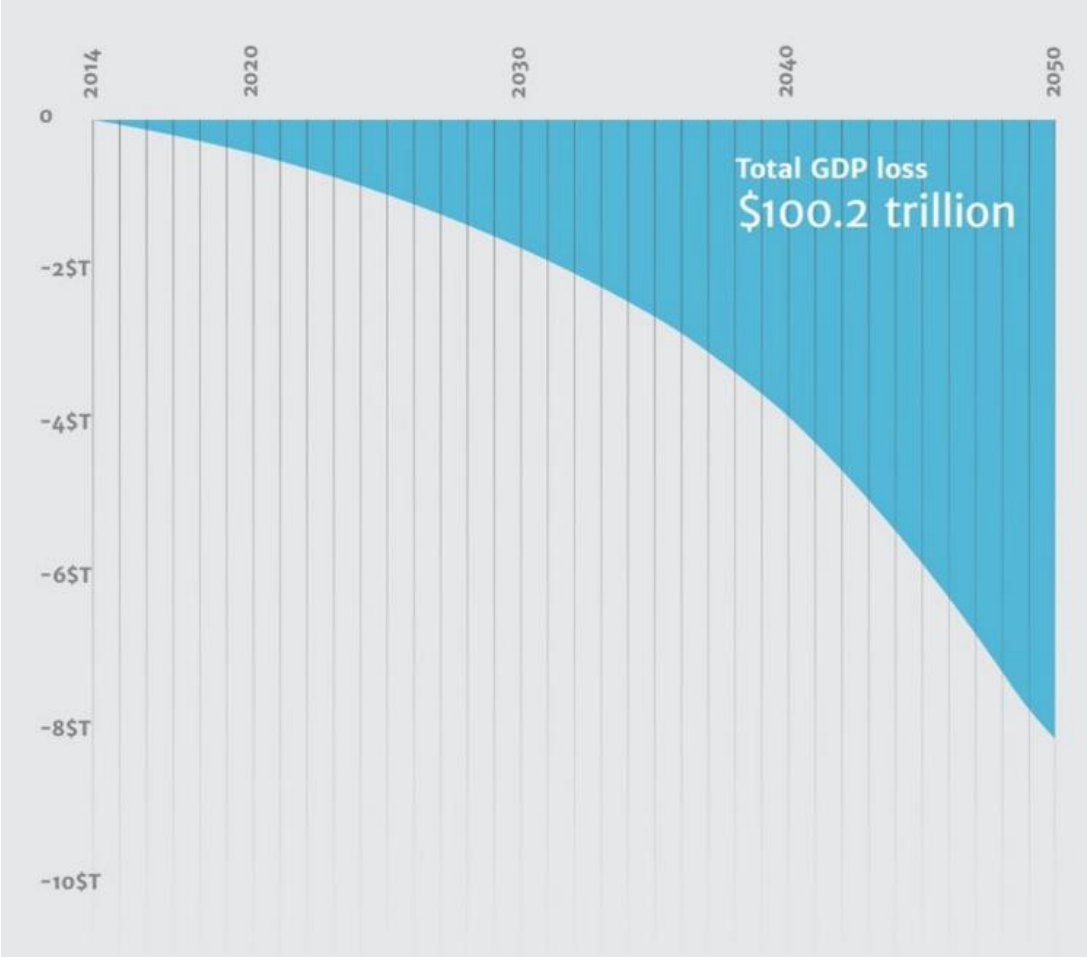


Figure 2: Schematic illustration of financial Antimicrobial Resistance's impact on World GDP in trillion of USD, by 2050 (Source: O'Neill, 2014).

The slow pace of new antibiotics discovery as well as the decline in investment by both industry and public funders in the area, have exacerbated the AMR problem. Indeed, for the last three decades from 1986 to 2016, not a single new class of antibiotic drug has been discovered and even when a new discovery does happen, it can take around 15-20 years for it to be developed from initial discovery to a clinical drug (Birkett et al., 2010). In the private sector, pharmaceutical companies and venture capital funds both tend to favour areas that have a higher commercial return. For example, the O'Neill 2016 report highlighted the comparison of development pipeline between oncology and antibiotics in 2014, with 800 new products in the development pipeline in oncology while fewer than 50 products were in the antibiotic pipeline. The same is

true for venture capital funds, the same method for focusing research on lucrative area is also observable, as highlighted in the O'Neill 2016 report which reported that of the 38 billion USD venture capital invested in pharmaceutical R&D between 2003 and 2013, only 1.8 billion USD was invested in antimicrobials research.

1.1.2 List of AMR Priorities

The Center for Disease Control and Prevention (CDC) released an antibiotic resistance threats report containing a list of bacteria and fungi which are of urgent, serious or concerning threats to human health (AR Threats Report, 2019) and in parallel, the WHO released a global list of human microbial pathogens which are of critical, high, and medium priority for R & D for new antibiotics based on specific prioritization criteria such as mortality, community burden, health-care burden, proof/prevalence of resistance, transmissibility and treatability (WHO antifungal report, 2020 ; WHO, 2020; Asokan et al.2019 ; WHO, 2017; Fair & Tor, 2014*)*. By combining these two reports it became clear that the major pathogens belonged to the ESKAPE group including *Escherichia coli, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterobacteriaceae*; however other pathogens such as *Clostridium difficile, Candida ssp,* Human Immunodeficiency Virus (responsible for acquired immunodeficiency syndrome AIDS) and *Plasmodium falciparum* (responsible for malaria) were all important targets for attention for treatment by novel antimicrobials (Table 1).

Micro-organism	**Pathogens**	**Antibiotic- Problematic resistance (WHO)**	**Risk level-WHO-priority level**	**Threat level-from CDC and WHO**
Bacteria	*Acinetobacter baumannii*	carbapenem-resistant	**Critical**	**Urgent threats**
Bacteria	*Pseudomonas aeruginosa*	carbapenem-resistant	**Critical**	**Urgent threats**
Fungal	*Candida auris*	fluconazole, amphotericin B, voriconazole, caspofungin (WHO)		**Urgent threats**
Bacteria	*Clostriduim difficile*	rifamycins, quinolones, aminoglycosides (Fair & Tor, 2014)		**Urgent threats**
Virus	Drug resistant HIV- HIVDR	NNRTIs (non-nucleoside reverse transcriptase inhibitors)		**Urgent threats**
Bacteria	*Enterobacteriaceae* *including:* - *Escherichia coli* - *Enterobacter spp.* - *Klebsiella pneumoniae* - *Serratia spp.,* - *Proteus spp.,* - *Providencia spp.,* - *Morganella spp* - *Salmonellae ssp*. - *Shigella ssp.*	carbapenem-resistant ESBL-producing (third generation cephalosporin-resistant) - fluoroquinolone-resistant - fluoroquinolone-resistant	**Critical** - Critical - Critical - Critical - Critical - Critical - Critical - Critical - High - Medium	**Urgent threats**

Bacteria	*Neisseria gonorrhoeae*	cephalosporin-resistant, fluoroquinolone-resistant β-lactams, quinolones, tetracyclines, macrolides (Fair & Tor, 2014)	**HIGH**	**Urgent threats**
Bacteria	*Mycobacterium tuberculosis*	rifampicin resistant TB-multi drug resistant TB (who) all classes except polymyxins (Fair & Tor, 2014)	**HIGH***	**Serious threats**
Bacteria	*Campylobacter ssp*.	fluoroquinolone-resistant	**HIGH**	**Serious threats**
Fungal	*Candida ssp* (excluding *C. auris*)	fluzocanole resistant Azole resistant		**Serious threats**
Bacteria	*Enterococcus faecium,*	vancomycin-resistant	**HIGH**	**Serious threats**
Bacteria	*Enterococci sp*.	vancomycin-resistant β-lactams, glycopeptides, aminoglycosides (Fair & Tor, 2014)	**HIGH**	**Serious threats**
Bacteria	*Helicobacter pylori*	clarithromycin-resistant	**HIGH**	
Parasite	*Plasmodium falciparum*	sulfadoxine pyrimethamineresistance artemisin resistance		Could become a major threat
Bacteria	*Staphylococcus aureus*	β-lactams, glycopeptides (Fair & Tor, 2014) methicillin-resistant, vancomycin-intermediate and resistant	**HIGH**	**Serious threats**
Bacteria	*Streptococcus pneumoniae*	penicillin-non-susceptible	**MEDIUM**	**Serious threats**

		β-lactams, macrolides, quinolones (Fair & Tor, 2014)		
Virus	*Haemophilus influenzae*	ampicillin-resistant	**MEDIUM**	
Bacteria	*Streptococcus* group A *Streptococcus* group B	erythromycin resistant clindamycin resistant		**Concerning threats**
Fungal	*Aspergillus fumigatus*	azole-resistant		**Watch list**
Bacteria	*Mycoplasma genitalium*	n/a		**Watch list**
Bacteria	*Bordella pertussis*	n/a		**Watch list**

Table 1: Emergent microbial threats. Adapted from Fair & Tor, 2014 including data from 2019 AR Threats Report; Asokan et al,. 2019; WHO antifungal report; WHO 2020; WHO 2017.

* *Mycobacterium tuberculosis*, the cause of human tuberculosis, was originally not subjected to review for inclusion in the WHO 2020 prioritization list as it is already a globally established priority for which innovative new treatments are urgently needed. However it is included in this table as part of the urgent threats.

1.1.3 Counter methods including new drug discovery

To counter the increasing rise in AMR, a global united multisectoral approach across many countries is required. As the threat is worldwide, the solution needs to be worldwide as well. International public health organisations and associations around the world, such as the WHO, UN and others, have recommended a coordinated action to brings together multiple sectors and stakeholders engaged in human, terrestrial and aquatic animal and plant health, food and feed production and the environment to communicate and work together in the design and implementation of programmes, policies, legislation and research to attain better public health outcomes (WHO, 2020).

Therefore the WHO recommendations, the IAGG report and the O'Neill report, have established 10 fronts on which to tackle AMR, as part of "One health response to antimicrobial resistance" are as follow: (1) public awareness; (2) sanitation and hygiene; (3) antibiotics in agriculture and the environment; (4) vaccines and alternative; (5) surveillance; (6) rapid diagnostics; (7) human capital; (8) new drug discovery; (9) global innovation found; and finally (10) International coalition for action (O'Neill., 2016).

(1) Public awareness involves creating global campaigns or events, such as the World Antimicrobial Awareness Week, that aims to raise awareness of antimicrobial resistance worldwide and encourage best practices among the general public, health workers and policy makers to slow the development and spread of drug-resistant infections (WHO, 2020). The improvement of AMR awareness is targeting the concept that patients and farmers do not demand, and clinicians and veterinarians do not prescribe, antibiotics when they are not needed (O'Neill., 2016).

(2) Sanitation and hygiene, the lack of hygiene and poor sanitation increases the risk of infection and therefore the use of antibiotics which will lead to drug resistance. The less people get infected, the less they need to use medicines such as antibiotics, and the less drug resistance arises (O'Neill, 2016). This step involves expanding access to clean water and sanitation in the developing world; reducing infections in health and care settings, including limiting superbugs in hospitals; and proper hand washing for everyone (Wellington et al., 2013).

(3) Reducing antibiotics in agriculture and the environment, is aimed at reducing the over-use of antibiotics in livestock and fish farming which will later be released into the environment or ingested by human. It is well understood that antibiotics are required in agriculture and aquaculture, for maintaining animal welfare and food security. However, much of their global use is not for treating sick animals, but rather to prevent infections or simply to promote growth (O'Neill, 2016). Antibiotics reach the environment in many ways, including through sewage systems and run-off from food-producing units such as farms, or through effluent from factories on AMR nearby water systems (O'Neill, 2016). This step involves regulations to set minimum standards for the treatment and release of manufacturing waste; and manufacturers to drive higher standards through their supply chains to take responsibility and correct any unnecessary environmental pollution (O'Neill, 2016; Wellington et al., 2013).

(4) Vaccines and alternative can prevent infections and therefore reduce the use of antimicrobials which will slow the rise of drug resistance. This step involves the development of existing vaccine to use against AMR, the development of new vaccines against AMR and founding for early-stage research (O'Neill, 2016).

(5) Surveillance ; this front include the Global Antimicrobial Resistance and Use Surveillance System (GLASS) which provides a standardized approach to the collection, analysis, interpretation and sharing of data by countries, territories and areas, and monitors the status of existing and new national surveillance systems, with an emphasis on representativeness and quality of data collection (WHO, 2020).

(6) Rapid diagnostics can change the way we are currently using antimicrobials in both humans and animals. Using data and testing technologies effective in informing the doctor's judgement to prescribe will improve the diagnostic and reducing unnecessary use of antibiotics, which will slow AMR and so making existing drugs last longer (O'Neill., 2016; WHO, 2020).

(7) Investing in human capital to improve the numbers, pay and recognition of people working in infectious disease to make the work attractive again and increase the candidates to work in AMR by improving career paths and rewards in these fields. From nurses and pharmacists in hospitals trained to improve stewardship, to microbiologists and laboratory scientists doing surveillance, diagnostic testing and R&D in academia, governments, public sector organisations or companies focusing on AMR-related specialties (O'Neill., 2016; WHO, 2020).

(8) New drug discovery is essential to counter AMR. The number of effective antimicrobial drugs to defeat infections that have become resistant to existing medicines must be dramatically increased. In addition, to increase this number, the investment in new drug discovery must improve as well as improving investment for existing ones (O'Neill., 2016). In addition, harmonised regulations and clinical trial networks can play an important role in this area to lower drug development costs and increase the investment (O'Neill., 2016; Wellington et al., 2013).

(9) and (10) Establish a global innovation fund for early-stage and non-commercial research, and build a global coalition for real action – via the G20 and the UN (O'Neill, 2016). AMR is a worldwide threat and thus cannot be solved by any one country or by

any one region as we live in a connected world where people, animal, food and microbes regularly travel around the globe (O'Neill, 2016; WHO, 2020).

New drug discovery

The marked increase in AMR together with the lack of innovation and effort in the discovery of new antibiotics drugs could result in humanity going back to "the dark ages of science and medicine" in a world without antibiotics (O'Neill, 2016; WHO, 2020). This could potentially result in a world where a simple infection would be fatal and a world where any attempt to save a life could be more life threatening than the disease itself. To counter this, there is a renewed interest globally in drug discovery, which is being spearheaded by the discovery of new natural product (NPs) including secondary metabolites. Secondary metabolites, including antibiotics, are widely believed to have been originally produced by bacteria as part of their chemical diversity, to allow them to compete and to survive against other bacteria in a competitive environment. As previously mentioned, NPs of microbial origin, are among many of our common antibiotics but have to date mainly been isolated from bacteria and fungi from terrestrial environments. To expand this repertoire scientists are now looking for NPs and secondary metabolites from as yet under or unexplored environment and ecosystems. This has led scientists to explore other more extreme environments, such as marine environments (with extremes in pressure), Antarctic environments (extreme cold temperature) and volcanic environments (extreme hot temperature).

1.2 Marine Genetic Resources (MGR)

1.2.1 Marine biodiscovery pipeline

Marine environments cover >70% of the earth's surface, and contain 95% of the biosphere, including millions of phylogenetically divergent microorganisms, and therefore represent an abundant source of biological and chemical diversity (Romano et al., 2018; Jensen et al., 1996; Hughes et al., 2010) which makes marine natural products a "treasure of the oceans", offering molecules with potential novel bioactivities and chemical biodiversity. The marine environment consists of ecosystems which possess a large variety of different environmental conditions. These involve ecosystems with variations in temperature, oxygen levels, nutrient availability, salinity and pressure, which has resulted in marine microorganisms which have evolved both phenotypically and genetically to survive in these extreme environments. This has resulted in many of these microorganisms producing enzymes and small bioactive molecules with potential use in a variety of different biotechnological applications (Reen et al., 2015). In addition, marine environments have peculiar geochemical features when compared with other land-based ecosystems, which is often reflected in the chemistry and properties of marine NPs produced by microorganisms inhabiting these marine ecosystems (Romano et al., 2018; Gribble et al., 2015).

The marine biodiscovery pipeline, from sample collection to drug lead is schematically depicted in Figure 3. It detailed all the step for marine drug biodiscovery starting with the collection of the sample at the source, either from existing partner collections (e.g. universities, culture collection, etc.) or most often, from an underexplored or new marine ecosystem (Jaspars et al., 2016a). These samples, such as, an entire organism, tissue, or microbial culture are then subjected to some form of screening for biological activity, as a preliminary mechanism to

confirm their activity (Firsova et al., 2017). When the source is of microbial origin and confirmed to be active, a medium to large-scale culture is then extracted by a chemical liquid-liquid extraction using organic solvents followed by a second confirmatory screen for bioactivity and or a screen for secondary metabolites on the crude extracts. When crude extracts are confirmed to be active or potentially novel, lastly these lead extracts follow the pipeline onto fractionation, purification, and chemical structural elucidation which ultimately confirms the bioactivity and or novelty of the lead.

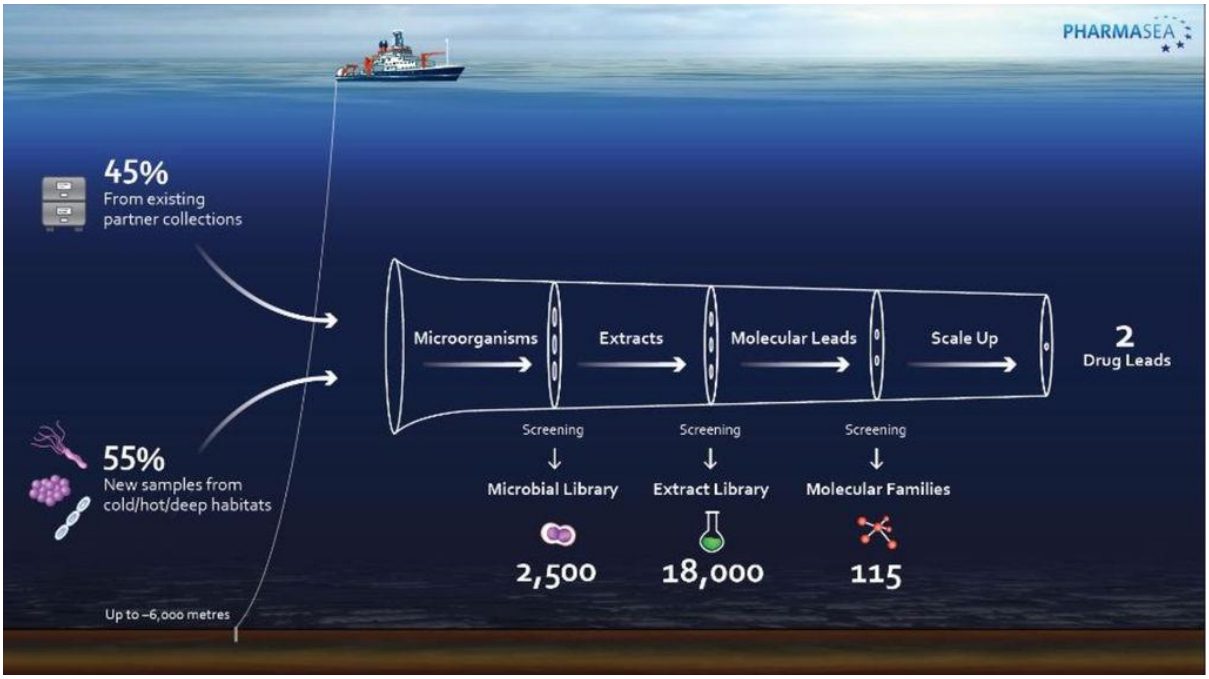


Figure 3: A typical Marine biodiscovery pipeline as envisaged by EU funded project Pharmasea. (Source: http://www.ibp.cnr.it/research/donatella-de-pascale). The collaboration project brought researchers to some of the deepest and coldest places on the planet. Scientists from all around the globe worked together to collect and screen samples of mud and sediment from huge previously untapped oceanic trenches. The large-scale, four-year Pharmasea project was backed by more than €9.5 millions of EU funding and brought together 24 partners from 13 countries from industry, academia and non-profit organisations (Accessed on : http://www.ibp.cnr.it/research/donatella-de-pascale).

1.2.2 Marine natural products (MNPs) as a source of new bioactive secondary metabolites

As detailed previously, marine natural products are often termed "a treasure of the oceans", offering novel bioactivities and chemical biodiversity (Romano et al., 2018; Jensen et al., 1996; Hughes et al., 2010). In particular it is well established that, microbes which are associated with marine sponges are exposed to highly competitive environments both physiologically and nutritionally which is likely to promote the production of novel secondary metabolites (Antoraz et al., 2015; Tang Xi et al., 2015; Lui et al., 20017; Jackson et al., 2018). In our case, this study focused on samples obtained from marine sponges, including deep sea sample and shallow water samples from *Lissodendoryx diversichela* (Porcupine Bank) and from *Halicona simulans* (Galway Bay) respectively (Figure 4).

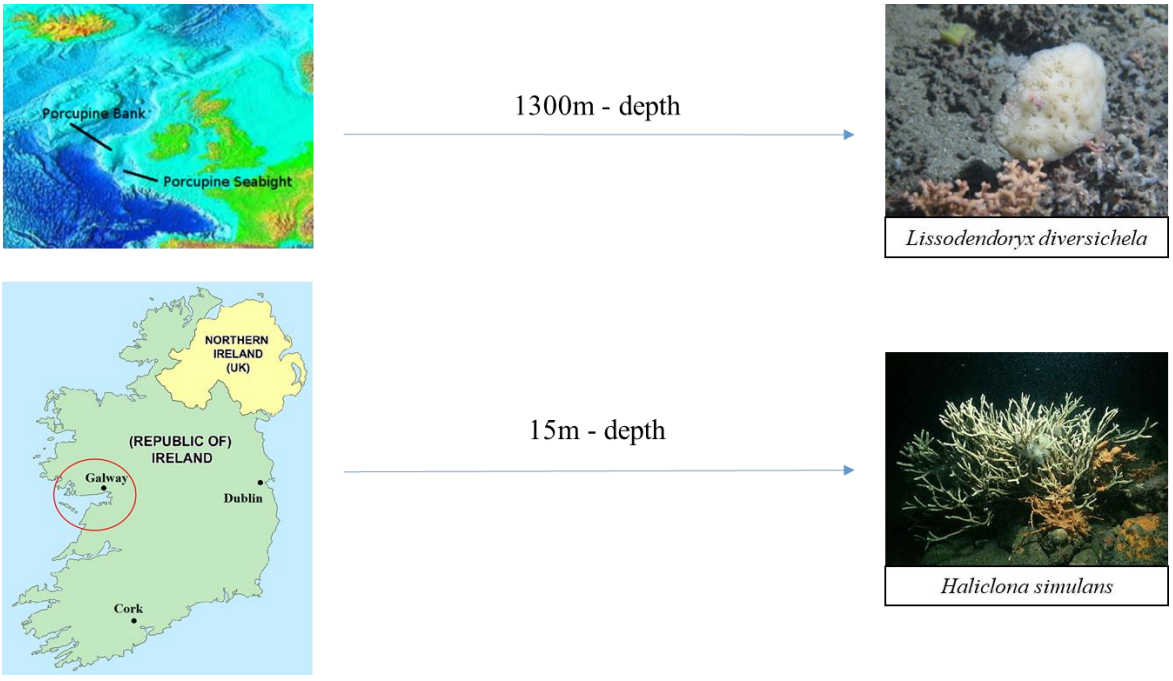


Figure 4: Origin of samples.

Porcupine banc - image credit available at https://en.wikipedia.org/wiki/Porcupine_Bank ; Galway bay/ Ireland map: Image adapted from Politics of Ireland - Wikipedia ; *Lissodendoryx diversichela* - Image credit: Crossman *et al.*, 2013. Available at Norwich Research Park Science (nrp.ac.uk); *Halicona stimulans*- Image credit: (Johnston, 1842); Picton, B.E. & Morrow, C.C. (2016). *Haliclona simulans* (Johnston, 1842). *Encyclopedia of Marine Life of Britain and Ireland.* Available at http://www.habitas.org.uk/marinelife/species.asp?item=C8630.

Both sponges (*Lissodendoryx diversichela* and *Halicona simulans*) harbour their own unique bacterial communities and are classified as low microbial abundance (LMA) sponges on the basis of these microbial communities (Kennedy et al., 2014; Giles et al., 2013; Marra 2019). These microbial communities are not only diverse but also appear to possess real potential as a source for the discovery of new medically relevant biological active agents (Kennedy et al., 2008) and play an important role in sponge biology (Kennedy et al., 2009) which make them relevant for this study. Secondary metabolites are generally produced by enzymes encoded by co-localized genes, forming biosynthetic gene clusters (BGCs) (Romano et al., 2018). Recent advances in genome sequencing have revealed that microbes often harbour various "cryptic" BGCs, which are not associated with currently chemically characterized molecules but are thought to be "silent" (not expressed) under standard laboratory conditions (Romano et al., 2018). Indeed, researchers have found that different growth conditions such as nutrients, culturing methods, and culturing parameters, together with other factors can affect the metabolic capacity and secondary metabolite biosynthetic potential of many microorganism. For example, *Streptomyces coelicolor* was initially known to encode BGCs for the production of 13 different classes of SMs, but the subsequent analysis of the newly sequenced genome revealed the presence of 16 additional BGCs which likely produce molecules with novel structures conditions (Romano et al., 2018; Bentley et al., 2002).

Overall, microbial diversity within marine ecosystems appears even greater than previously imagined, and has been described as "the tip of the microbial NP iceberg" (Romano et al., 2018). But despite the undoubted potential that marine microrganisms offer, problems continue to exist with the culturing of many of these microbes under laboratory conditions, resulting in

the phenomenon which is often referred to as “uncultivable microbes” (Romano et al., 2018). Indeed, only a tiny fraction of the marine microbial diversity present in the oceans can be suitably cultured in laboratory. In addition even if these microorganism can be cultivated under laboratory conditions, it does not always result in the synthesis of many of the secondary metabolites encoded within the biosynthetic gene clusters (BGCs) present in their genomes. This was initially highlighted by Williamson et al., 2006, who reported on the inconsistency between the number of biosynthetic gene clusters (BGCs) identified using bioinformatics approaches as potentially producing secondary metabolites and the actual number of chemically characterized secondary metabolites produced by any given microorganism. Therefore it is clear that the conventional approach of cultivating marine microorganisms under laboratory conditions must be complemented with alternative approaches in order to explore a larger portion of this potential microbial diversity. A number of approaches have been undertaken in an attempt to induce expression from these “silent” or “cryptic” BGCs which are present in the genome of theses marine microorganisms, but which are not expressed under typical laboratory conditions. Specific culture conditions are required to induce expression of these silent BGCs to unlock the chemical diversity that marine diversity can provide, leading to the discovery of novel bioactive molecules (Williamson et al., 2006). This has led to the development of both cultivation-dependent and independent methods to help unlock the potential of these “silent” BGCs (Romano et al., 2018). One of the methods that is typically used to awaken cryptic microbial biosynthetic pathways, is the One Strain MAny Compounds (OSMAC) approach, which is underpinned by the fact that one single microbial strain can produce different compounds when grown under different conditions, using different parameters of fermentation including nutrient source (e.g. salts, C-, N-, P-, S-sources), temperature, pH, aeration, co-culturing etc. This approach will be further detailed in chapter 1.

1.2.3 Marine sponges

Marine sponges (phylum: Porifera), are one of the oldest extant metazoans having branched off from other metazoans at least 650 million years ago (Rajendran, 2019; Kamalakkannan, 2015; Li *et al*., 1998). They are sessile filter feeding organisms, which display an impressive capacity to pump thousands of litres of seawater per kilogram of sponge bodyweight per day. Once the water has entered the sponge the microorganisms present in the seawater are either phagocytosed or are retained either as commensals or symbionts in the sponge mesophyll tissues. In this way sponges typically harbour dense microbial communities of up to 38% of their biomass (Vacelet 1975). This association between sponges and their microbiota appears to be unique and selective, with sponge symbionts possessing the ability to produce secondary metabolites, many of which have been shown to be chemically diverse bioactive compounds (Naughton et al., 2017; Faulkner, 2002; Laport et al., 2009; Hertiani et al., 2010; Turk et al., 2013). For example, 283 novel compounds were reported from the phylum Porifera in 2014 alone (Naughton et al., 2017; Blunt et al., 2014). It is believed that these metabolites may be part of a sophisticated armory of defence chemicals that help protect the sponge; who do not have complex immune systems, from predation by fish and other invertebrates. These metabolites may not only deter predators but may also help prevent growth on their surfaces by competitive bacterial species (Naughton et al., 2017). Most of the novel compounds that are produced by sponges are from a diverse number of structural classes, including steroids, terpenes, polyketides and peptides. (Kamalakkannan, 2015).

The chemical diversity of these bioactive compounds is immense, ranging from nucleosides, cyclic peptides and fatty acids to amino acid derivatives, and to terpenoids, macrolides, porphyrins, aliphatic cyclic peroxides and sterols (Naughton et al., 2017; Kamalakkannan, 2015) possessing pharmacological, cytotoxic, antibacterial, anticancer, anti-inflammatory,

immunomodulatory, neurosuppressive ,and anti-viral properties (Naughton et al., 2017; Sagar et al., 2010; Villa and Gerwick, 2010; Frota et al., 2012; Mehbub et al., 2014; Kamalakkannan, 2015). In the last number of years there has been mounting evidence suggesting that microbial-symbionts of sponge species are in fact the true producers of these bioactive molecules (Naughton et al., 2017; Donadio et al., 2007; Piel, 2009). Marine sponges and their associated microbes, together with seaweeds, soft and hard corals, gorgonians, and cnidarians continue to be important rich sources of structurally novel bioactive secondary metabolites, with potential biopharmaceutical applications (Rajendran, 2019; Kamalakkannan, 2015).

Members of the *Streptomyces* genus, which will be discussed later in this review, are known to be among the most prodigious producers of antibiotics and bioactive secondary metabolites. By focusing on marine *Streptomyces* samples taken from sponges in marine environments our hope in this study was to identify novel bioactive secondary metabolites.

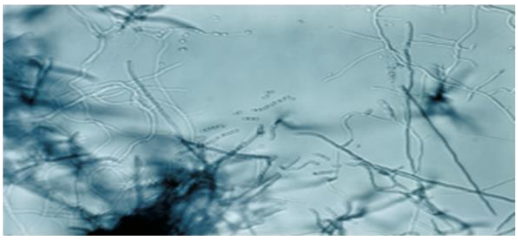
1.2.4 Marine *Streptomyces*

Streptomyces is a genus of gram positive bacteria from the Actinomycetaceae family, the Actinomycetales order and the Actinobacteria phylum (Waksman et al., 1943; Watve et al., 2001). They resembles fungi due to their filamentous form and display specific morphological differentiation which involves the formation of a layer of hyphae that can differentiate into a chain of spores (Figure 5). They inhabit a variety of different habitats, mostly terrestrial such as soil but they have also been isolated from different marine environments such as in fish, molluscs, sponges, seaweeds and mangroves, seawater and sediments. *Streptomyces* species exhibit an unusual, developmentally complex life cycle, an extreme chemical diversity and are well known producers of bioactive secondary metabolites with antifungal, antiviral, anti-

tumor, anti-hypertensive, anti-suppressant and anti-insecticidal activities; as well as being potent producers of antibiotics; which makes them of significant industrial and commercial importance (Jones et al., 2017). They are also known to produce enzyme inhibitors and pigments, as well as having biodegradation potential by producing biopolymer degradative enzymes (Malik et al., 2020); while marine *Streptomyces* are often used as probiotics in marine aquaculture applications (Mazón-Suástegui et al., 2020ab ; Tan et al., 2016; Das et al., 2010).

Streptomyces species

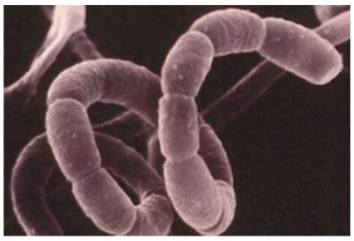
Scientific classification	
Kingdom:	Bacteria
Phylum:	Actinobacteria
Class:	Actinomycetes
Order:	Actinomycetales
Family:	Actinomycetaceae
Genus:	***Streptomyces***



Slide culture of a *Streptomyces* species – on microscope



Streptomyces coelicolor – on a petri dish



Streptomyces ipomoea – on electron micrograph

Figure 5: Scientific classification and illustrations of *Streptomyces* species

Picture credits below. Picture of Slide culture of *Streptomyces* on microscope obtained from the CDC Public Health Image Library, Image credit: CDC/Dr. David Berd (PHIL #2983), 1972. Available at https://phil.cdc.gov/Details.aspx?pid=2983. Picture of *Streptomyces coelicolor* on a plate (petri dish): Mervyn Bibb and Andrew Davis (John Innes Centre), Norwich Research Park (**NRP-68**); Available at http://images.norwichresearchpark.ac.uk/imagedetails.aspx?imgid=68. Picture of *Streptomyces ipomoea* on an electron micrograph: Individual spores of a spiral spore chain of *Streptomyces ipomoea*, are approximately 1 µm in diameter and 1-2 µm long. Image credit: Amanda Lawrence and C. A. Clark (Clark *et al*. 1981). Morphology of spore-bearing structures in *Streptomyces ipomoea*. Canadian Journal of Microbiology 27: 575-579). Available at

https://projects.ncsu.edu/cals/course/pp728/Sreptomyces_ipomoea/Streptomyces_ipomoea.htm

They have linear genomes which have a high GC content, often over 70% (Lee, *et al*. 2020; Hopwood, *et al.* 2006; Lee *et al*. 2019), which make them quite difficult to annotate and highlights the necessity to obtain high quality genome sequences when attempting to annotate their genomes *(*Madigan *et al*., 2005; Hwang *et al. 2019;* Lee, *et al*. 2020*).* Indeed since *Streptomycetes* have linear chromosome, it is often difficult to confirm the completeness of the assembled chromosome and one of the major obstacles in obtaining high quality genomic information from *Streptomycetes* is the low fidelity of sequencing techniques when dealing with high G C genomes and frequently repetitive sequences such as terminal inverted repeats *(*Harrison *et al.2014;* Lee, et al. 2020*).* Good quality whole genome sequence together with genome assembly and annotation are essential when attempting to predict the presence of secondary metabolite biosynthetic gene clusters (smBGCs). As of late 2019, of the 1,614 *Streptomyces* genomes that had been deposited in the NCBI Assembly database only 189 and 35 assemblies were designated as complete genome level and chromosome level, respectively. In addition more than 86% of the assemblies were draft-quality genome sequences, which contained fragmented multiple contigs or ambiguous sequences (Lee et al., 2020).

Several complete genome sequences of various *Streptomyces* species organisms, such as, *Streptomyces coelicolor* M145, as well as *Streptomyces rochei* 7434AN4, and others has been published and studied in an attempt to elucidate metabolic and regulatory networks, to elucidate complex smBGC regulation, to increase the production of commercially important compounds and to facilitate the discovery of novel and or new secondary metabolites (Paradkar et al., 2003; Lee, et al. 2020; Nindita. Et al, 2019). In addition the genomes of other *Streptomyces* strains, including *Streptomyces* sp. GS93–23, *Streptomyces* sp. 3211–3, and *Streptomyces* sp. S3–4, have also been published in studies contributing to the ecology and evolution of disease-suppressive microbial communities as in addition to be a source of new natural product

discovery, they are also used as soil inoculants in agriculture where they are known to function in promoting plant growth and in protecting plants from disease (Heinsch et al., 2019).

Approximately two thirds of all known natural antibiotics have been isolated from actinomycetes, with around 75% of these from the *Streptomyces* genus (Franco-Correa et al. 2010). It is also known that *Streptomyces* has the capacity to produce around a total of about 7600 bioactive compounds, making this genus the major antibiotic producer used for drug discovery and production in the pharmaceutical industry (Olanrewaju & Babalola, 2019). Thus given the aforementioned AMR crisis and the urgent need for new drug discovery; and knowing that the marine ecosystem is a huge potential source of novel isolates and new bioactive compounds coupled with the previous track record of *Streptomyces* as a potent source of antibiotics; there is a renewed interest in marine *Streptomyces.* A number of research groups are currently focusing on the genus and investigating their abilities to produce novel bioactive compounds (Cheng et al., 2020). Furthermore, recent advances in genome mining have shown that marine *Streptomyces* isolates possess a large number of unexplored silent secondary metabolite biosynthetic gene clusters (smBGCs) which makes them an invaluable natural source of new drug discovery (Xu et al., 2019; Lee, et al. 2020).

Indeed, not only are marine *Streptomyces* drawing more and more attention as a promising source of new natural products, but they also represent a valuable pool of genes for combinatorial biosynthesis of secondary metabolites. Genome sequencing and metabolic profiling of a number of diverse marine *Streptomyces* isolates has already resulted in reports that they produce new secondary metabolites and to the identification of the gene clusters necessary for their biosynthesis, which overall facilitates the discovery of new secondary metabolites. Among these, *Streptomyces* sp. MP131-18, has been reported to contain 36 gene

clusters involved in the biosynthesis of 18 different types of secondary metabolites which were predicted using AntiSMASH analysis and led to the identification of several new biologically active compounds including lagunapyrones D and E metabolites (Paulus et al, 2017). Another example is the discovery of a new metabolite, namely, dentigerumycin E, and its putative PKS-NRPS biosynthetic gene cluster, produced by *Streptomyces* sp strain JB5 when co-cultured with *Bacillus* sp strain GN1 (Shin et al., 2018). In addition, other marine *Streptomyces* have also displayed the potential to produce novel bioactive secondary metabolites, such as, *Streptomyces* sp. DUT11 which has the potential to produce novel lassopeptides and lantibiotics due to its production of tunicamycin I, V, and VII metabolites and biosynthetic gene clusters sharing similarities with tunicamycin and nonactin have also been identified in the genome of this strain (Xiao-Na Xu et al., 2018). Also *Streptomyces sp. IOPU* isolated from the marine sponge *Iotrochoto purpurea* which produced ten diverse bioactive compounds, namely, hexahydro-menaquinone MK-9, borrelidin, ferulic acid, N-acetylanthranilic acid, uracil, uridine**,** thymidine, sitosteryl-3β-D-glucoside, linoleic acid and methyl linoleate, have been reported and have been investigated as antioxidant and antimicrobial agents (Abd-Ellatif et al., 2019).

1.3 "Omics" and bioinformatics methods for biodiscovery of marine microbial natural products.

Up until recently, computational based bioinformatic analysis of large data sets using machine learning programs was traditionally the preserve of researchers who were computer literate with particular expertise in both the informatics and bioinformatics fields. Indeed,

computational data analysis using languages such as R and Python, required some level of fluency in writing instructions together with familiarity of relevant functions and packages; as well as some degree of programming skills to perform any integrative omics study such as data processing, integration, modelling and processing the large quantities of available data. However, thanks to the advances in genomics technology, increases in computational speed, memory storage capability to store big and large data, advances in analytical techniques and to the multitude of recent available resources to generate high-throughput data, such as using tools or software or programs like AntiSMASH, GNPS, Cytoscape, KNIME, Galaxy, Ugene, FinchTV, Antibase, ChemSpider, MZmine2, BLAST, GenBank, MG-RAST; genome mining has become the preserve of a larger cohort of scientists. These advances, together with a growing number of tutorials and courses available at workshops, webinars, or on public online platforms such as YouTube, have allowed a much greater participation by scientists in the analysis of bacterial genomic datasets. This so called "omics revolution", has made the analysis of large genomic datasets much more accessible to biologists and while still challenging, quite complex computational analyses is becoming much easier with time.

What is "omics"? Omics refers to the branches of sciences which regroup various discipline in biology that ends in the suffix – omic (Figure 6a). It combines genomics (including cognitive genomics, comparative genomics, functional genomics, metagenomics, neurogenomics, pangenomics and personal genomics), epigenomics, transcriptomics (including metatranscriptomics), lipidomics, proteomics (including immunoproteomics, nutriproteomics, and proteogenomics), metabolomics (including metabolomics and metabonomics) and ionomics. It is used to analyse the whole makeup of a given biological structure or function, to characterise, quantify, explore roles and relationships, at different levels, including at the gene (genomics), protein (proteomics) or metabolites levels (metabolomics) – (Figure 6a and 6b).

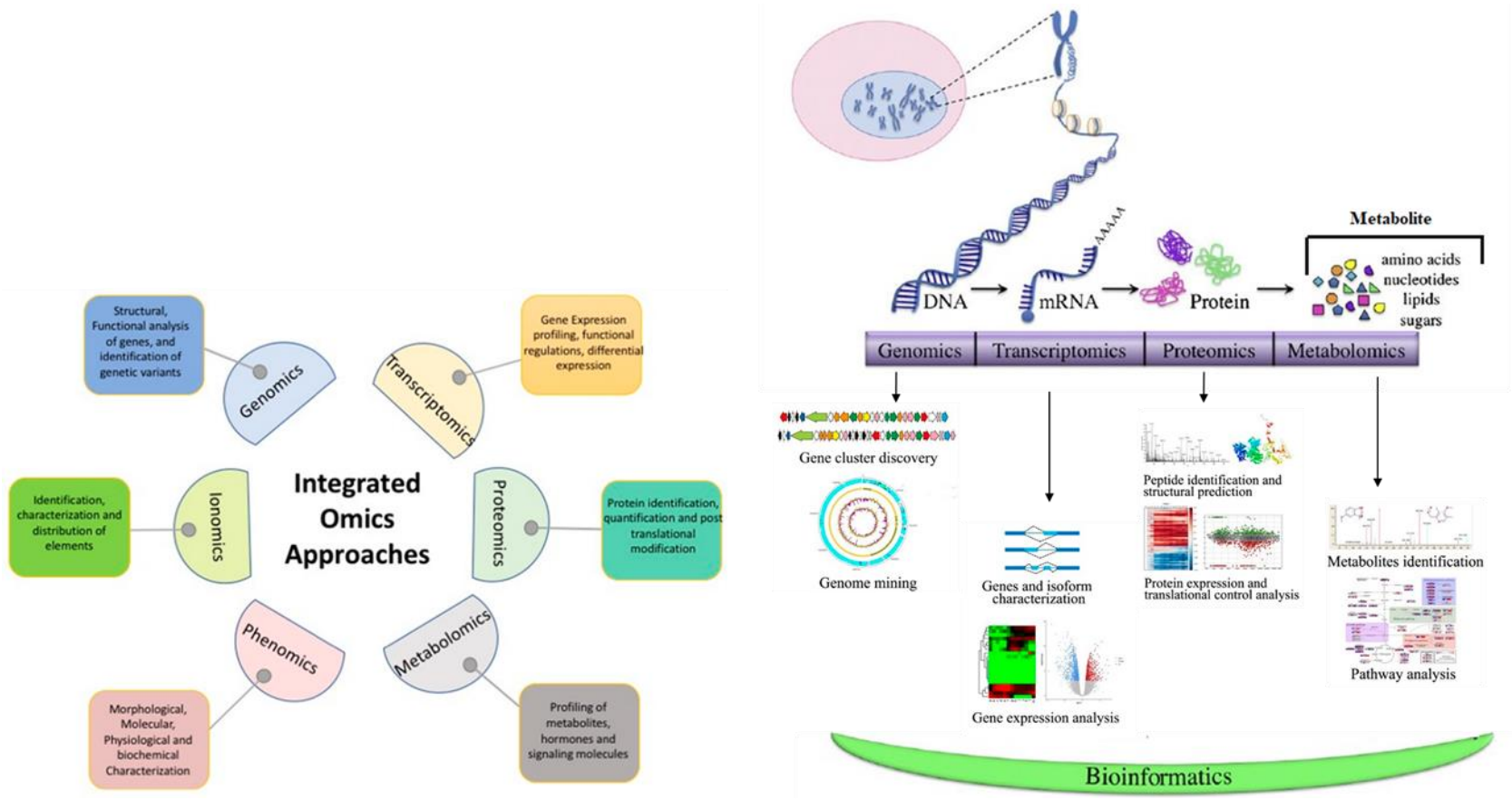


Figure 6: Integrated Omics approaches general descriptions (a) and applications (b).

Representation of different omics branches being used individually or in an integrated manner in science (a); applications and explanation of bioinformatics ‘‘omics’’ in marine biology (b) from a gene level to a metabolite level. (a) Image Source: Chaudhary *et al*., 2019. (b) Figure b adapted from both Akbar *et al*., 2020 and Fontana *et al*., 2012.

In marine biotechnology research, "-omics" based methodologies are typically used to investigate and ultimately exploit different marine bacterial species at different scales, from genomics with whole genome sequencing (WGS) to the elucidation of the genes and BGCs of microorganisms, while also providing a reference baseline for other omics based approaches. These include both transcriptomics and proteomics which are used to understand gene regulation and to investigate the functional properties of the microorganism; together with metabolomics, which is used to gain a fuller understanding of the phenotypic effects of gene expression with respect to metabolites and metabolic pathways involved in both primary and secondary metabolism (Ambrosino et al., 2019) (Figure 6b). In addition other omics based approaches can also be employed in marine science, including metagenomics which is typically used to gain a better understanding of the microbial populations associated with different marine ecosystems such as marine sponges as well as metatransriptomics to identify the genes being expressed and which are likely to be functional within these microbial populations (Kennedy et al., 2010). Researchers in this area has been greatly enhanced by access to the recently established Marine Metagenomics Portal (MMP) whose aim is to provide high-quality curated and accessible microbial metagenomics resources (https://mmp.sfb.uit.no/).

These multi-level (multi-omics) approaches, based on molecular and bioinformatic investigations at the genomic, transcriptomic, proteomic, and metabolomic levels are important in the discovery of marine natural products with novel bioactivities and to explore the key molecular processes involved in their production. Indeed, omics approaches, accompanied by the associated bioinformatic resources and computational tools for molecular analyses and modelling, have boosted the rapid advancement in biotechnology in general as well as in the area of drug discovery (Ambrosino et al., 2019).

1.3.1 Genomics

The capacity for an organism to produce bioactive secondary metabolites is embedded within its own genome with the presence of specific genes and BGCs involved in the biosynthesis of the metabolites. Therefore an analysis of the genomes of these organisms is often the first step in uncovering their potential in the marine drug discovery pipeline.

Genomics based studies typically involves sequencing marine bacterial genomes, using Next Generation Sequencing (NGS) and or Whole Genome Sequencing (WGS) based techniques following by the use of various computational based tools. These typically involve homology-based searches which are often conducted with BLAST (Basic Local Alignment Search Tool); resulting in the identification of the genetics basis of desirable traits, and knowledge of the genes responsible for a given phenotype; as well as allowing for phylogenetic analysis to establish the evolutionary relationships together with identifying potential Biosynthetic Gene Clusters (BGCs) including the discovery of silent biosynthetic gene clusters. In this respect bioinformatics tools such as AntiSMASH (antibiotics and Secondary Metabolites Analysis SHell) have been particularly useful (Blin et al., 2019). AntiSMASH is used to identify smBGCs within a genome and relies on the highly conserved sequences within the smBGCs to map their location. (Lee et al, 2020). AntiSMASH achieves this by employing a sequence alignment-based profile in a Hidden Markov Model (HMM) of genes that are specific for certain types of smBGCs (Lee et al, 2020; Blin et al., 2019). For example, AntiSMASH identifies smBGCs based on the highly conserved core biosynthetic enzymes and evaluates the results using a set of manually curated BGC cluster rules, followed by the discarding of false positives using negative models (e.g., fatty acid synthases are homologous to PKSs).

The most recent updated version of AntiSMASH (version 5), can predict up to 52 different types of smBGCs and can provide rapid gene annotation when bacterial genome data sets are submitted in FASTA format (Lee et al, 2020).

In this respect genomics has proven very useful for the characterization of marine species that are important in the production of secondary metabolites of interest for industrial, pharmaceutical, and green biotechnology applications (Kumar et al., 2018).

1.3.1.1 Secondary metabolites Biosynthetic Gene Clusters sm(BGC).

Secondary metabolites are produced from the multi enzyme complexes encoded by secondary metabolite biosynthetic gene clusters (smBGCs), which generally contain whole pathways that facilitate precursor biosynthesis, assembly, modification, resistance, and regulation of their product (Lee et al, 2020). The expression of these clusters is controlled by complex regulatory networks governed by both biotic and abiotic stresses which are typically found in the bacteria's natural habitat (Lee et al, 2020).

BGCs are classified based on their product as; saccharides, terpenoids, ribosomally synthesized and post-translationally modified peptides (RiPPs), non-ribosomal peptide synthetases (NRPS), and polyketide synthases (PKS) (Naughton et al., 2017). NRPS, PKS and Hybrid NRPS-PKS BGCs (Figure 7), are gaining more and more attention due to the large number of clinically relevant bioactive natural products that have been discovered from Non-ribosomal peptides and polyketides molecules which were are synthesized in a modular fashion by large multi-enzyme complexes (Naughton et al., 2017; Doroghazi and Metcalf, 2013; Xiong et al.,

2013). These multi-enzyme complexes, which consist of functional units known as modules, which contain at least three essential domains; (i) a catalytic domain responsible for selection of a specific monomer (ii) a carrier protein domain which facilitates attachment of the monomer after thioesterification and (iii) a second catalytic domain which functions in chain elongation (Naughton et al., 2017; Fischbach and Walsh, 2006), (Figure 7).

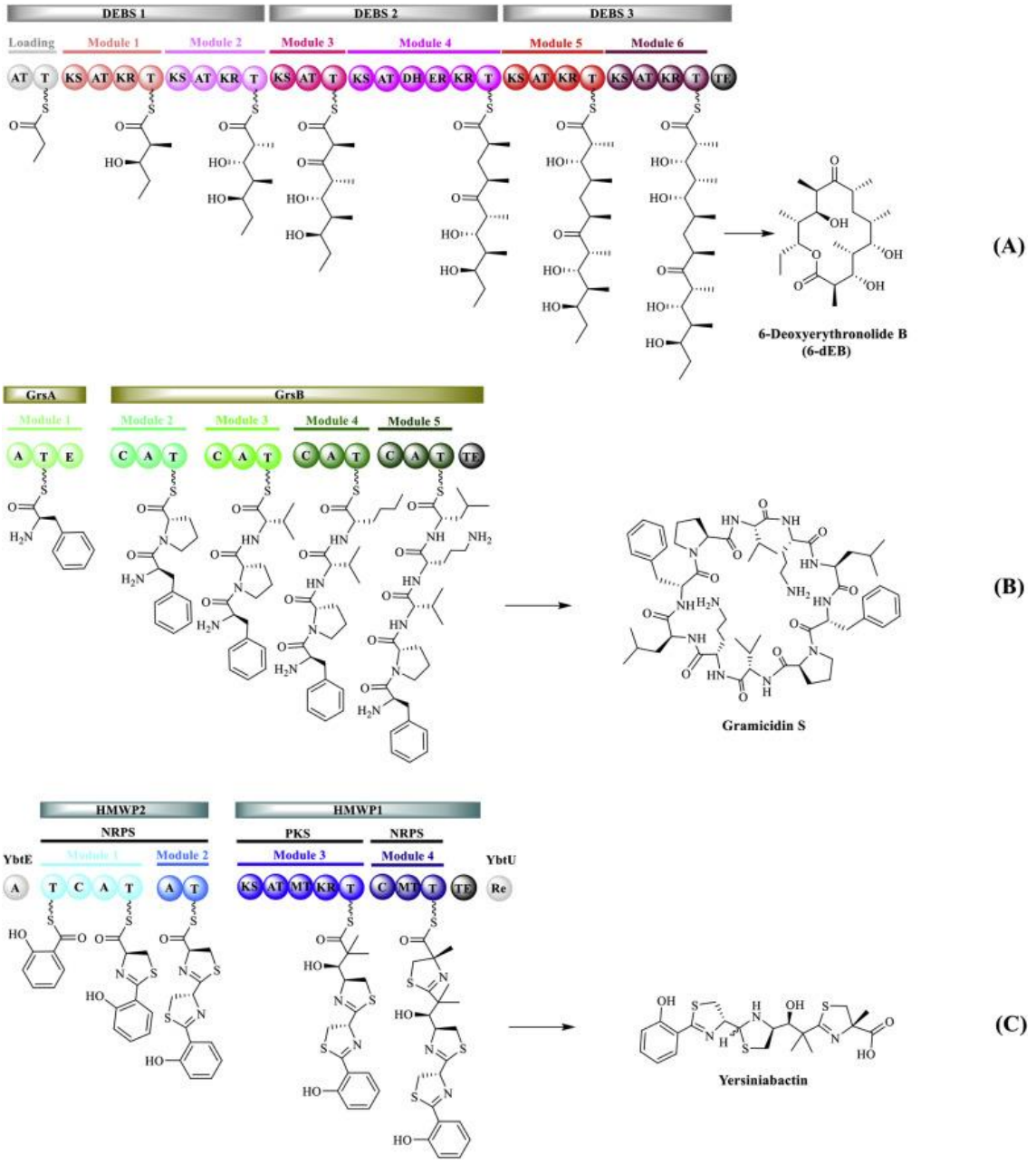


Figure 7: Representative biosynthetic pathways of (A) PK: 6-deoxyerythronolide B (6-dEB), (B) NRP: gramicidin S, and (C) PK/NRP hybrid: yersiniabactin (Source: Li et al., 2018).

Catalytic domains: AT, acyltransferase; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoylreductase; TE, thioesterase; A, adenylation; C, condensation; T, thiolation; E, epimerase; MT, methyltransferase; Re, reductase (Li *et al*,. 2018).

In a typical NRPS module, illustrated in Figure 7B, these domains are represented by a condensation domain (C), an adenylation domain (A) and a peptidyl carrier protein (PCP) domain (Naughton et al., 2017). Analysis of the amino acids which line the binding pocket of the A domain, facilitates prediction of the specific monomer incorporated at this site, which allows the prediction of the amino acid backbone structure of the resulting peptide (Naughton et al., 2017; Finking *et al*,. 2004).

In a typical PKS module, illustrated in Figure 7A, the domain are as follow, a ketosynthase (KS) domain, an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain (Naughton et al., 2017). However, there are 3 types of PKSs, classified by the organization of their catalytic domains (Naughton et al., 2017; Shen, 2003). These 3 types of PKSs are further detailed and illustrated in Figure 8. Type I PKS consist of multi-domain polyproteins which can be classified as either modular biosynthetic complexes, consisting of a series of enzymatic domains for each chain extension and modification step; or iterative, wherein a single set of enzymatic domains are reused several times during polyketide biosynthesis (Naughton et al., 2017). Type II PKS differ from Type I PKS, by the separate and discrete proteins which form the putative multienzyme complexes (Naughton et al., 2017; Shen, 2003). Lastly Type III PKS, consist of a simple homodimeric architecture and function essentially as condensing enzymes (Austin et al, 2003). In addition, tailoring domains such as heterocyclization and epimerisation (E) domains (in the case of NRPS modules) or β-ketoreductase (KR) and dehydrogenase (DH) domains (as in the case of PKS modules) which contribute to the overall structural diversity of the resulting final products, may be associated with each of these module (Naughton et al., 2017). Furthermore, NRPS- PKS Hybrid products (Figure 7C) can possess the structural and

functional similarities shared between NRPS and PKS gene clusters, which offers an even greater variety of secondary metabolites (Miller et al., 2002; Liu et al., 2004. Naughton et al., 2017).

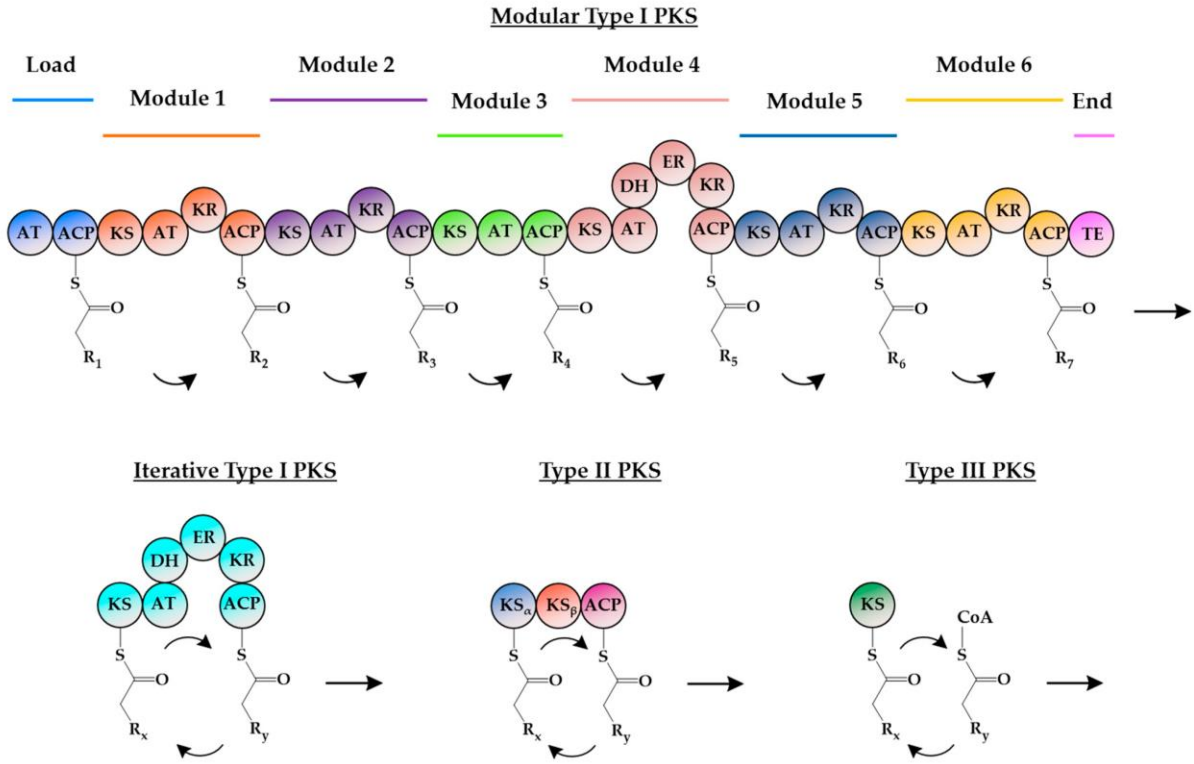


Figure 8: Domain organization of the different type of PKSs. (Source: Lim *et al*,. 2016).

Putative domains are represented by circles, in modular type I PKSs, functional domains are organized into several modules, with each module being responsible for a single decarboxylative condensation step in polyketide formation (Lim et al. 2016). For iterative type I PKSs, the functional domains are clustered in a single module, and each domain is used repeatedly during polyketide synthesis (Lim et al. 2016). Type II PKSs are dissociable multi-enzyme complexes, with each protein bearing a single and independent catalytic domain that is used iteratively during polyketide formation (Lim et al. 2016). Reactions by type III PKSs are also iterative but do not require an ACP for the attachment of the growing polyketide chain (Lim et al, 2016). AT: Acyltransferase; ACP: Acyl carrier protein; KS: Ketosynthase; KR: Ketoreductase; DH: Dehydratase; ER: Enoyl reductase; TE: Thioesterase; CoA: Coenzyme A.

1.3.2 Metabolomics

The metabolome of an organism, refers to the complete set of small-molecule metabolites that can be found within the organism and it directly correlates with gene expression and associated metabolite production, typically representing the phenotypical responses of the organism to a vast range of physiological and environmental stimuli (Yan et al., *2015;* Ambrosino et al., 2019; Bundy et al., 2005). Any disturbance or perturbation to a microbial system, will typically cause changes in the metabolite profile of that system (Yan et al., 2015). Metabolomics is the study of the whole repertoire of molecules in microbial cells, which represents a significant and speedily evolving part of the new systems biology area (Chandra Mohana *et al.2018).*

Relative to different omics fields, metabolomics remains a comparatively young discipline but has become a "key field" in the study of marine biodiversity, particularly due not only to the extreme variability in chemical and physical conditions that the marine environment presents (Ambrosino et al., 2019; Romano et al., 2017), but also due to the ongoing advances being made in metabolomics technologies that are leading to the generation of more accurate data. Indeed, instrumentation has become much quicker, more sensitive with increased reliability and automation resulting in the reduction of errors in the measurement of constituents, while also allowing for the application of increased statistical analyses for data validation (Chandra Mohana *et al.2018).* In addition, metabolomics provides data that is typically, more precise, germane and quantitative than genomics, transcriptomics or proteomics; by accurately quantifying the spectrum of biochemical changes, allowing the mapping of changes to various metabolic pathways within bacterial cells (Chandra Mohana *et al.2018).* While genome mining approaches aid in the prediction and identification of biosynthetic gene clusters from genes that are likely to produce molecules, metabolomic approaches aid in the analysis of the chemical diversity of a microorganism, helping with visualization of the metabolic pathways within the

cell and linking metabolite production to genes within gene clusters or operons. Thus metabolomics approaches complement genomics, as a way to confirm the predictions that the genomics based approaches can offer, with respect to the metabolites that are produced from the different microbial gene clusters (Chandra Mohana *et al.*, 2018; Patti *et al.*, 2012).

Metabolomic approaches in research are based on the direct profiling of molecules via Mass spectrometry (MS) techniques (Ambrosino et al., 2019; Fuhrer *et al.*, *2011;* Zampieri *et al.*, 2016). It involves chemical based techniques and analysis, such as in most of cases, (1) chemical extraction of molecule, (2) MS based techniques, (3) manual dereplications strategies, and (4) Nuclear magnetic resonance (NMR), to allow the chemically extraction, analysis, identification and characterisation of the metabolites being targeted. These approaches also involve the use of bioinformatic based skills , by employing bioinformatics resources such as ChemSpider, GNPS, MZmine 2, Cytoscape, AntiSMACH, Jupiter notebook among others, to visualize the metabolome together with the chemical diversity and the metabolite clusters on a (5) molecular network and or a (6) bioactivity map.

1.3.2.1 (1) Chemical extraction

In the field of marine natural products biodiscovery, there are several methods that can be employed to chemically extracts molecules or compounds, however, the most commonly used method is Liquid–liquid extraction (LLE), also known as partitioning. It is a useful method for separating compounds within a mixture, based on their relative solubilities in two different immiscible liquids, usually water (polar) and an organic solvent (non-polar). In marine natural products discovery, samples containing potential bioactive compounds, which are often from bacterial isolates from samples obtained from different marine environments; that are grown in different media; are mixed into a funnel with an organic

solvent such as ethyl acetate, methanol or dichloromethane, thereby producing crude extracts from 2 separate phases: the aqueous phase and organic phase. Potential bioactive compounds can be present in either of these phases. In general most of compounds tend to be in the organic phase which later facilitates the LC-MS analysis. However, some compounds may also be present in the aqueous phase, which can complicate the LC-MS analysis as the aqueous phase contain a lot of waste which make more background noise in the LC-MS.

1.3.2.2 (2) MS- based techniques

In untargeted metabolomics approaches, biological samples are most often processed, through an initial liquid chromatography (LC) or gas chromatography (GC) phase and subsequently subjected to MS analysis, leading to the detection of thousands of metabolites in a single eluate (Cox et al., 2014). There is a lot of different MS based techniques that can be used in metabolomics, but in general they fall into five categories: Gas Chromatography – Mass Spectrometry (GC- MS), Liquid Chromatography - Mass Spectrometry (LC - MS), Mass Spectrometry (MS), Nuclear Magnetic Resonance (NMR) and Integrated Applications (IA) (Cox et al., 2014). However, the most widely used technique is LC-MS. LC-MS Liquid Chromatography - Mass Spectrometry is an analytical technique that combines the physical separation capabilities of liquid chromatography with the mass analysis capabilities of mass spectrometry. LC-MS is a technique that has become routine with the development of electrospray ionisation (ESI) providing a simple and robust interface (Pitt, 2009) and has been used frequently in natural product discovery programs (Floros et al., 2016). Pitt, 2009, reported that LC-MS can not only provide superior specificity and sensitivity, but that it can also be used to develop highly accurate and

reproducible assays when combined with stable isotope dilution. Another advantage of LC-MS assays is the capacity to multiplex several analytes within a single analytical run with minimal incremental cost, which gives the potential to simplify laboratory set up and provide additional useful information including metabolite profiles, which are very useful in metabolomics (Pitt, 2009). Indeed, it can detect hundreds of compounds in a single run due to fast scanning speeds, thereby demonstrating a high degree of multiplexing possibilities (Pitt,. 2009). Modern mass spectrometers are highly sensitive and with the advances in technologies, the sensitivity of MS continues to improve (Pitt, 2009) and LC-MS have now reached an unparalleled sensitivity (Dunn et al., 2005) making them one of the cornerstones in marine natural product discovery. The goal of such metabolomics based efforts is typically to allow the identification of both known and unknown molecules in complex backgrounds (Floros et al., 2016).

Cox and co workers have previously described the commonly employed liquid chromatographic methods used in metabolomic analysis: normal-phase, reverse-phase, and hydrophilic interaction liquid chromatography (Cox et al., 2014). Normal-phase liquid chromatography (NPLC) readily separates highly non-polar metabolites (e.g. fatty acids, sterols, triacylglycerols, etc.) while reverse-phase liquid chromatography liquid chromatography (RPLC) is better suited for the separation of metabolites of medium to low polarity (e.g. alkaloids, flavonoids, glycosylated steroids, phenolic acids, etc.) (Cox et al., 2014). Hydrophilic interaction liquid chromatography (HILIC) can be used to separate a broad range of metabolites depending upon the bonded-phase of chromatographic media used (e.g. amino bonded, cationic boned, zwitterionic bonded, etc.) (Cox et al., 2014; Dunn et al., 2005; Lu W et al., 2008, Zhou B et al., 2012). Ultra-high performance liquid chromatography (UPLC) employs a short column with a small diameter particle size packing material and can operate at pressures of 5000 psi, for short elution times, but at the

expense of separation resolving power (Cox et al., 2014; Dunn et al., 2005, Zhou et al., 2012); while capillary liquid chromatography incorporates much longer capillary columns, which significantly increase the resolution of chromatographic separation but necessitate increased analysis time (Cox et al., 2014; Dunn et al., 2005). Multidimensional liquid chromatography (MDLC) combines two or more of the aforementioned chromatographic modes to enhance metabolite separation (Cox et al., 2014; Lu et al., 2008; Zhou et al., 2012).

In LC-MS, the series of metabolites produced following chromatography are subsequently exposed to an ion source which will dissociates them into ionic derivatives required for mass analysis (Cox et al., 2014). The ionization sources can include atmospheric pressure chemical ionization (APCI), atmospheric pressure photoionization (APPI), fast atom bombardment (FAB) and electrospray ionization (ESI). However, in metabolomics studies, ESI is by far the most common ionization source employed, due to its ability to ionize both polar and semi-polar compounds within a wide molecular weight range (Cox et al., 2014; Zhou et al., 2012). Mass spectrometers can be categorized as fourier transform ion cyclotron resonance (FTICR), ion trap (IT), orbitrap, quadrupole and time-of-flight (QTOF). Hybrid/tandem mass spectrometers are constructed by using two or more of the previously described mass analyzers in the same LC-MS system (Cox et al., 2014). For metabolomics studies, IT and QTOF are most commonly used, due to their ability to provide the accurate high mass resolution of MS/MS analysis (Cox et al., 2014). In the study that will be described later in this thesis, High Resolution LC-MS/MS analysis was performed with a Bruker MAXIS II QTOF instrument which runs in tandem with HPLC using a C18 column.

The flexible nature of LC-MS technologies coupled with the extreme diversity of metabolite chemo-physical properties or concentration, within a sample, can result in the introduction of a clear source of variation in the metabolomics data that is obtained (Cox et al., 2014). Furthermore, the ability to customize parameters to separate, detect and target a wide range of diverse molecules at low concentrations (e.g. pg mL^{-1}), makes LC-MS an ideal tool for metabolomic analysis (Cox et al., 2014; Shen et al., 2005; Kaddurah-Daouk et al., 2008). For example, Shen and co workers have reported the detection of over 5000 metabolites from a single experiment utilizing a capillary-RPLC-IT-TOF-MS/MS platform in a study on *Shewanella oneidensis* (Shen et al., 2005).

1.3.2.3 (3) Manual dereplication strategies

Manual dereplication strategies, including manual annotation of peak ions, are employed on metabolites from the LC-MS data to assess the possibilities of identification of the detected metabolites, following specific criteria and libraries. Part of this criteria is deduction of the possible molecular weights and molecular formulas using Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma); and by comparison with the relevant natural product library databases. The absence of any "hits" following searches on these libraries databases will deduce the presence of potential new metabolites, which would later need to be confirmed by NMR.

Metabolomics based approaches for natural product discovery, including manual dereplication, depends on the aforementioned natural product libraries which enable rapid characterization of known metabolites (Chandra Mohana et al., 2018). Libraries such as Chemspider, Antibase, The Dictionary of Natural Products, Reaxys and others, provide access to the structure information of a huge diversity of metabolites, including chemical

formula, molecular weight, monoisotopic mass and pharmacological activity (The_Royal_Society_of_Chemistry, 2015). In addition, the open-access platforms sharing of raw, processed, or liquid chromatography with tandem mass spectrometry data; such as the global natural products social molecular networking (GNPS) is very useful in the identification of the metabolites and their related metabolic pathways (Ambrosino et al., 2019; Wang et al., 2016). One of the often-perceived hurdles involved in natural product drug discovery is the chance of redundancy and many dereplication procedure have been established to deal with this potential drawback (Chandra Mohana et al., 2018). The structural information on metabolites, obtained at an early stage, provided by NMR and high-resolution MS-based platform, followed by its comparison with literature, can lead to identification without the need for laborious isolation processes, if structures have been previously reported (Chandra Mohana et al. 2018).

1.3.2.4 (4) Nuclear magnetic resonance (NMR).

NMR spectroscopy is a traditional powerful analytical tool that is essential for the structure elucidation of natural products (Mohamed et al., 2020; Kim et al., 2011; Kim et al., 2010; Mahrous et al., 2015). It is also a very useful technique for the structural elucidation of novel and/or unexpected compounds (i.e. identifiable by NMR but not LC-MS) including those with identical masses and/or different isotopomer distributions (Mohamed et al., 2020; Robinette et al., 2012; Tawfike et al., 2013). Unlike MS-based techniques, NMR is less biased as the results do not rely on the type of ionization condition or the preferences of the instruments that are used (Mohamed et al., 2020; Cai et al., 2007). NMR also allows the detection of both primary and secondary metabolites, simultaneously (Mohamed et al., 2020; Kim et al., 2009).

1.3.2.5 (5) Molecular networks

The data obtained from chemical analysis using MS-based techniques, including raw spectrum files, are typically then processed by bioinformatics tool such as GNPS. GNPS is a web-based mass spectrometry ecosystem that aims to be an open-access knowledge base for community-wide organization and facilitates the sharing of raw, processed, or annotated fragmentation mass spectrometry data (MS/MS). Thus it aids in identification and discovery throughout the entire life cycle of data; from initial data acquisition/analysis to post publication (Wang et al., 2016). GNPS will analyse the chemical diversity and clusters compounds with similar structural features together, which translate into similar fragmentation patterns and into groups of molecular families (Hamed et al., 2020). The resulting visualisation, followed by specific annotations can then be performed by metabolome mining using Cytoscape (https://cytoscape.org/), an open source software platform for visualizing complex networks and integrating these with any type of attribute data; allowing the final version of the molecular network which will allow complete visualization of the chemical diversity of the metabolome. An example of such as molecular is in illustrated in Figure 13 in Chapter 2, which consists of nodes which correspond to parent ions and are linked into groups with edges which represent a cosine similarity score and provides effective and rapid dereplication of large and complex MS/MS datasets (Hamed et al., 2020). Molecular networks have and continue to be successfully implemented in natural product research to accelerate targeted isolation of potentially new metabolites (Hamed et al., 2020. Ding et al., 2018, Yang et al., 2013, Li et al., 2018). The molecular network concept, will be further developed and detailed later in this thesis,

specifically when employed in the study of the metabolome and chemical diversity of the *Streptomyces* B226SN104, SM3 and SM9 isolates.

1.3.2.6 (6) Bioactivity mappings

A second method for molecular networking, called Feature-based Molecular Networking (FBMN) is also available in the GNPS online platform, where a feature detection and alignment tool (MZmine, OpenMS, XCMS, etc.) is used to pre-process the LC-MS/MS data (Buedenbender et al,. 2020). This pre-processing involves filtering data for noise, duplication of peaks and isomers which all together allow relative quantification within the FBMN and furthermore makes it possible to integrate biological activity data into FBMN in a complementary workflow termed "bioactive molecular networking" (BMN) (Buedenbender et al,. 2020). BMN correlate the features of chromatographic profiles of a fractionated extract derived through for example MZmine, (http://mzmine.github.io/) with the bioactivity levels of the respective sample (extracts) leading to the calculation of a bioactivity score for each molecule in the network. The bioactivity score is based on the statistical Pearson correlation coefficients between chemical features and observed bioactivity and can be mapped out on the molecular network to predict potential bioactive compounds or chemical families and subsequently guide their targeted isolation (Buedenbender et al,. 2020). Bioactivity mapping will also be further detailed in the thesis, where the BMN concept was employed in an attempt to map the antimicrobial components in the *Streptomyces* SM9 metabolome.

It is important to note that bioinformatics analysis relies not only on the quality and accuracy of the chemical analysis and results, including the spectrometry data; but also on the metabolome annotation component of the metabolome mining, as an accurate annotation is necessary for data interpretation and for visualisation. Indeed, the computational workflows

for metabolomic interpretation, such as, high-throughput metabolite profiling and annotation, are highly challenging tasks, with fast evolving metabolomics datasets (Ambrosino et al., 2019; Rawlings et al., 2017; Chaleckis, et al., 2019).

The importance of the use of chemical analysis (LC-MS/NMR) coupled with bioinformatics involving genomics and metabolomics in marine drug discovery is reflected in the increased number of scientific papers being published in the area. Amongst these numerous papers, characterization of a 7-prenylisatin antibiotic from *Streptomyces* Species MBT28, which is effective against *Bacillus subtilis* is typical (Wu et al., 2015). The used of NMR-based metabolomics together with a proteomics based approach facilitated the identification of a gene cluster which was subsequently shown to encode an indole prenyltransferase enzyme that catalyzed the conversion of tryptophan into 7-prenylisatin (Chandra Mohana *et al.2018*; Wu et al., 2015). Another example, from the same research group is the reported identification of novel C-glycosylpyranonaphthoquinones in *Streptomyces* sp. MBT76, utilising a combined NMR based metabolomics approach coupled to genome mining (Chandra Mohana *et al.2018;* Wu et al., 2017). This led to the discovery that these C-glycosylpyranonaphthoquinones were produced as a result of the activation of the *qin* gene cluster involving a type II PKS pathway and that these molecules were effective against Gram-positive bacteria, which ultimately provided valuable insights into the pyranonaphthoquinone group of antibiotics (Chandra Mohana *et al.*2018*;* Wu *et al*., 2017). Finally another example is the report by Cheng and co workers who isolated 64 actinomycetes from 12 different marine sponge species and subjected them to metabolomic analysis using high-resolution LC-MS and NMR for dereplication purposes (Cheng et al., 2015). This LC-MS/NMR based detection allowed the prioritization of two isolates which exhibiting distinct chemical profiles belonging to the genera *Streptomyces* (SBT348) and *Micromonospora* (SBT687); as well as, both possessing potent anti-trypanosomal activities (Cheng et al., 2015; Chandra Mohana *et al.*2018).

In this study, we aimed to contribute to the global fight against AMR by employing a large scale based OSMAC approach, involving changes in both temperature and nutrient source parameters, together with new marine "omics" based methods, to study and characterize new marine *Streptomyces* isolates from sponges. In addition, specific bioactivity towards clinical pathogens that have been prioritised by the WHO list (Table 1) has also been targeted for this study, as well as, other activities such as anti-cancer, anti-inflammatory and antioxidant activities which are also important and of biomedical interest. This study employed a multi-fields perspectives, to investigate on the production of secondary metabolites, of potential new metabolites and of potential new activity (Figure 9). We investigated and described potential novel biosynthetic pathways that may encode for any potential new bioactivities, using molecular networks and bioactivity mappings. With this study, we hope to increase the efficiency of future novel bioactive discovery studies.

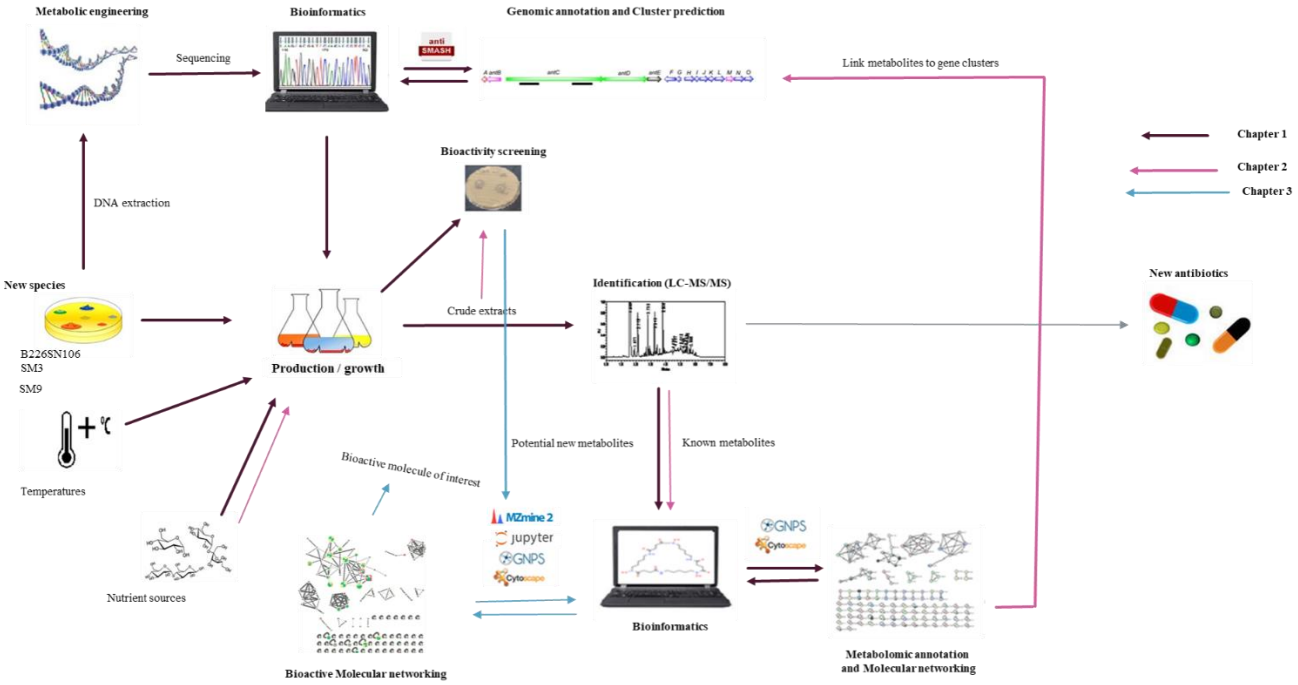


Figure 9: Design of OSMAC based approach developed in this study. (Adapted from Antoraz et al., 2015).

The One Strain MAny Compounds approach using genomics analysis, nutrient sources and temperatures changes in parameters of growth conditions. Bioactivity screening and metabolomic analysis additional steps has been added to the original OSMAC approach to enhance further the production and the selection of potential bioactive novel metabolites though the different chapters of this thesis.

References

Editorial: ChemSpider--a tool for Natural Products research. (2015). *Nat Prod Rep, 32*(8), 1163-1164. doi:10.1039/c5np90022k

Akbar, M. A., Mohd Yusof, N. Y., Tahir, N. I., Ahmad, A., Usup, G., Sahrani, F. K., & Bunawan, H. (2020). Biosynthesis of Saxitoxin in Marine Dinoflagellates: An Omics Perspective. *Marine Drugs, 18*(2). doi:10.3390/md18020103

Ambrosino, L., Tangherlini, M., Colantuono, C., Esposito, A., Sangiovanni, M., Miralto, M., Sansone, C., Chiusano, M. L. (2019). Bioinformatics for Marine Products: An Overview of Resources, Bottlenecks, and Perspectives. *Marine Drugs, 17*(10). doi:10.3390/md17100576

Antoraz, S., Santamaria, R. I., Diaz, M., Sanz, D., & Rodriguez, H. (2015). Toward a new focus in antibiotic and drug discovery from the *Streptomyces* arsenal. *Front Microbiol, 6*, 461. doi:10.3389/fmicb.2015.00461

AR Threats Report, 2019. Biggest threat and data. 2019 AR Threat Report. Center for Disease Control and Prevention CDC's Antibiotic Resistance Threats in the United States, 2019. https://www.cdc.gov/drugresistance/biggest-threats.html

Asokan, G. V., Ramadhan, T., Ahmed, E., & Sanad, H. (2019). WHO Global Priority Pathogens List: A Bibliometric Analysis of Medline-PubMed for Knowledge Mobilization to Infection Prevention and Control Practices in Bahrain. *Oman Med J,*

34(3), 184-193. doi:10.5001/omj.2019.37

Austin, M. B., & Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep, 20*(1), 79-110. doi:10.1039/b100917f

Aya E. S., Abd-Ellatif ., Ahmed, S., Abdel-Razek., Abdelaaty Hamed., Maha, M., Soltan Hesham S. M. Soliman Mohamed Shaaban. Bioactive compounds from marine Streptomyces sp.: Structure identification and biological activities. First published: 25 October 2019 https://doi.org/10.1002/vjch.201900108.

Bansal, A. K. (2005). Bioinformatics in microbial biotechnology--a mini review. *Microb Cell Fact, 4*, 19. doi:10.1186/1475-2859-4-19

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. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417*, 141–147.

Birkett, D., Brosen, K., Cascorbi, I., Gustafsson, L. L., Maxwell, S., Rago, L.,. Sjoqvist, F. (2010). Clinical pharmacology in research, teaching and health care: Considerations by IUPHAR, the International Union of Basic and Clinical Pharmacology. *Basic Clin Pharmacol Toxicol, 107*(1), 531-559. doi:10.1111/j.1742-7843.2010.00602.x

Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. (2015). Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol, 13*(1), 42-51.

doi:10.1038/nrmicro3380

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res, 47*(W1), W81-W87. doi:10.1093/nar/gkz310

Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H., & Prinsep, M. R. (2014). Marine natural products. *Nat Prod Rep, 31*(2), 160-258. doi:10.1039/c3np70117d

Buedenbender, L., Astone, F. A., & Tasdemir, D. (2020). Bioactive Molecular Networking for Mapping the Antimicrobial Constituents of the Baltic Brown Alga Fucus vesiculosus. *Marine Drugs, 18*(6). doi:10.3390/md18060311

Bundy, J. G., Willey, T. L., Castell, R. S., Ellar, D. J., & Brindle, K. M. (2005). Discrimination of pathogenic clinical isolates and laboratory strains of Bacillus cereus by NMR-based metabolomic profiling. *FEMS Microbiol Lett, 242*(1), 127-136. doi:10.1016/j.femsle.2004.10.048

Cai, S. S., Short, L. C., Syage, J. A., Potvin, M., & Curtis, J. M. (2007). Liquid chromatography-atmospheric pressure photoionization-mass spectrometry analysis of triacylglycerol lipids--effects of mobile phases on sensitivity. *J Chromatogr A, 1173*(1-2), 88-97. doi:10.1016/j.chroma.2007.10.008

Chaleckis, R., Meister, I., Zhang, P., & Wheelock, C. E. (2019). Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. *Curr Opin*

Biotechnol, 55, 44-50. doi:10.1016/j.copbio.2018.07.010

Chandra Mohana, N., Yashavantha Rao, H. C., Rakshith, D., Mithun, P. R., Nuthan, B. R., Satish, S. (2018). Omics based approach for biodiscovery of microbial natural products in antibiotic resistance era. *J Genet Eng Biotechnol, 16*(1), 1-8. doi:10.1016/j.jgeb.2018.01.006

Chaudhary, J., Khatri, P., Singla, P., Kumawat, S., Kumari, A., R, V., Vikram, A., Jindal, S. K., Kardile, H., Kumar, R., Sonah, H., Deshmukh, R. (2019). Advances in Omics Approaches for Abiotic Stress Tolerance in Tomato. *Biology (Basel), 8*(4). doi:10.3390/biology8040090

Cheng, C., MacIntyre, L., Abdelmohsen, U. R., Horn, H., Polymenakou, P. N., Edrada-Ebel, R., & Hentschel, U. (2015). Biodiversity, Anti-Trypanosomal Activity Screening, and Metabolomic Profiling of Actinomycetes Isolated from Mediterranean Sponges. *PLoS One, 10*(9), e0138528. doi:10.1371/journal.pone.0138528

Cheng, M. M., Tang, X. L., Sun, Y. T., Song, D. Y., Cheng, Y. J., Liu, H., . . . Li, G. Q. (2020). Biological and Chemical Diversity of Marine Sponge-Derived Microorganisms over the Last Two Decades from 1998 to 2017. *Molecules, 25*(4). doi:10.3390/molecules25040853

Clark, C. A., & Lawrence, A. (1981). Morphology of spore-bearing structures in Streptomyces ipomoea. *Can J Microbiol, 27*(6), 575-579. doi:10.1139/m81-087

Cox, D. G., Oh, J., Keasling, A., Colson, K. L., & Hamann, M. T. (2014). The utility of metabolomics in natural product and biomarker characterization. *Biochim Biophys Acta, 1840*(12), 3460-3474. doi:10.1016/j.bbagen.2014.08.007

Crossman, L., Hutchings, M., Bibb, M., Kennedy, J., Dobson, A. (2013). Mining Biodiversity within Marine sponges for Bioactive compounds). Available at Norwich Research Park Science (nrp.ac.uk).

Dadgostar, P. (2019). Antimicrobial Resistance: Implications and Costs. *Infect Drug Resist, 12*, 3903-3910. doi:10.2147/IDR.S234610

Das, S., Ward, L.R., Burke, C. Screening of marine Streptomyces spp. for potential use as probiotics in aquaculture. Aquaculture Volume 305, Issues 1–4, 1 July 2010, Pages 32-41. https://doi.org/10.1016/j.aquaculture.2010.04.001.

Ding, C. Y. G., Pang, L. M., Liang, Z. X., Goh, K. K. K., Glukhov, E., Gerwick, W. H., & Tan, L. T. (2018). MS/MS-Based Molecular Networking Approach for the Detection of Aplysiatoxin-Related Compounds in Environmental Marine Cyanobacteria. *Marine Drugs, 16*(12). doi:10.3390/md16120505

Donadio, S., Monciardini, P., & Sosio, M. (2007). Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep, 24*(5), 1073-1109. doi:10.1039/b514050c

Doroghazi, J. R., & Metcalf, W. W. (2013). Comparative genomics of actinomycetes with a

focus on natural product biosynthetic genes. *BMC Genomics, 14*, 611. doi:10.1186/1471-2164-14-611

Drlica, K., & Zhao, X. (1997). DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol Mol Biol Rev, 61*(3), 377-392.

Dunn, W.B., Ellis, D.I. (2005). Metabolomics: current analytical platforms and methodologies. TrAC-Trend Anal. Chem. 2005;24:285–294.

Fadela, 2019. New report calls for urgent action to avert antimicrobial resistance crisis: International organizations unite on critical recommendations to combat drug-resistant infections and prevent staggering number of deaths each year. https://www.who.int/news/item/29-04-2019-new-report-calls-for-urgent-action-to-avert-antimicrobial-resistance-crisis#:~:text=If%20no%20action%20is%20taken,2008%2D2009%20global%20financial%20crisis.

Faulkner, D. J. (2002). Marine natural products. *Nat Prod Rep, 19*(1), 1-48. doi:10.1039/b009029h

Finking, R., & Marahiel, M. A. (2004). Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol, 58*, 453-488. doi:10.1146/annurev.micro.58.030603.123615

Firsova, D., Mahajan, N., Solanki, H., Morrow, C. and Thomas, O.P. (2017). Current Status and Perspectives in Marine Biodiscovery. In Bioprospecting (pp. 29-50). Springer,

Cham, Switzerland.

Floros, D. J., Jensen, P. R., Dorrestein, P. C., & Koyama, N. (2016). A metabolomics guided exploration of marine natural product chemical space. *Metabolomics, 12*(9). doi:10.1007/s11306-016-1087-5

Fontana, J. M., Alexander, E., & Salvatore, M. (2012). Translational research in infectious disease: current paradigms and challenges ahead. *Transl Res, 159*(6), 430-453. doi:10.1016/j.trsl.2011.12.009

Franco-Correa, M., Quintana, A., Duque, C., Suarez, C., Rodríguez M.X., Barea J.M. Evaluation of actinomycete strains for key traits related with plant growth promotion and mycorrhiza helping activities. *Appl Soil Ecol.* 2010;45(3):209–217. 10.1016/j.apsoil.2010.04.007

Frota, M. J., Silva, R. B., Mothes, B., Henriques, A. T., & Moreira, J. C. (2012). Current status on natural products with antitumor activity from Brazilian marine sponges. *Curr Pharm Biotechnol, 13*(1), 235-244. doi:10.2174/138920112798868674

Fuhrer, T., Heer, D., Begemann, B., & Zamboni, N. (2011). High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal Chem, 83*(18), 7074-7080. doi:10.1021/ac201267k

Giles, E. C., Kamke, J., Moitinho-Silva, L., Taylor, M. W., Hentschel, U., Ravasi, T., & Schmitt, S. (2013). Bacterial community profiles in low microbial abundance

sponges. *FEMS microbiology ecology*, *83*(1), 232–241. https://doi.org/10.1111/j.1574-6941.2012.01467.x

Gribble, G. W. (2015). Biological Activity of Recently Discovered Halogenated Marine Natural Products. *Marine Drugs, 13*(7), 4044-4136. doi:10.3390/md13074044

Hamed, A. A., Soldatou, S., Qader, M. M., Arjunan, S., Miranda, K. J., Casolari, F., Ebel, R. (2020). Screening Fungal Endophytes Derived from Under-Explored Egyptian Marine Habitats for Antimicrobial and Antioxidant Properties in Factionalised Textiles. *Microorganisms, 8*(10). doi:10.3390/microorganisms8101617

Harrison, J., & Studholme, D. J. (2014). Recently published Streptomyces genome sequences. *Microb Biotechnol, 7*(5), 373-380. doi:10.1111/1751-7915.12143

Heinsch, S. C., Hsu, S. Y., Otto-Hanson, L., Kinkel, L., & Smanski, M. J. (2019). Complete genome sequences of Streptomyces spp. isolated from disease-suppressive soils. *BMC Genomics*, *20*(1), 994. doi:10.1186/s12864-019-6279-8

Hertiani, T., Edrada-Ebel, R., Ortlepp, S., van Soest, R. W., de Voogd, N. J., Wray, V., Hentschel, U., Kozytska, S., Muller, W. E., Proksch, P. (2010). From anti-fouling to biofilm inhibition: new cytotoxic secondary metabolites from two Indonesian Agelas sponges. *Bioorg Med Chem, 18*(3), 1297-1311. doi:10.1016/j.bmc.2009.12.028

Hopwood, D. A. (2006). Soil to genomics: the Streptomyces chromosome. *Annu Rev Genet, 40*, 1-23. doi:10.1146/annurev.genet.40.110405.090639

Hughes, C. C., & Fenical, W. (2010). Antibacterials from the sea. *Chemistry, 16*(42), 12512-12525. doi:10.1002/chem.201001279

Hwang, S., Lee, N., Jeong, Y., Lee, Y., Kim, W., Cho, S., Palsson, B. O., Cho, B. K. (2019). Primary transcriptome and translatome analysis determines transcriptional and translational regulatory elements encoded in the Streptomyces clavuligerus genome. *Nucleic Acids Res, 47*(12), 6114-6129. doi:10.1093/nar/gkz471

IACG.2019- No time to wait: securing the future from drug-resistant infections. Interagency Coordination Group on Antimicrobial Resistance. https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG_final_report_EN.pdf?ua=1

Jackson, S. A., Crossman, L., Almeida, E. L., Margassery, L. M., Kennedy, J., & Dobson, A. D. W. (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene Clusters in the Genomes of Marine Sponge Derived Streptomyces spp. Isolates. *Marine Drugs, 16*(2). doi:10.3390/md16020067

Jacoby, G. A. (2005). Mechanisms of resistance to quinolones. *Clin Infect Dis, 41 Suppl 2*, S120-126. doi:10.1086/428052

Jaspars, M., De Pascale, D., Andersen, J.H., Reyes, F., Crawford, A.D. and Ianora, A. (2016).The marine biodiscovery pipeline and ocean medicines of tomorrow. Journal of the Marine Biological Association of the United Kingdom, 96(1), pp.151-158.

Jaspars, M., De Pascale, D., Andersen, J.H., Reyes, F., Crawford, A.D. and Ianora, A. (2016). The marine biodiscovery pipeline and ocean medicines of tomorrow. Journal of the Marine Biological Association of the United Kingdom, 96(1), pp.151-158.

Jensen, P.R., Fenical, W. (1996). Marine bacterial diversity as a resource for novel microbial products. J. Ind. Microbiol. Biotechnol. 1996, 17, 346–351.

Jones, S. E., & Elliot, M. A. (2017). Streptomyces Exploration: Competition, Volatile Communication and New Bacterial Behaviours. *Trends Microbiol, 25*(7), 522-531. doi:10.1016/j.tim.2017.02.001

Kamalakkannan, P. (2015). Marine sponges a good source of bioactive compounds in anticancer agents. International Journal of Pharmaceutical Sciences Review and Research 31(2):132-135.

Kaddurah-Daouk, R., Kristal, B. S., & Weinshilboum, R. M. (2008). Metabolomics: a global biochemical approach to drug response and disease. *Annu Rev Pharmacol Toxicol, 48*, 653-683. doi:10.1146/annurev.pharmtox.48.113006.094715

Kennedy, J., Codling, C.E., Jones, B.V., Dobson, A.D.W. and Marchesi, J.R. (2008), Diversity of microbes associated with the marine sponge, *Haliclona simulans*, isolated from Irish waters and identification of polyketide synthase genes from the sponge metagenome. Environmental Microbiology, 10: 1888-1902. https://doi.org/10.1111/j.1462-2920.2008.01614.x

Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., Bourke, M. B., Rea, M. C., O'Connor, P. M., Ross, R. P., Hill, C., O'Gara, F., Marchesi, J. R., & Dobson, A. D. (2009b). Isolation and analysis of bacteria with antimicrobial activities from the marine sponge Haliclona simulans collected from Irish waters. *Marine biotechnology (New York, N.Y.), 11*(3), 384–396. https://doi.org/10.1007/s10126-008-9154-1

Kennedy, J., Flemer, B., Jackson, S. A., Lejon, D. P., Morrissey, J. P., O'Gara, F., & Dobson, A. D. (2010). Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Marine Drugs, 8*(3), 608-628. doi:10.3390/md8030608

Kennedy, J., Flemer, B., Jackson, S.A., Morrissey, J.P., O'Gara, F., Dobson ADW (2014) Evidence of a Putative Deep Sea Specific Microbiome in Marine Sponges. PLoS ONE 9(3): e91092. doi:10.1371/journal.pone.0091092

Kim, H. K., Choi, Y. H., & Verpoorte, R. (2010). NMR-based metabolomic analysis of plants. *Nat Protoc, 5*(3), 536-549. doi:10.1038/nprot.2009.237

Kim, H. K., Choi, Y. H., & Verpoorte, R. (2011). NMR-based plant metabolomics: where do we stand, where do we go? *Trends Biotechnol, 29*(6), 267-275. doi:10.1016/j.tibtech.2011.02.001

Kim, H. K., Saifullah, Khan, S., Wilson, E. G., Kricun, S. D., Meissner, A., Goraler, S., Deelder, A. M., Choi, Y. H., Verpoorte, R. (2010). Metabolic classification of South American Ilex species by NMR-based metabolomics. *Phytochemistry, 71*(7), 773-784. doi:10.1016/j.phytochem.2010.02.001

Kumar, A., Sorensen, J. L., Hansen, F. T., Arvas, M., Syed, M. F., Hassan, L., Benz, J. P., Record, E., Henrissat, B., Poggeler, S., Kempken, F. (2018). Genome Sequencing and analyses of Two Marine Fungi from the North Sea Unraveled a Plethora of Novel Biosynthetic Gene Clusters. *Sci Rep, 8*(1), 10187. doi:10.1038/s41598-018-28473-z

Laport, M. S., Santos, O. C., & Muricy, G. (2009). Marine sponges: potential sources of new antimicrobial drugs. *Curr Pharm Biotechnol, 10*(1), 86-105. doi:10.2174/138920109787048625

Lee, N., Hwang, S., Kim, J., Cho, S., Palsson, B., & Cho, B. K. (2020). Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Comput Struct Biotechnol J, 18*, 1548-1556. doi:10.1016/j.csbj.2020.06.024

Lee, N., Hwang, S., Lee, Y., Cho, S., Palsson, B., & Cho, B. K. (2019). Synthetic Biology Tools for Novel Secondary Metabolite Discovery in Streptomyces. *J Microbiol Biotechnol, 29*(5), 667-686. doi:10.4014/jmb.1904.04015

Lee, N., Kim, W., Hwang, S., Lee, Y., Cho, S., Palsson, B., & Cho, B. K. (2020). Thirty complete Streptomyces genome sequences for mining novel secondary metabolite biosynthetic gene clusters. *Sci Data, 7*(1), 55. doi:10.1038/s41597-020-0395-9

Li, C. W., Chen, J. Y., & Hua, T. E. (1998). Precambrian sponges with cellular structures. *Science (New York, N.Y.), 279*(5352), 879–882. https://doi.org/10.1126/science.279.5352.879

Li, F., Janussen, D., Peifer, C., Perez-Victoria, I., & Tasdemir, D. (2018). Targeted Isolation of Tsitsikammamines from the Antarctic Deep-Sea Sponge Latrunculia biformis by Molecular Networking and Anticancer Activity. *Marine Drugs, 16*(8). doi:10.3390/md16080268

Li, J., Zhang, L., & Liu, W. (2018). Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. *Synth Syst Biotechnol, 3*(2), 83-89. doi:10.1016/j.synbio.2018.02.002

Lim, Y. P., Go, M. K., & Yew, W. S. (2016). Exploiting the Biosynthetic Potential of Type III Polyketide Synthases. *Molecules, 21*(6). doi:10.3390/molecules21060806

Liu, F., Garneau, S., & Walsh, C. T. (2004). Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: minimal requirements at N and C termini of EpoB for elongation. *Chem Biol, 11*(11), 1533-1542. doi:10.1016/j.chembiol.2004.08.017

Liu, M., Seidel, V., Katerere, D. R., & Gray, A. I. (2007). Colorimetric broth microdilution method for the antifungal screening of plant extracts against yeasts. *Methods, 42*(4), 325-329. doi:10.1016/j.ymeth.2007.02.013

Lu, W., Bennett, B.D., Rabinowitz, J.D. (2008). Analytical strategies for LC–MS-based targeted metabolomics. J. Chromatogr. B. 2008;871:236–242.

Madigan, M., Martinko, J., eds. *(2005).* Brock Biology of Microorganisms (11th ed.).

Prentice Hall. ISBN 978-0-13-144329-7

Mahrous, E. A., & Farag, M. A. (2015). Two dimensional NMR spectroscopic approaches for exploring plant metabolome: A review. *J Adv Res, 6*(1), 3-15. doi:10.1016/j.jare.2014.10.003

Malik, A., Kim, Y. R., Jang, I. H., Hwang, S., Oh, D. C., & Kim, S. B. (2020). Genome-based analysis for the bioactive potential of Streptomyces yeochonensis CN732, an acidophilic filamentous soil actinobacterium. *BMC Genomics, 21*(1), 118. doi:10.1186/s12864-020-6468-5

Marra, M.V., 2019. PhD thesis. Investigation of biological factors that may contribute to bioactivity in *Haliclona* (Porifera, Haplosclerida). Availalble at Investigation of biological factors that may contribute to bioactivity in Haliclona (Porifera, Haplosclerida) (nuigalway.ie).

Mazon-Suastegui, J. M., Salas-Leiva, J. S., Medina-Marrero, R., Medina-Garcia, R., & Garcia-Bernal, M. (2020). Effect of Streptomyces probiotics on gut microbiota of Pacific white shrimp. Global Aquaculture Alliance, Health & Welfare. Available at https://www.aquaculturealliance.org/advocate/effect-of-streptomyces-probiotics-on-gut-microbiota-of-pacific-white-shrimp/

Mazon-Suastegui, J. M., Salas-Leiva, J. S., Medina-Marrero, R., Medina-Garcia, R., & Garcia-Bernal, M. (2020). Effect of Streptomyces probiotics on the gut microbiota of Litopenaeus vannamei challenged with Vibrio parahaemolyticus. *Microbiologyopen,*

9(2), e967. doi:10.1002/mbo3.967

Mehbub, M. F., Lei, J., Franco, C., & Zhang, W. (2014). Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. *Marine Drugs, 12*(8), 4539-4577. doi:10.3390/md12084539

Miller, D. A., Luo, L., Hillson, N., Keating, T. A., & Walsh, C. T. (2002). Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem Biol, 9*(3), 333-344. doi:10.1016/s1074-5521(02)00115-1

Mulvey, M. R., & Simor, A. E. (2009). Antimicrobial resistance in hospitals: how concerned should we be? *CMAJ, 180*(4), 408-415. doi:10.1503/cmaj.080239

Murray, P.R., Baron, E.J., Jorgensen, J.H., Pfaller, M.A. and Yolken, R.H. (2003). Manual of clinical microbiology. 2003. American Society for Microbiology, Washington, DC.

Mushtaq, A. (2016). UN commits to tackling antimicrobial resistance. Lancet Infect Dis. 2016;16(11):1229–1230

Nahrgang, S., Nolte, E., Rechel, B. Antimicrobial resistance. The role of public health organizations in addressing public health problems in Europe: The case of obesity, alcohol and antimicrobial resistance [Internet]. Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2018. (Health Policy Series, No. 51.) 4. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536193/

Naughton, L. M., Romano, S., O'Gara, F., & Dobson, A. D. W. (2017). Identification of Secondary Metabolite Gene Clusters in the Pseudovibrio Genus Reveals Encouraging Biosynthetic Potential toward the Production of Novel Bioactive Compounds. *Front Microbiol, 8*, 1494. doi:10.3389/fmicb.2017.01494

Neu, H. C. (1992). The crisis in antibiotic resistance. *Science, 257*(5073), 1064-1073. doi:10.1126/science.257.5073.1064

Nindita, Y., Cao, Z., Fauzi, A. A., Teshima, A., Misaki, Y., Muslimin, R., Arakawa, K. (2019). The genome sequence of Streptomyces rochei 7434AN4, which carries a linear chromosome and three characteristic linear plasmids. *Sci Rep, 9*(1), 10973. doi:10.1038/s41598-019-47406-y

O'Neill, J. 2014. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. 2014. https://amr-review.org/Publications.html

O'Neill, J. 2015. Securing new drugs for future generations : the pipeline of antibiotics. 2015. https://amr-review.org/Publications.html

O'Neill, J. 2016. Tackling drug-resistant infections globally: final report and recommendations the review on antimicrobial resistance chaired by JIM O'NEILL; 2016. https://amr-review.org/Publications.html

Olanrewaju, O. S., & Babalola, O. O. (2019). Streptomyces: implications and interactions in plant growth promotion. *Appl Microbiol Biotechnol, 103*(3), 1179-1188.

doi:10.1007/s00253-018-09577-y

Paradkar, A., Trefzer, A., Chakraburtty, R., & Stassi, D. (2003). Streptomyces genetics: a genomic perspective. *Crit Rev Biotechnol, 23*(1), 1-27. doi:10.1080/713609296

Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol, 13*(4), 263-269. doi:10.1038/nrm3314

Paulus, C., Rebets, Y., Tokovenko, B., Nadmid, S., Terekhova, L. P., Myronovskyi, M., . . . Luzhetskyy, A. (2017). New natural products identified by combined genomics-metabolomics profiling of marine Streptomyces sp. MP131-18. *Sci Rep, 7*, 42382. doi:10.1038/srep42382

Picton, B.E. & Morrow, C.C. (2016). *Haliclona simulans* (Johnston, 1842). *[In] Encyclopedia of Marine Life of Britain and Ireland.* Available at http://www.habitas.org.uk/marinelife/species.asp?item=C8630.

Piel, J. (2009). Metabolites from symbiotic bacteria. *Nat Prod Rep, 26*(3), 338-362. doi:10.1039/b703499g

Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev, 30*(1), 19-34.

Procopio, R. E., Silva, I. R., Martins, M. K., Azevedo, J. L., & Araujo, J. M. (2012). Antibiotics produced by Streptomyces. *Braz J Infect Dis, 16*(5), 466-471.

doi:10.1016/j.bjid.2012.08.014

Proksch, P. (1994). Defensive roles for secondary metabolites from marine sponges and sponge-feeding nudibranchs. *Toxicon, 32*(6), 639-655. doi:10.1016/0041-0101(94)90334-4

Rajendran, S. (2019). Marine Sponges: Repositories of Bioactive compounds with Medicinal applications. International Journal of ChemTech Research 12(01):26-48. DOI:10.20902/IJCTR.2019.120103. https://www.researchgate.net/publication/330217609_Marine_Sponges_Repositories_of_Bioactive_compounds_with_Medicinal_applications

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res, 46*(D1), D624-D632. doi:10.1093/nar/gkx1134

Reen, F. J., Gutierrez-Barranquero, J. A., Dobson, A. D., Adams, C., & O'Gara, F. (2015). Emerging concepts promising new horizons for marine biodiscovery and synthetic biology. *Marine Drugs, 13*(5), 2924-2954. doi:10.3390/md13052924

Robinette, S. L., Bruschweiler, R., Schroeder, F. C., & Edison, A. S. (2012). NMR in metabolomics and natural products research: two sides of the same coin. *Acc Chem Res, 45*(2), 288-297. doi:10.1021/ar2001606

Romano, S., Jackson, S. A., Patry, S., & Dobson, A. D. W. (2018). Extending the "One Strain Many Compounds" (OSMAC) Principle to Marine Microorganisms. *Marine Drugs, 16*(7). doi:10.3390/md16070244

Sagar, S., Kaur, M., & Minneman, K. P. (2010). Antiviral lead compounds from marine sponges. *Marine Drugs, 8*(10), 2619-2638. doi:10.3390/md8102619

Salem, M. A., Perez de Souza, L., Serag, A., Fernie, A. R., Farag, M. A., Ezzat, S. M., & Alseekh, S. (2020). Metabolomics in the Context of Plant Natural Products Research: From Sample Preparation to Metabolite Analysis. *Metabolites, 10*(1). doi:10.3390/metabo10010037

Santamaría Ramón, I., Díaz, M., Sanz, D. and Rodríguez, H. (2015). Toward a new focus in antibiotic and drug discovery from the Streptomyces arsenal. Front. Microbiol. 4,461.

Santoro-Lopes, G., & de Gouvea, E. F. (2014). Multidrug-resistant bacterial infections after liver transplantation: an ever-growing challenge. *World J Gastroenterol, 20*(20), 6201-6210. doi:10.3748/wjg.v20.i20.6201

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol, 7*(2), 285-295. doi:10.1016/s1367-5931(03)00020-6

Shen, Y., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., Zhao, R., Livesay, E. A., Udseth, H. R., Smith, R. D. (2005). Automated 20 kpsi RPLC-MS and MS/MS

with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. *Anal Chem, 77*(10), 3090-3100. doi:10.1021/ac0483062

Shin, D., Byun, W. S., Moon, K., Kwon, Y., Bae, M., Um, S., Lee, S. K., Oh, D. C. (2018). Coculture of Marine Streptomyces sp. With Bacillus sp. Produces a New Piperazic Acid-Bearing Cyclic Peptide. *Front Chem, 6*, 498. doi:10.3389/fchem.2018.00498

Tan, L. T., Chan, K. G., Lee, L. H., & Goh, B. H. (2016). Streptomyces Bacteria as Potential Probiotics in Aquaculture. *Front Microbiol, 7*, 79. doi:10.3389/fmicb.2016.00079

Tang, X., Li, J., Millan-Aguinaga, N., Zhang, J. J., O'Neill, E. C., Ugalde, J. A., Jensen, P. R., Mantovani, S. M., Moore, B. S. (2015). Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem Biol, 10*(12), 2841-2849. doi:10.1021/acschembio.5b00658

Tawfike, A. F., Viegelmann, C., & Edrada-Ebel, R. (2013). Metabolomics and dereplication strategies in natural products. *Methods Mol Biol, 1055*, 227-244. doi:10.1007/978-1-62703-577-4_17

TB Alliance, 2019. Global pandemic. TB alliance. https://www.tballi ance.org/why-new-tb-drugs/global-pandemic. Accessed September 17, 2019.

Tenover, F. C. (2006). Mechanisms of antimicrobial resistance in bacteria. *Am J Med, 119*(6 Suppl 1), S3-10; discussion S62-70. doi:10.1016/j.amjmed.2006.03.011

The Economist, 2019. Drug-resistant tuberculosis: worth the investment. https://www.eiu.com/graphics/market ing/pdf/Drug-resistant-tuberculosis-Article.pdf. The Economist. Intelligence Unit. Healthcare. Accessed September 17, 2019.

Turk, T., Ambrozic Avgustin, J., Batista, U., Strugar, G., Kosmina, R., Civovic, S., Janussen, D., Kauferstein, S., Mebs, D., Sepcic, K. (2013). Biological activities of ethanolic extracts from deep-sea Antarctic marine sponges. *Marine Drugs, 11*(4), 1126-1139. doi:10.3390/md11041126

UN Ad hoc Interagency Coordinating Group on Antimicrobial Resistance, 2019. World Health Organization. UN Interagency Coordination Group (IACG) on Antimicrobial Resistance. Office of the UN Secretary-General. https://www.who.int/antimicrobial-resistance/interagency-coordination-group/en/

Vacelet, J. (1975) Étude en microscopiie électronique de l'association entre bactéries et spongiaires du genre *Verongia* (Dyctioceratida). J Micros Biol Cell 23:271–288

Villa, F. A., & Gerwick, L. (2010). Marine natural product drug discovery: Leads for treatment of inflammation, cancer, infections, and neurological disorders. *Immunopharmacol Immunotoxicol, 32*(2), 228-237. doi:10.3109/08923970903296136

Waksman, S. A., & Henrici, A. T. (1943). The Nomenclature and Classification of the Actinomycetes. *J Bacteriol, 46*(4), 337-341. doi:10.1128/JB.46.4.337-341.1943

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D.,

Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crusemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderon, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., P, C. A. B., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodriguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. F., Alexandrov, T., Litaudon, M., Wolfender, J. L. Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Muller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B. O., Pogliano, K., Linington, R. G., Gutierrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C., Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol, 34*(8), 828-837. doi:10.1038/nbt.3597

Watve, M. G., Tickoo, R., Jog, M. M., & Bhole, B. D. (2001). How many antibiotics are

produced by the genus Streptomyces? *Arch Microbiol, 176*(5), 386-390. doi:10.1007/s002030100345

Webber, M. A., & Piddock, L. J. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *J Antimicrob Chemother, 51*(1), 9-11. doi:10.1093/jac/dkg050

Wellington, E. M., Boxall, A. B., Cross, P., Feil, E. J., Gaze, W. H., Hawkey, P. M., Johnson-Rollings, A. S., Jones, D. L., Lee, N. M., Otten, W., Thomas, C. M., & Williams, A. P. (2013). The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria. The Lancet. Infectious diseases, 13(2), 155–165. https://doi.org/10.1016/S1473-3099(12)70317-1

Williamson, N.R., Fineran, P.C., Leeper, F.J. and Salmond, G.P. 2006. The biosynthesis and regulation of bacterial prodiginines. Nature Reviews Microbiology, 4(12), p.887.

WHO, 2011. EUR/RC61/14. Copenhagen: WHO Regional Office for Europe; 2011. Regional Action Plan on Antibiotic Resistance. Sixty-first session of the Regional Committee for Europe.

WHO, 2015. Geneva: World Health Organization; 2015a. Global Action Plan on Antimicrobial Resistance.

WHO, 2017. Geneva: World Health Organization; 2017. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf

WHO, 2020. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

WHO antifungal report, 2020. WHO antifungal meeting report. First Meeting of the WHO Expert Group on Identifying Priority Fungal Pathogens. Meeting report. World Health Organization. June 2020. Electronic version ISBN 978-92-4-000635-5. Available at https://www.who.int/publications/i/item/9789240006355

Wu, C., Du, C., Gubbens, J., Choi, Y. H., & van Wezel, G. P. (2015). Metabolomics-Driven Discovery of a Prenylated Isatin Antibiotic Produced by *Streptomyces* Species MBT28. *J Nat Prod, 78*(10), 2355-2363. doi:10.1021/acs.jnatprod.5b00276

Wu, C., Du, C., Ichinose, K., Choi, Y. H., & van Wezel, G. P. (2017). Discovery of C-Glycosyl pyranonaphthoquinones in Streptomyces sp. MBT76 by a Combined NMR-Based Metabolomics and Bioinformatics Workflow. *J Nat Prod, 80*(2), 269-277. doi:10.1021/acs.jnatprod.6b00478

Xiong, Z. Q., Wang, J. F., Hao, Y. Y., & Wang, Y. (2013). Recent advances in the discovery and development of marine microbial natural products. *Marine Drugs, 11*(3), 700-717. doi:10.3390/md11030700

Xu, L., Ye, K. X., Dai, W. H., Sun, C., Xu, L. H., & Han, B. N. (2019). Comparative Genomic Insights into Secondary Metabolism Biosynthetic Gene Cluster Distributions of Marine *Streptomyces*. *Marine Drugs, 17*(9).

doi:10.3390/md17090498

Xu, X. N., Chen, L. Y., Chen, C., Tang, Y. J., Bai, F. W., Su, C., & Zhao, X. Q. (2018). Genome Mining of the Marine Actinomycete Streptomyces sp. DUT11 and Discovery of Tunicamycins as Anti-complement Agents. *Front Microbiol, 9*, 1318. doi:10.3389/fmicb.2018.01318

Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., . . . Dorrestein, P. C. (2013). Molecular networking as a dereplication strategy. *J Nat Prod, 76*(9), 1686-1699. doi:10.1021/np400413s

Zampieri, M., Sekar, K., Zamboni, N., & Sauer, U. (2017). Frontiers of high-throughput metabolomics. *Curr Opin Chem Biol, 36*, 15-23. doi:10.1016/j.cbpa.2016.12.006

Zhou, B., Xiao, J. F., Tuli, L., & Ressom, H. W. (2012). LC-MS-based metabolomics. *Mol Biosyst, 8*(2), 470-481. doi:10.1039/c1mb05350g

2 Chapter 2: Utilization of OSMAC, Genomic mining, LC-MS and Metabolomic analysis to detect and identify potential new bioactive metabolites in marine *Streptomyces* strains, from deep-sea sponge isolate B226SN104 and shallow water sponge isolate SM3.

Abstract

Microbes associated with marine sponges are exposed to highly competitive environments both physiologically and nutritionally which is likely to promote the production of novel secondary metabolites. This study focuses on the OSMAC (One Strain MAny Metabolites) approach from a number of bacterial strains that were originally isolated from marine sponges, including one from a deep sea sponge, for the production and identification of potential new bioactive metabolites. Six *Streptomyces* strains were functionally characterized by screening for anti-bacterial and anti-fungal activities and were genetically characterized. Full genome sequence were generated for each strain using Illumina technology and approximately 159 potential secondary metabolism biosynthetic gene clusters in total were identified using the AntiSMASH program and genome mining approaches. In addition, two of these strains, B226SN104 and SM3, which displayed the highest number of gene clusters of interest, were subjected to different growth conditions using the OSMAC approach to determine the optimal nutrient media and the optimal temperature for the production of potential novel secondary metabolites. Cell-free supernatants were chemically extracted using Liquid Liquid Extraction (LLE) with

ethyl-acetate. Crude extracts containing the potential secondary metabolites from both aqueous and organic phases were analyzed using LC-MS, molecular networking and manual de-replication for the chemical prediction of the molecules involved. This resulted in the identification a high number of interesting metabolites and a potential new metabolite, produced in specific fermentation conditions which highlight the good candidacy of these strains for the study of potential "silent" gene clusters. To our knowledge this is the first OSMAC study tested on a large scale nutrient sources (seven media per strain tested) while simultaneously comparing the two most used optimum temperature for marine bacteria (i.e. 28ºC and 30ºC). For ours strains, OSMAC revealed that the best condition of fermentation -to produce potential new metabolites from these marine bacterial strains was 28ºC as the optimum temperature, and mannitol with peptone as the optimum nutrient sources.

Keywords: OSMAC, secondary metabolites, potential new metabolite, marine bacterial strain, *Streptomyces*, biosynthetic gene clusters, chemical diversity.

Introduction

The One Strain MAny Metabolites (OSMAC) principal involves growing a microbial strain under a variety of different environmental conditions, which typically results in the production of many different types of metabolites; ultimately leading to the potential discovery of new bioactive metabolites such as for example antibiotics (Figure 10) (*Antoraz et al. 2015; Romano et al., 2018; Pan et al., 2019)*. It is highly efficient in inducing chemical diversity through the provision of variations in cultivation parameters (Fan *et al*., 2019). Factors such as nutrient content, growth temperature, pH and culture aeration, are all known to alter the physiology of microbial strains and thereby alter their secondary metabolism profiles (*Ying et al., 2018*). Parameters such as changes in nutrient sources, involving the modification of both carbon and nitrogen sources in the culture media are known to be particularly useful in altering the metabolite production profile in bacterial strains (*Sproule et al., 2019*). Other parameters include signals that can activate regulatory systems to trigger antibiotic production (*Antoraz et al. 2015; Pan et al., 2019*), while other parameters, including the use of antibiotics and hormones which can also boost the production of secondary metabolites (*Antoraz et al. 2015*). Finally, the use of co-culture approaches with the strain of interest being cultured with another fungal or bacterial strain to stimulate interspecies interactions can also result in the induction of the secondary metabolites (Fdhila *et al*., 2003; Wang *et al*., 2014; *Antoraz et al. 2015*). Changes in these parameters tends to promote, induce or stimulate the production of bioactive secondary metabolites and potentially new metabolites (Reen *et al*., 2015; Pan *et al*., 2019). Following use of the OSMAC strategy, the culture media is typically extracted generating crude extracts from the organic and-or aqueous phases. Bioassay guided approaches are then employed on these crude extracts for the chemical detection of metabolites that are produced, and together with dereplication strategies, result in the identification of these metabolite with the overall aim of determining whether these are potential new metabolites or derivatives of

existing metabolites (Weller *et al.*, 2012). For a unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques, see Weller *et al*., 2012.

In marine environments, bacteria are exposed to quite unique environmental conditions such as temperature, pressure, dissolved oxygen and low nutrient availability, which often results in the production of structurally diverse and biologically important metabolites (Petersen *et al*., 2020). Marine *Streptomyces* species have in particular being shown to produce a wide variety of different bioactive metabolites, which are rich in biological activity. For example marine *Streptomyces* derived metabolites have been reported to exhibit anticancer, antifungal, cytotoxic and antibacterial activity (Niu *et al*., 2011; Zotchev, 2012; Song *et al.*, 2014; Jackson *et al*., 2018).

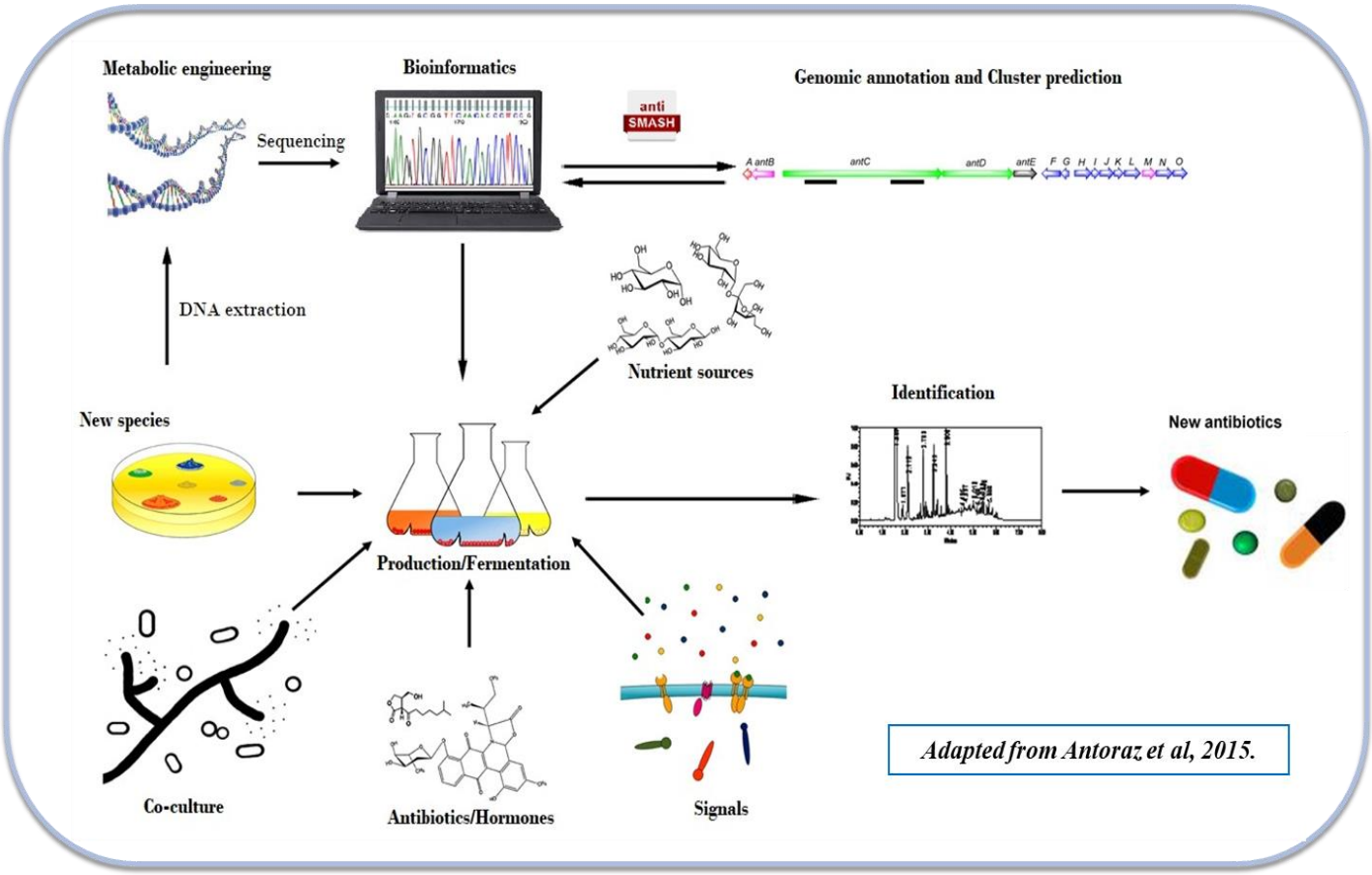


Figure 10: The OSMAC approach. Figure adapted from (Antoraz et al., 2015).

Nutrient sources are known to regulate secondary metabolism in *Streptomyces* strains (Martin *et al*., 2011) with for example antibiotic SBR-22 production in the marine sediment isolate *Streptomyces psammoticus* BT-408 being reported to be regulated by different nitrogen sources (Sujatha et al., 2005) . In addition Surugamide A production in the marine sponge isolate *Streptomyces* sp. SM17 is also known to be regulated by different carbon sources (Saha *et al*., 2005; Almeida *et al*., 2019). Thus in this study we employed the OSMAC approach to investigate the effect of seven different culture conditions with differing carbon and nitrogen nutrient sources, on secondary metabolite production in two marine *Streptomyces* strains; namely *Streptomyces* sp. SM3 isolated from the sponge *Haliclona simulans* (Kennedy *et al.*, 2009) and *Streptomyces* sp. B226SN104 isolated from the deep sea sponge *Stelletta normani* (Borchert *et al.*, 2016).

In addition while a number of studies have reported the effect of temperature on secondary metabolite production in marine bacteria, such as *Pseudoalteromonas*, *Pseudomonas*, *Serratia* and *Vibrio* species; with the optimal temperature being either 28°C or 30°C, within the temperature range from 25°C to 40°C (Wietz *et al., 2011; Darabpour et al. 2012*); there is little information on the effect of temperature on secondary metabolite production in marine *Streptomyces* species. Thus given that there is some evidence that secondary metabolite production may be affected by temperature in terrestrial *Streptomyces* strains, with Jadomycin B production in *Streptomyces venezuelae* ISP5230 being reported to increases following a 42°C heat shock (Doull *et al.*, 1994), and that temperature may regulate validamycin A production in *Streptomyces hydroscopicus*; we decided to also investigate the potential effect of temperature on secondary metabolite production when both SM3 and B226SN104 were grown on the aforementioned different media at two commonly used temperature conditions for the production of secondary metabolites in marine bacterial strains namely 28°C and 30°C.

Results and Discussion

2.1 Genomics and microbiological activity of the *Streptomyces spp* isolates

2.1.1 Secondary metabolite gene clusters of the *Streptomyces spp* isolates.

The isolation and characterization of 18 marine sponge derived *Streptomyces* strains, which displayed a range of antimicrobial and antifungal activities against a number of clinical pathogens has previously been reported (Kennedy *et al.*, 2009; Jackson *et al.*, 2018). The genomes of five of these strains namely *Streptomyces* sp SM3, SM6, SM9 and B226SN104 which displayed the best range of bioactivities were subsequently sequenced and together with genome sequence data available for another marine sponge derived *Streptomyces* strain namely SM14 (Almeida *et al.*, 2019[a]), the six genomes were analysed for the presence of potential secondary metabolite gene clusters (Figure 11).

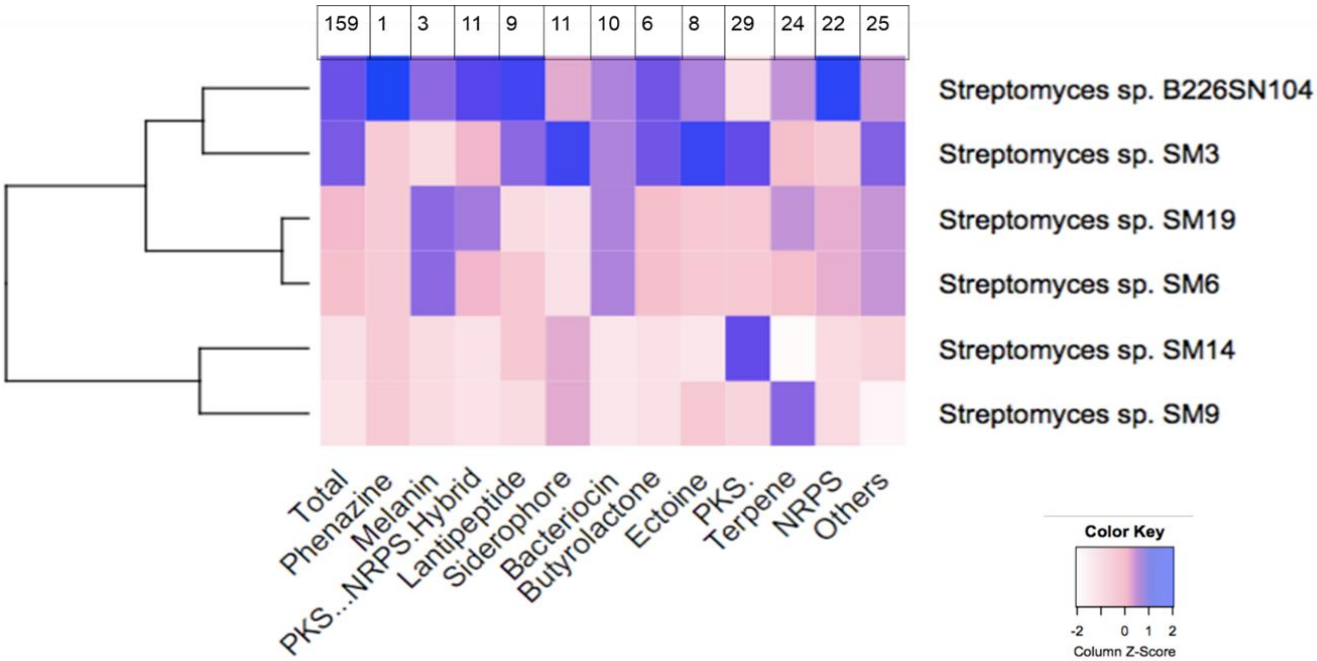


Figure 11: Heat map of the predicted secondary metabolite gene clusters in the genomes of *Streptomyces* isolates using AntiSMASH and heatmap function in R.

Values were automatically converted from the normal distribution of the number of gene clusters to the standard normal distribution by "standardizing" with the use of z-scores. The z-score – (indicated by colour scale) is positive if the value lies above the mean, and negative if it lies below the mean. The

dendrogram was generated automatically according to the default parameters ("predicted clusters") which shows two distinct clusters, representing the genomes which contain the lowest number of gene clusters in one group cluster, and those which contain the higher number of gene clusters in the other group cluster defined by two sub-group clusters. Normal distribution values of the total number of gene clusters identified among the six genomes isolates were manually added on the top of the heat map.

Among these six *Streptomyces* genomes, a total of 159 secondary metabolite gene clusters were predicted to be present, including a total of 29 PKS, 24 Terpene, 22 NRPS, 11 Siderophore and 11 PKS-NRPS hybrids. Following this analysis, two isolates namely B226SN104 and SM3 appeared to have the highest number of predicted gene clusters in their genomes. Thus having identified these two strains as possessing the most promising genomes for the potential production of secondary metabolites, we selected them for analysis using the OSMAC approach, to monitor production of secondary metabolites following growth in different culture conditions including different parameters such as nutrient sources and temperatures. The culture conditions were chosen taking into account well established media for the culture of marine Actinobacteria strains (*Goodfellow et al., 2010*) general media for culturing *Streptomyces* strains (Romano *et al*., 2018; Almeida *et al*., 2019b) together with media known to induce sporulation and antibiotics production in *Streptomyces* (*Tian et al. 2013*) (Table 2).

Media	Composition
M19	Mannitol 2%, Peptone 2% (Goodfellow *et al*., 2010).
M400	Glucose 1%, Meat extract 0.3%, Peptone 0.3%, Starch 2%, Yeast extract 0.5%, CaCO3 0.3% (Goodfellow *et al*., 2010).
MMM	Glucose 1%, Tryptone 0.5%, Starch 0.5%, CaCO3 0.1% (Goodfellow *et al*., 2010).
OM	Oatmeal 2% (Goodfellow *et al*., 2010; Almeida *et al*., 2019b).

SGG	Glucose 1%, Glycerol 1%, Corn steep solid 0.25%, Peptone 0.5%, Starch 1%, Yeast extract 0.2%, CaCO3 0.3%, NaCl 0.1%. (Goodfellow *et al*., 2010).
SM	Soymeal 2%, Mannitol 2%, pH=7 (Tian *et al. 2013*; Almeida *et al*., 2019b]
SYP	Starch 1%, Yeast extract 0.4%, Peptone 0.2% (Romano *et al*., 2018; Almeida *et al*., 2019b)
M410	Glucose 1%, Glycerol 1%, Casaminoacid 1.5%, Oatmeal 0.5%, Peptone 1%, Yeast extract 0.5%, CaCO3 0.1% (Goodfellow *et al*., 2010).

Table 2: Composition of the different media used for the fermentation of the *Streptomyces* sp SM3 and *Streptomyces* sp B226SN104 strains.

2.1.2 Biological activities of B226SN104 and SM3 isolates

To confirm that both isolates were producing bioactivity when cultured under these particular growth conditions, the isolates were screened for antimicrobial activities against relevant clinical pathogens species when grown in M410 media. When grown on this medium the deep sea isolate B226SN104 displayed bioactivity against *C. glabrata* (CBS138) while the shallow water isolate SM3 displayed bioactivity against *P. aeruginosa* (PAO1) and *B. subtilis* (IE32), in well diffusion assays (Table 3).

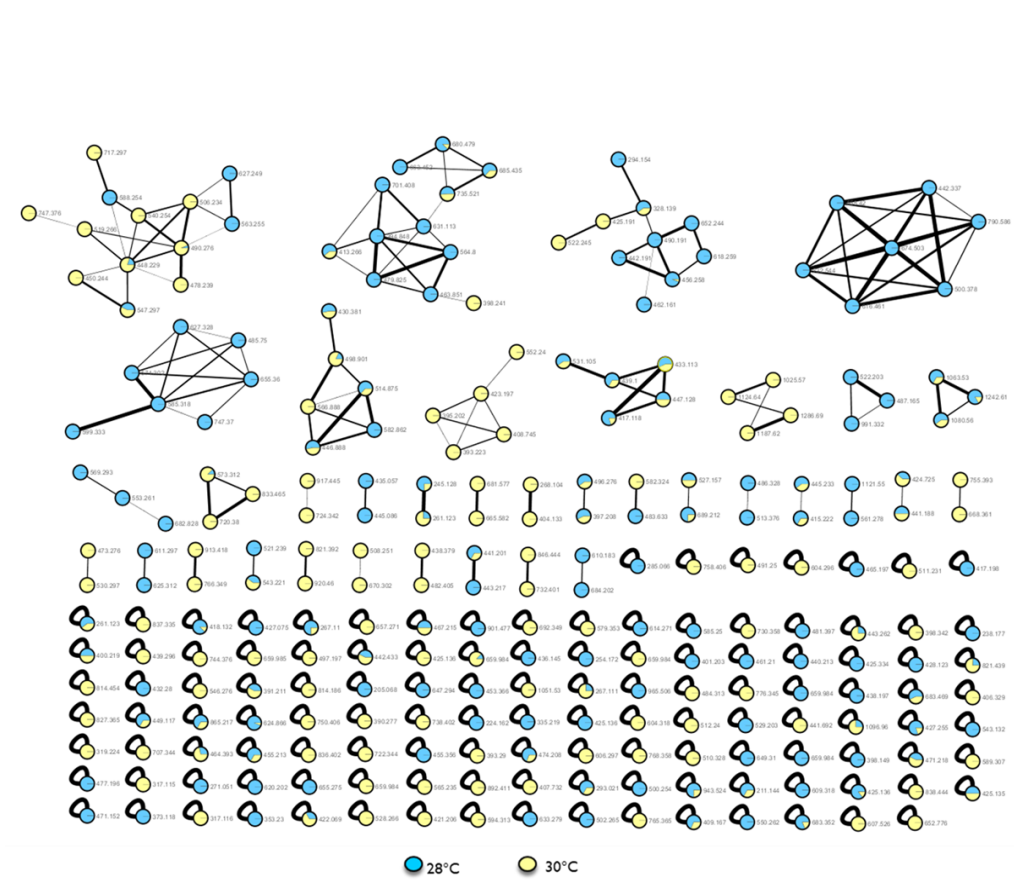
Isolate ID	Bioactivities	Against Relevant pathogenic species
B226SN104	Anti-fungal activity	*C. glabrata (CBS138)*
SM3	Anti-bacterial activity	*P. aeruginosa (PAO1)* *B. subtilis(IE32)*

Table 3: bioactivity table of the deep sea isolate B226SN104 and the shallow water isolate SM3 both tested by well diffusion assay against the relevant pathogenic species: *Staphylococcus aureus* (NCIMB 9518), *Pseudomonas aeruginosa* (PAO1), *Bacillus subtilis* (IE32), *Escherichia coli* (NCIMB 12210), *Candida glabrata* (CBS138), and *Candida albicans* (Sc5314).

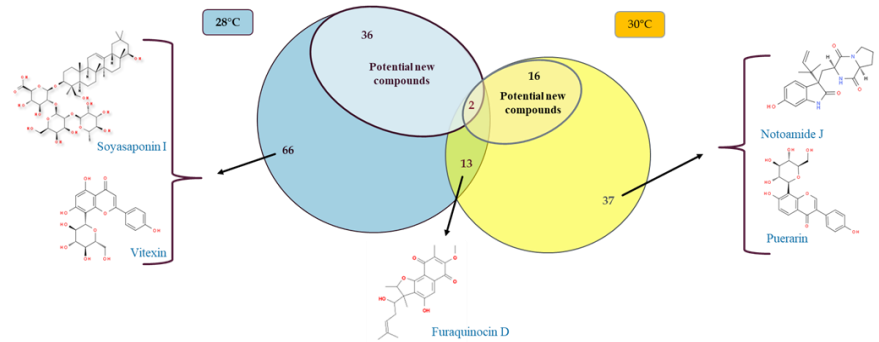
2.2 Metabolomics and chemical analysis

2.2.1 Temperatures involvement in the production of secondary metabolites from both B226SN104 and SM3 isolates.

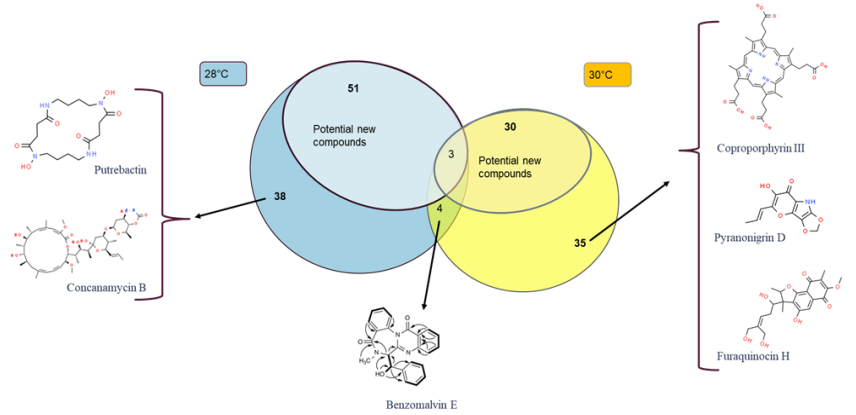
We then employed metabolomics analysis through molecular networking on the crude organic and aqueous extracts of SM3 cultures that had been fermented in the seven different culture media namely- M19, M400, MMM, OM, SGG, SM and SYP (Table 2) at both 28°C and 30°C; to examine the SM3 metabolome and to visualize the chemical diversity of the metabolites produced by the strain growing under these different conditions, by comparing the mass spectrometry profiles present in the extracts (Figure 12a).



a)



b)



c)

Figure 12: Molecular network of SM3 fermented at 28°C and 30°C (a). Venn diagram of the specific and shared detected compounds, including potential new compounds, for SM3 (b) and B226SN104 strain (c) fermented at 28°C and 30°C respectively.

The node (molecular network) and metabolite (venn diagram) originated from media/temperature controls were excluded from the network and venn diagrams to enable visualization of only the metabolites potentially produced by our isolates cultured in these specific parameters.

Molecular network analysis was performed using the MS/MS based Global Natural Product Molecular Networking GNPS online platform, which employs rapid and automated mass spectral mining of large number of samples by using high-resolution mass spectrometry parent ion fragmentation data (MS/MS); with parent ions being connected based on the chemical architecture of the metabolite (Watrous *et al*., 2012; Fan *et al*., 2019; Oppong-Danquah *et al*., 2018). In addition it has also been reported that molecular networks can be used to link known and unknown compounds belonging to the same molecular family within the network and deliver more in-depth information on the chemical inventory of the producing organism (Fan *et al*., 2019 and Oppong-Danquah *et al*., 2018). Molecular networks can also be used to compare and/or optimize the culture conditions, and the extraction methods from both organic and aqueous phases for bacterial natural products (Fan *et al*, 2019).

Figure 12a represents the metabolites detected by GNPS-Cytoscape in all the samples fermented at 28°C and 30°C respectively, with the blue and yellow nodes corresponding to the metabolites, with the edges corresponding to the relatedness of the individual nodes within a cluster based on the cosine score. The cosine score determines the relationship between individual nodes to each other which is indicated by the visual thickness of the edge between related nodes in the network. The higher the cosine score, the thicker the edge will be and therefore the more related the individual nodes are to each other (Wang *et al*., 2016). It is clear that temperature is an important parameter with respect to the production of metabolites in SM3, with different metabolites being produced when the strain is grown in the same growth medium, but at different temperatures Figure 12a. These metabolites are visible as single blue or yellow colours on the nodes in the network, which demonstrates that if fermented in the same media but at different temperatures, different metabolites are produced. In contrast multi-coloured blue and yellow nodes represent metabolites which are detected at both temperatures,

indicating that they are produced when SM3 is cultured at both at 28°C and 30°C (Figure 12a). When focusing on the specific groups of metabolites observed when SM3 is grown at 28°C, one group composed of seven metabolites that are strongly related to each other; based on the edges thickness visible on the network, is clearly evident. The presence of this group of metabolites, both single and multi-coloured, with high cosine scores, suggests that these molecules or "metabolite cluster" may potentially have been produced through the same synthetic route and thus may originate from the same gene cluster and be regulated at least to some extent by temperature, particularly given that they are only produced when SM3 is grown at 28°C and not 30°C.

Further analysis involving the total manual de-replication of metabolites detected from all the organic and aqueous phase extracts when SM3 was cultured in all 7 culture media at both temperatures indicated that only 9.1% (15 of 165 metabolites) could be detected in extracts from SM3 when the culture was grown at both 28°C and 30°C. In contrast 102 (61.8%) and 53 (32.1%) metabolites were only present in extracts of SM3 grown at 28°C or at 30°C respectively (Figure 12b). Manual de-replication revealed that the production of metabolites such as Vitexin and Soyasaponin I only occurs at 28°C, while the production of other metabolites such as Notoamide J and Puerarin in SM3 only occurs at 30°C (Figure 12b).

When a similar analysis was performed on the manually de-replicated metabolites detected in all organic phase extracts produced by the B226SN104 strain at both temperatures, it appeared that 4.3% (7 of 161 metabolites) were produced at both temperatures, while 55.3% (89 of 161 metabolites) and 40.4% (65 of 161 metabolites) were only produced at 28°C and 30°C respectively (Fig 12c). Thus as with SM3 it also appears that production of some specific metabolites only occurs in B226SN104 at certain temperatures with Putrebactin and Concanamycin B only being produced at 28°C, and Coproporphyrin III, Pyranonigrin D and Furaquinocin H only being produced at 30°C (Fig 12c). Furthermore, when specifically

focusing on potential new metabolites (Fig 12b and 12c), and comparing the number of new metabolites detected in both strains at 28°C and 30°C; it appears that growth at 28°C resulted in an increase in the production of potential new metabolites. For example, in B226SN104 extracts, 51 and 30 potential new metabolites were detected when cultured specifically at 28°C and 30°C respectively compare to 3 when cultured at both temperatures. Expressed as a percentage of total potential new metabolites, these numbers correspond to: 21.8% (in the 28°C culturing condition), 9.7% (in the 30°C condition) and only 1.2% (cultured at both temperatures), respectively.

Temperature has previously been reported to effect metabolite production in different microorganisms. In fungi for example temperature is known to regulate aflatoxin production in *Aspergillus flavus* (Lind *et al.*, 2016), terpene T-2 toxin production in *Fusarium* species (Nazari *et al.*, 2016) as well as Trypacidin and Endocrocin production in *A. fumigatus* (Hagiwara *et al.*, 2017). Temperature also affects metabolite production in marine microorganisms. For example, the optimal temperature for antimicrobial metabolite production with anti-MRSA (Methicillin Resistant *Staphylococcus aureus*) activity in *Pseudoalteromonas pricida* PG02 has been reported to be 28°C (*Darabpour et al. 2012*), while optimal prodigiosin production in *Serratia marceans* occurs at 30°C (Thomson *et al.*, 2000) and antibiotic production in *Vibrio coralliticus* strains LMG10953 and LMG20984, was also found to be optimal at 30°C (*Wietz et al. 2011*). In addition, the optimal production of secondary metabolites with antibacterial, antifungal and antioxidant activities in the deep sea sediment isolate *Streptomyces sparsus* VSM-30 has been reported to occur at 30°C (Managamuri *et al.*, 2017). While valinomycin production in the cold desert isolate *Streptomyces lavendulae* is regulated at lower temperatures with optimal production at 10°C, with lower levels being produced at 20°C. This increased production has been shown to occur as a result of the

upregulation of the valinomycin biosynthetic genes at 10°C (Sharma *et al*., 2017). Temperature has also been shown to regulate validomycin production in *Streptomyces hygroscopicus*, with higher levels being produced at 37°C rather than at 30°C, which is the normal temperature for growth and sporulation. A SARP-family regulatory gene has been shown to be involved in the thermo-regulation of validamycin production in this strain (Wu *et al*., 2012). Thus it is perhaps not surprising that different metabolites were observed in both SM3 and in B226SN104 when both were cultured at different temperatures (Figure 12a, 12b, and 12c).

2.2.2 Nutrient source media involvement in the production of secondary metabolites from both B226SN104 and SM3 isolates

It is also important to consider that the observed effect of temperature on metabolite production in both strains is also likely to be influenced by the nutrient media in which the strain is being cultured at the given temperature. Indeed the specific conditions for the production of each metabolite differed by strain and/or media and/or temperature. For example, some metabolites were detected under a particular set of conditions, such as Cyclo (L-leucyl-L-propyl), Vitexin, Puerarin, Putrebactin, Furaquinocin H, Mannioside A, Coproporphyrin III, Serinolamide A, among others, which were only detected at a specific temperature when SM3 and/or B226SN104 are cultured in a specific nutrient media condition (Table 4). In contrast other metabolites were detected following growth of one strain in one specific condition regardless of the culture temperature, such as for example Furaquinocin D which was detected in SM3 following growth of the strain in OM media specifically, at both temperatures 28°C and 30°C. Other metabolites, however, such as Gibbestatin B were detected under all the media tested (M19, M400, MMM, OM, SGG, SYP and SM) and at both temperatures tested (28°C and 30°C) in SM3.

Strain	Temperature of growth	Nutrient source media	Extract Phases	$[M+H]^+$	Molecular formula	Compound name	Reported Biological activity	Additional reported information (Synthesis route, chemical group ect.)
SM3	28ºC	SM	Organic	433.1132	$C_{21}H_{20}O_{10}$	Vitexin	Anti-oxidant, Anti-cancer, Anti-inflammatory and Anti-hyperalgesic activity with also neuroprotective capacity (He, et al. 2016).	n/a
SM3	28ºC	SM	Organic	943.5263	$C_{48}H_{78}O_{18}$	Soyasaponin I	Potent and Specific Sialyltransferase inhibitor (Wu et al. 2001).	n/a
SM3	28ºC	M19, OM, MMM, SGG, SYP	Organic	663.4520	$C_{34}H_{48}O_{13}$	Sarmentoside B	Cardiac toxicity (Callow et al. 1952).	Sterol glycoside (Callow et al. 1952).
SM3	28ºC	M19	organic	211.1446	$C_{11}H_{18}N_2O_2$	Cyclo (L-leucyl-L-propyl)	Antileukemic and anti-vancomycin-resistant enterocci (VRE) (Rhee, 2002). Inhibitor of aflatoxin production (Yan et al. 2004).	n/a
SM3	28ºC	M19, SGG, SM, SYP	Organic and water	387.1811	$C_{23}H_{22}N_4O_2$	Fiscalin B	Antitumor activity against non-small cell lung cancer (NCI-H460) and colorectal adenocarcinoma (HCT-15) cell lines, neuroprotection capacity (Long et al. 2019).Inhibit the binding of radiolabeled substance P ligand to the human neurokinin (NK-1) receptor (Wong et al. 1993).	Quinazolinone Alkaloids (Long et al. 2019)
SM3	28ºC	M19, MMM, OM, M400, SGG, SM	organic	403.2321 to 403.2323	$C_{20}H_{34}O_8$	Botcinolide	Weak antifungal activity against Magnaporthe grisea (Tani et al. 2006).	n/a
SM3	28ºC - 30 ºC	all	Organic and water	387.1799 - 387.1806	$C_{22}H_{26}O_6$	Gibbestatin B	Inhibits the GA-induced expression of α-amylase (Hayashi et al. 2000).	n/a
SM3	28ºC - 30 ºC	OM	organic	387.1804	$C_{22}H_{26}O_6$	Furaquinocin D	cytocidal activities against HeLa S3 and B16 melanoma cells in vitro (Ishibashi et al. 1991).	PKS-Hybrid (Kumano et al. 2010)
SM3	30 ºC	MMM	water	836.403	$C_{38}H_{57}N_7O_{14}$	Formestin B	n/a	n/a
SM3	30 ºC	M400, OM, SGG, SYP	water	384.1923	$C_{21}H_{25}N_3O_4$	Notoamide J	Potential biosynthetic precursor and cytotoxicity against HeLa cells with an IC_{50} value of more than 50 µg/mL (Tsukamoto et al. 2008)]. Potential Pivotal Intermediate in the Biosynthesis of Several Prenylated Indole Alkaloids and anti-cancer activity (Ding et al. 2010; Finefield et al. 2010).	Prenylated indole alkaloid (sukamoto et al. 2008).
SM3	30 ºC	SM	organic	417.1185	$C_{21}H_{20}O_9$	Puerarin	Vasodilatation, cardioprotection, neuroprotection, antioxidant, anticancer, anti-inflammation, alleviating pain, promoting bone formation, inhibiting alcohol intake, and attenuating insulin resistance (Zhou et al. 2013).	Isoflavone C- glycoside (Zhou et al. 2013).
B226SN104	28ºC	M400, SYP	organic	852.5100	$C_{45}H_{73}NO_{14}$	Concanamycin B	Specific inhibitor of the vacuolar type H(+)-ATPase, inhibit PTH-stimulated osteoclastic pit formation, suppress the acidification of vacuolar organelles by V-ATPase in the osteoclasts (Woo et al. 1996.; Lee et al. 2000).	n/a
B226SN104	28ºC	M400	organic	373.2092	$C_{16}H_{28}N_4O_6$	Putrebactin	Iron ion binding (Codd et al. 2018)	Siderophore (Ledyard et al. 1997)
B226SN104	28ºC	MMM	organic	245.1286	$C_{14}H_{16}N_2O_2$	Cyclo (phenylalanyl-propyl)	Potential muscle relaxant, antibiotic and anticancer, active in prokaryote and eukaryote (Milne et al. 1998).	n/a
B226SN104	28ºC	M19	organic	201.1481	$C_{11}H_{20}O_3$	Okaspirodiol	Potential antifungal activity (Bender et al. 2006).	n/a

B226SN104	28ºC 30 ºC	M400, MMM, SGG, SYP	Organic Organic	398.1498 398.1502	$C_{24}H_{19}N_3O_3$	Benzomalvin E	Indoleamine 2,3-dioxygenase inhibitor, leads to apoptosis and immunosuppression [Jang et al., 2012].	n/a
B226SN104	30 ºC	SYP	organic	419.1709	$C_{22}H_{26}O_8$	Furaquinocin H	cytocidal activities against HeLa S3 and B16 melanoma cells in vitro (Ishibashi et al. 1991)	**Polyketide**-isoprenoid hybrid (meroterpenoid) (Kumano et al. 2010)
B226SN104	30 ºC	M400	organic	235.0516	$C_{11}H_9NO_5$	Pyranonigrin D	n/a	Potential PKS-NRPS hybrid (as Pyranonigrin E is pks-nrps hybrid (Awakawa et al. 2013)
B226SN104	30 ºC	OM	organic	739.4265	$C_{39}H_{62}O_{13}$	Mannioside A	In vivo anti-inflammatory effect (Tapondjou et al. 2008)	n/a
B226SN104	30 ºC	M19	organic	384.8468	$C_{23}H_{45}NO_3$	Serinolamide A	Cannabinomimetic agents, moderate agonist effect and selectivity for the CB_1cannabinoid receptor. (Gutiérrez et al. 2011)	Probable PKS (Gutiérrez et al. 2011)
B226SN104	30 ºC	M400	water	655.2770	$C_{36}H_{38}N_4O_8$	Coproporphyrin III	bound zinc in excess over other metals antibacterial activity (Cleary et al. 2018)	Siderophore (Cleary et al. 2018)
B226SN104	30 ºC	SM	water	415.1250	$C_{19}H_{18}N_4O_7$	Futalosine	Cytotoxic activity (Schneemann et al. 2010)	Nucleoside (Schneemann et al. 2010)
B226SN104	30 ºC	OM	organic	505.1775	$C_{18}H_{32}O_{16}$	Globotriose	Oral therapeutic agent with *in vivo* and in vitro effect, inhibit and reverse the binding and agglutination of a P-fimbriated strain of E. coli and is a feasible alternative therapy for urinary tract infections (Leach et al. 2005).	n/a

Table 4: Some compounds produced by SM3 and B226SN104 in different conditions following the OSMAC approach.

All compounds were identified by LC-MS, manual de-replication, Chemspider, Antibase, GNPS and Cytoscape. For some samples, different $[M+H]^+$ for different media extracts and or phases. n/a= not reported.

Another example of the combined effect of temperature and culture media is the production of Vitexin which was only detected in extracts from SM3 following growth in SM media at 28°C, while Serinolamide A was only detected when B226SN104 was grown in M19 media at 30°C (Table 4).

Based on the most promising number of potential new metabolites obtained from the manual de-replication in SM3, both aqueous and organic phase extracts from six of the seven media in which the strain had been grown were selected for further metabolomics analysis through the establishment of a Molecular Network (Figure 13a). It is clear that nutrient source plays an importance role in metabolite production in SM3, with some metabolites appearing to only be produced in specific media (single colour node), such as Cyclo (L-leucyl-L-propyl) which is only detected in the M19 extracts; while other metabolites appear to be produced in a variety of different media (multi-colour node), such as Sarmentoside B which was detected by GNPS and Cytoscape in extracts from five different media namely M19, MMM, OM, SGG and SYP.

Thus it is likely that SM3 contains a number of potential "silent" or "cryptic" gene clusters encoding for the production of specific secondary metabolites that are only produced under a specific set of nutrient conditions or at a specific temperature. For example Cyclo (L-leucyl-L-propyl) is only produced when the strain is fermented in M19 media, but is not produced in any of the other media tested when fermented at the same temperature. Furthermore, it may be interesting to investigate if the metabolite cluster involving the 7 metabolites with high cosine scores, that appear to be only produced in SM3 when grown in the M19 media (Figure 13a) are products of the same or a related gene cluster in the strain.

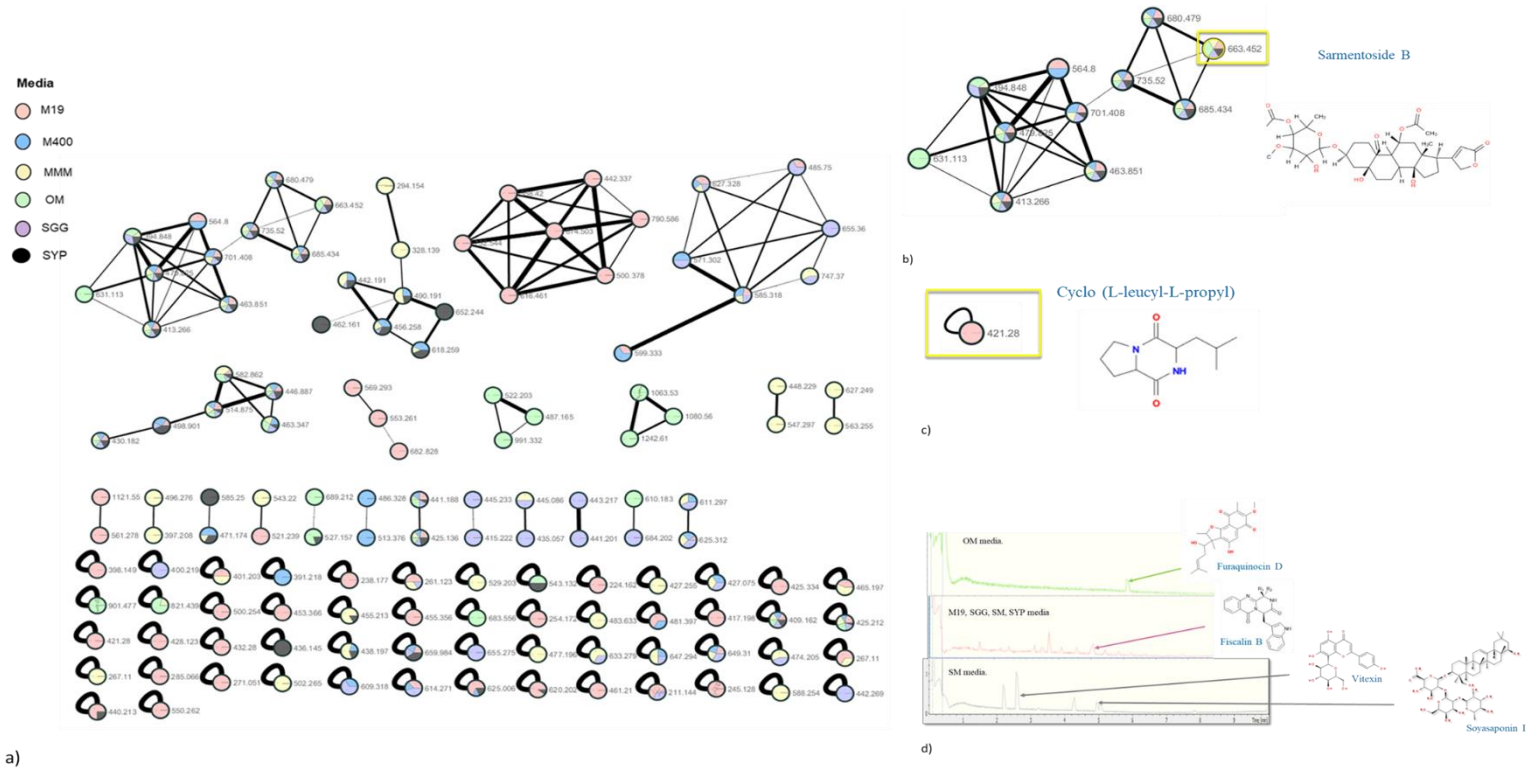


Figure 13: Molecular network of SM3 fermented in six media (a). Enlarged sub-network view of the compounds Sarmentoside B in the cluster (b) and of the compound Cyclo (L-leucyl-L-PROPYL) isolated in the network (c). Chemical structure of manually dereplicated compounds (d).

Results obtained using: Chemspider, Antibase, Reaxys, GNPS and Cytoscape. The nodes corresponds to the compounds detected by GNPS and Cytoscape. The edges corresponds to the individual node relatedness within a cluster based on the cosine score. This value determines the relationships between individual nodes to each other which is indicated by the visual thickness of the edge between related nodes in the network. The chemical structure pictures were generated from both the GNPS and Chemspider software.

Figure 13b represents the sub-network of Sarmentoside B metabolites that were identified, in a number of different culture media (namely M19, MMM, OM, SGG and SYP) but not when SM3 was grown in M400 media. Sarmentoside B is a sterol glycoside, which has previously been reported to display cardiac toxicity activity (Callow *et al.* 1952). While in contrast, Cyclo (L-leucyl-L-propyl) was only produced when SM3 was grown on M19 growth media which contains conditions mannitol and peptones as nutrient sources (Figure 13c).

While a number of metabolites were identified that could be linked to the metabolomic network others metabolites that were identified, such as Furaquinocin D, Fiscalin B, Vitexin and Soyasaponin I, could not be linked to the metabolomic network (Figure 13d). Of these Soyasaponin I and Vitexin were only detected when SM3 was grown in SM media, while Furaquinocin D was only detected when SM3 was grown in OM media (Figure 13d). Soyasaponin I is a potent and specific sialyltransferase inhibitor (Wu *et al.* 2001); while Vitexin is a metabolite which has been reported to have anti-oxidant, anti-cancer, anti-inflammatory and anti-hyperalgesic activity, while also possessing neuroprotective capacity (*He et al. 2016*). Furaquinocin D is known to be synthesised by a hybrid-PKS and to possess cytocidal activities (*Ishibashi et al. 1991; Kumano et al. 2010*), while Fiscalin B has been reported to possess antitumor activity and to display a neuroprotection capacity (*Long et al. 2019*)

2.3 Potential new metabolites detected from SM3 and B226SN104 crude extracts

Figure 14 represents a Venn diagram indicating the numbers of metabolites produced when SM3 was grown on 4 different growth media, namely M 19, Medium 400, Oatmeal and MMM media. The four media were selected from the results of the molecular network analysis on SM3 with the manual de-replication results together as the more promising media for the production of potential new compounds fermented at 28˚C. Growth of SM3 in M19 media resulted in the detection of the highest numbers of metabolites suggesting that a combination of mannitol-peptone in the growth medium promoted production of these metabolites (Figure 14a). In total 32 metabolites were identified in the M19 extracts, including 27 metabolites that were only present following growth in M19 (Fig 14a). In addition, a total of 19 potential new metabolites were detected in M19 extracts from SM3, while only 6, 6 and 7 potential new metabolites were detected in extracts following growth in OM, MMM and M400 respectively, indicating that of the 4 growth media analyzed, M19 appears to be optimal, for the production of potential new metabolites by SM3 (Fig 14b).

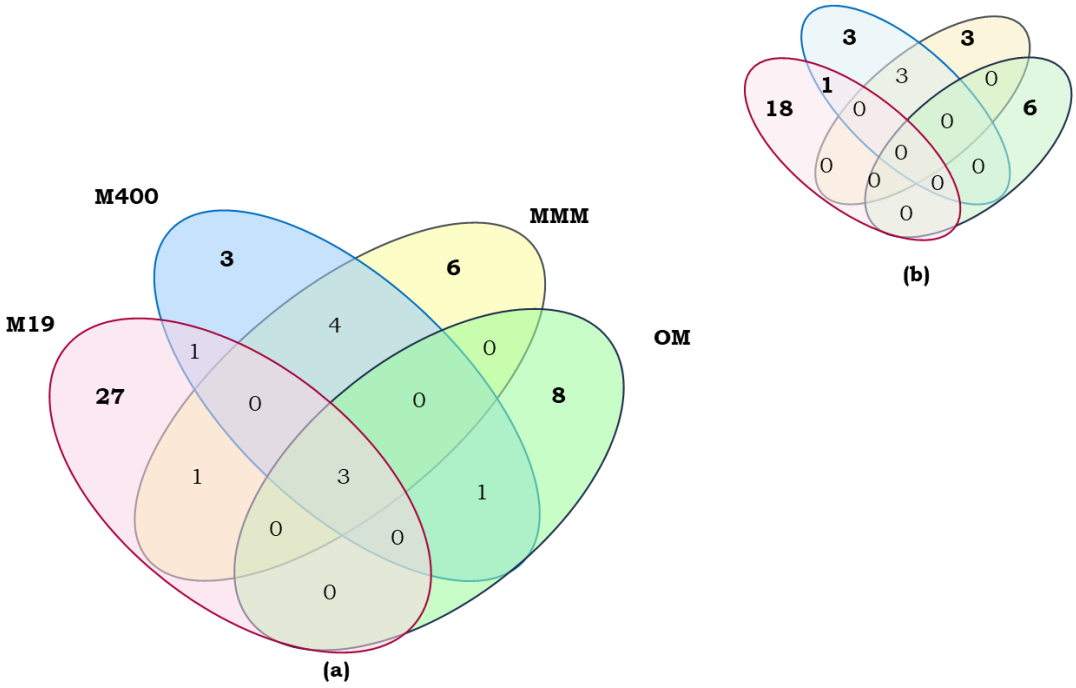


Figure 14: Venn diagram of the specific and shared compounds (a), including potential new compounds (b), produced by SM3, detected in the four most promising media in both aqueous and organic phase.

Based on the most promising number of potential new metabolites obtained from the manual de-replication in the deep sea isolate B226SN104, aqueous and organic phase extracts from the strain grown on four different growth media were analyzed. Growth of B226SN104 in M19, oatmeal (OM), MMM and SGG resulted in the detection of a total of 106 metabolites, including 77 metabolites which were only produced following growth in one specific media (Figure 15). Growth of B226SN104 on M19 media (56 metabolites) and the SMG media (48 metabolites) resulted in the highest numbers of metabolites being produced. Thus these two media appear better than OM and MMM growth media as nutrient sources to promote the production of secondary metabolites in B226SN104 (Fig 15a). In addition, when specifically focusing on the production of potential new metabolites, a total of 31 and 28 potential new metabolites were detected in M19 and SGG extracts respectively, while only 8 and 6 potential new metabolites being detected in the MMM and OM media, indicating that as with the production of total

metabolites by B226SN104, both M19 and SGG were again better than MMM and OM for production of potential new metabolites produced by the strain (Fig 15b).

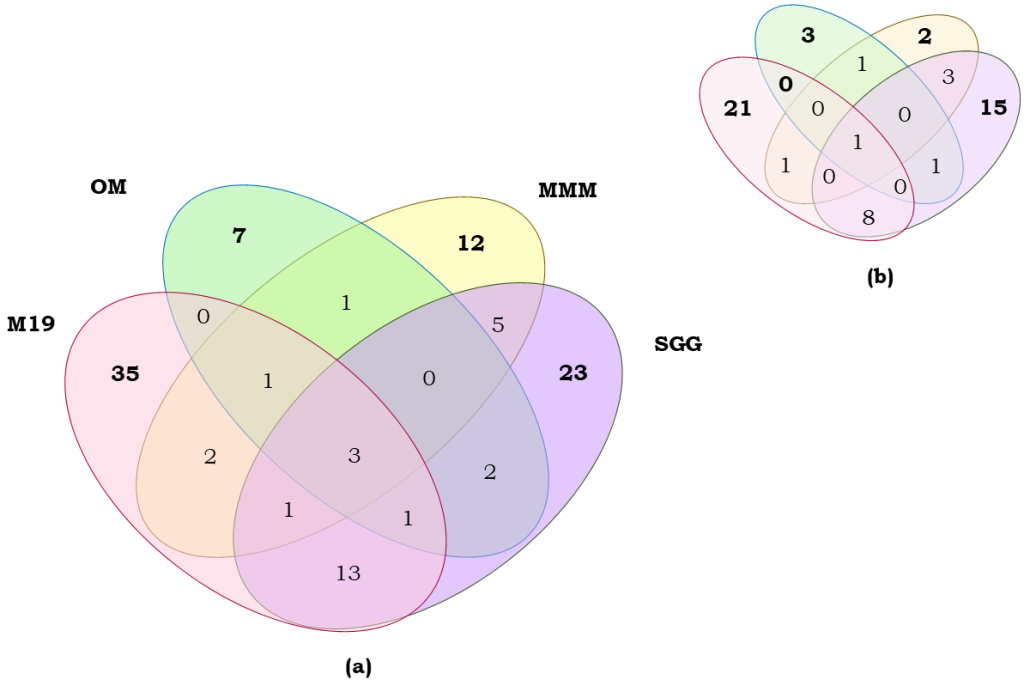
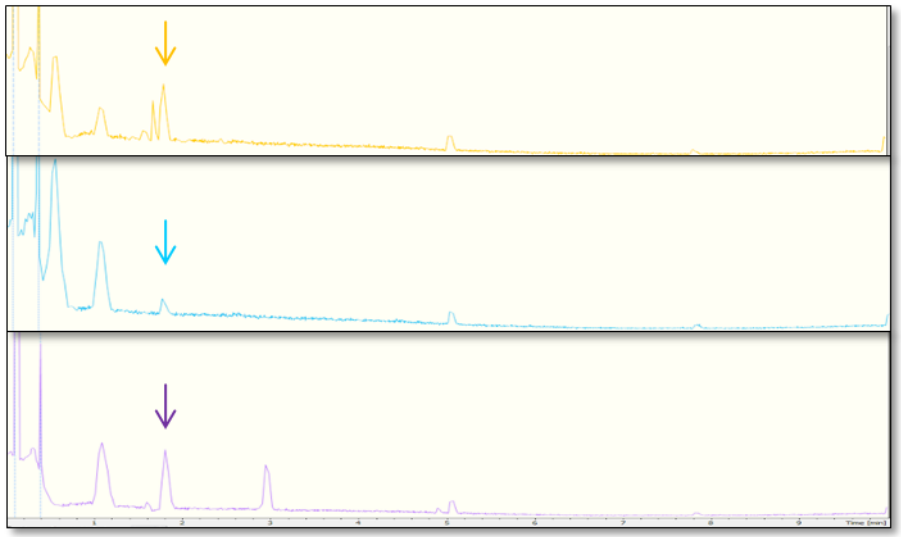
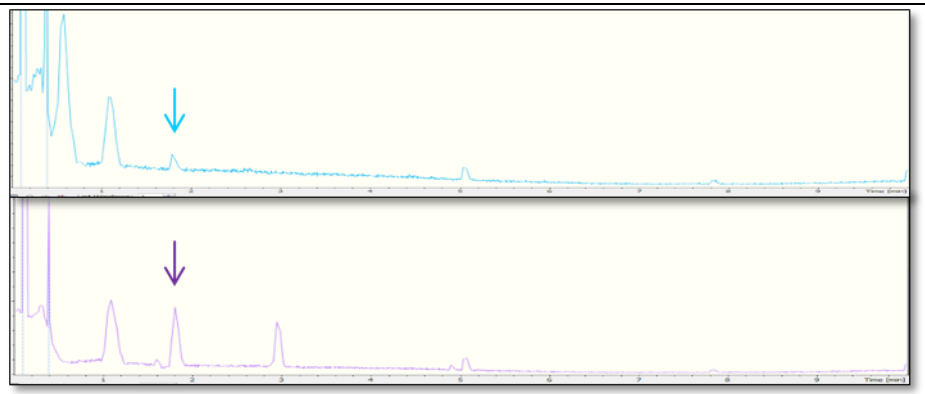
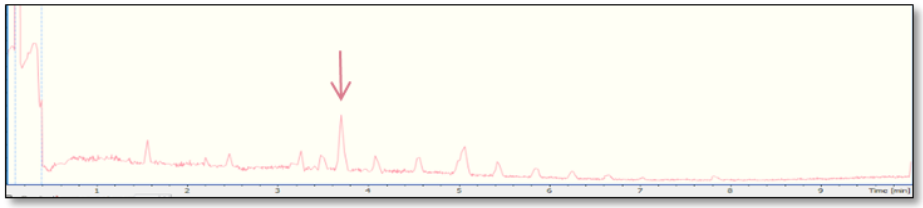
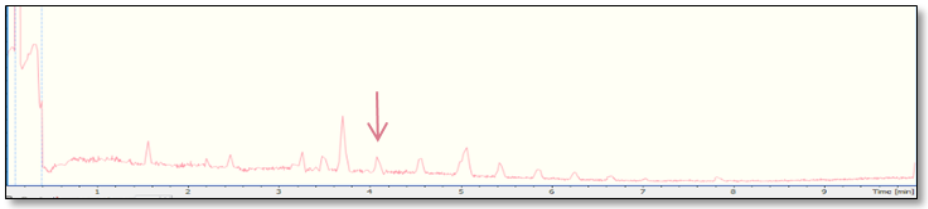
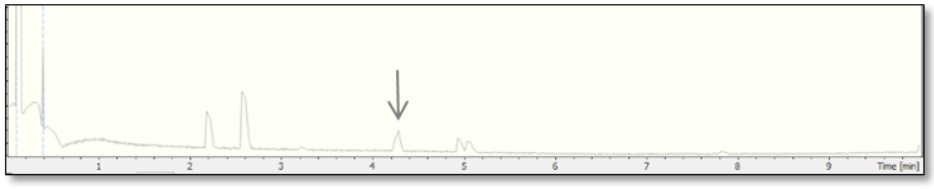


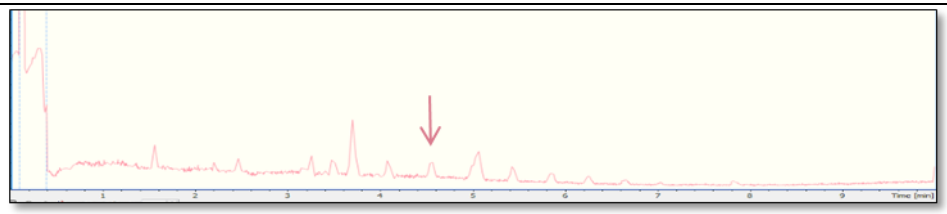
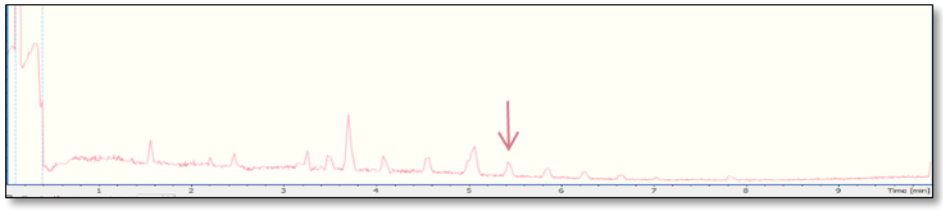
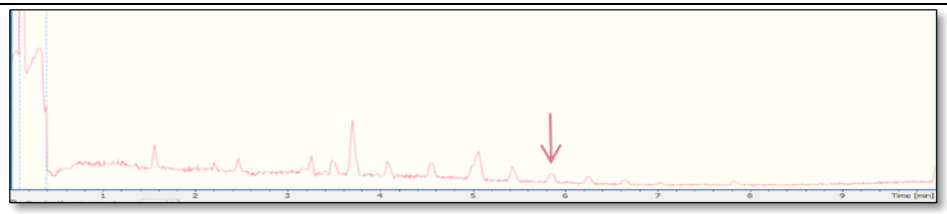
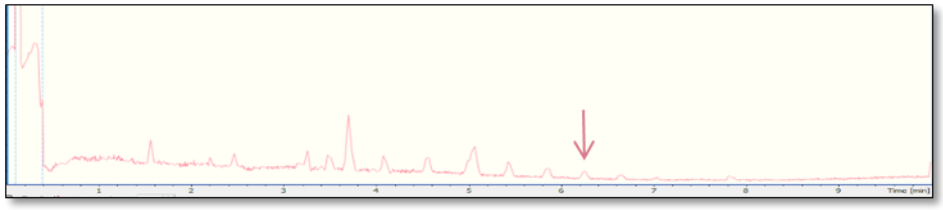
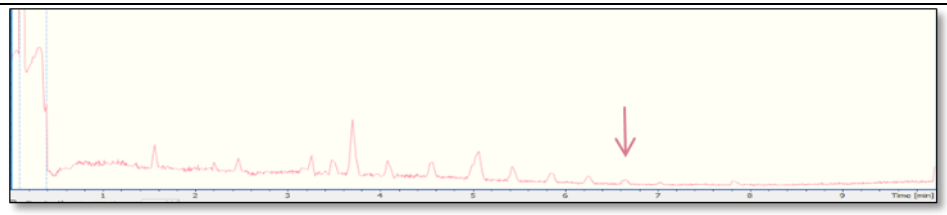
Figure 15: Venn diagram of the specific and shared compounds (a), including potential new compounds (b), produced by B226SN104 strain, detected in the four most promising media in organic phase.

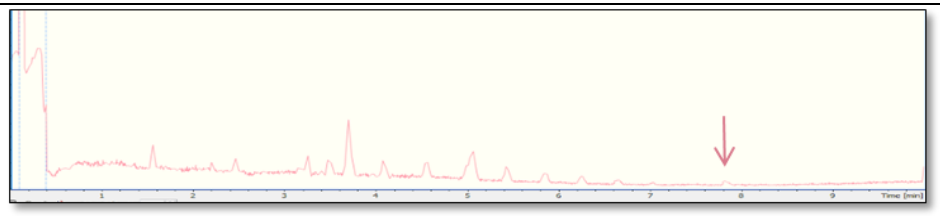
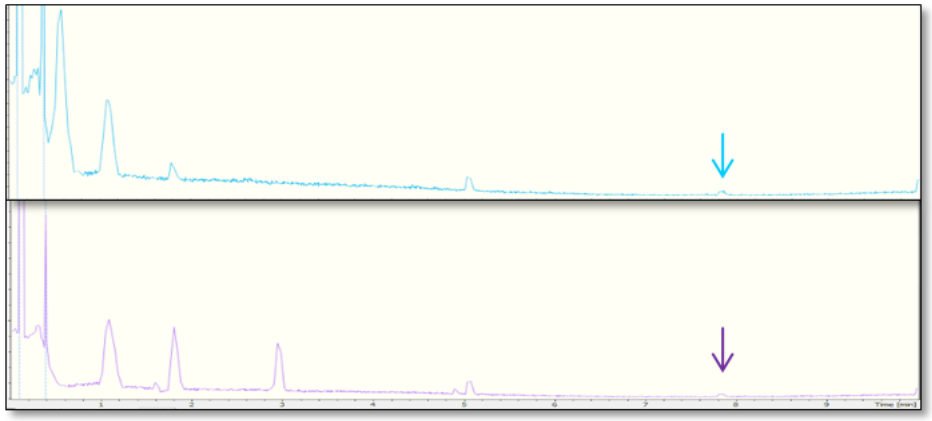
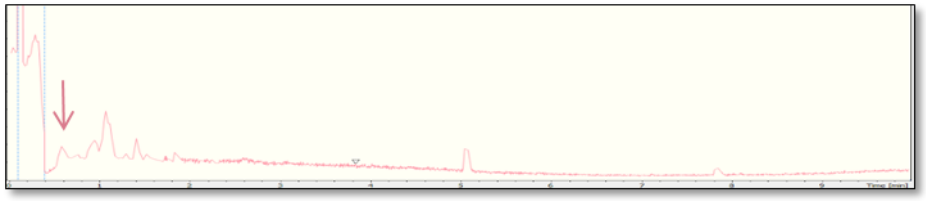
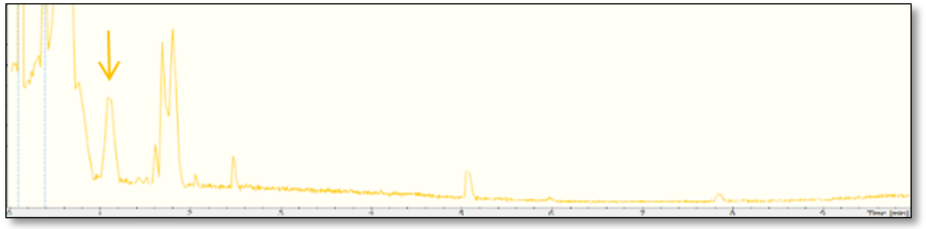
The most interesting metabolites detected during this OSMAC approach experiment from both SM3 and B226SN104 crude extracts with their culturing conditions are listed in Table 4, together with the potential new metabolites detected from SM3 crude extracts fermented at 28°C in seven media listed in Table 4. Both tables, (Table 4 and Table 5), highlight the fact that the production of many of the metabolites detected is regulated differently in each strain, and that subtle changes in nutrient source or in temperature, even if a change of only 2 degree (28-30°C) can result in not only marked change in the metabolites been produced and in the metabolites not been produced, but also to changes in the resulting bioactivity being targeted. For example, this is demonstrated in Table 4, where both Vitexin and Soyasaponin I

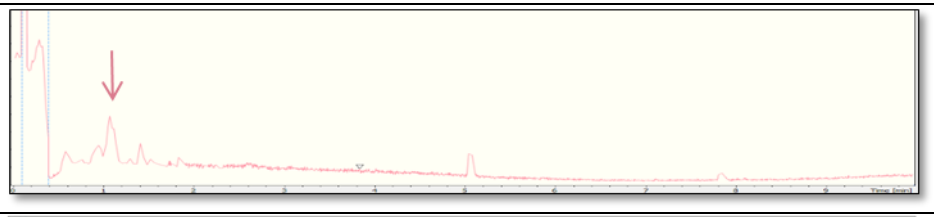
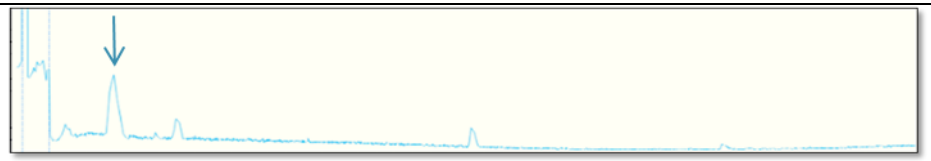
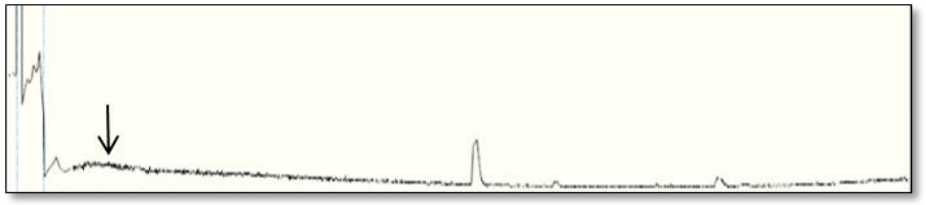
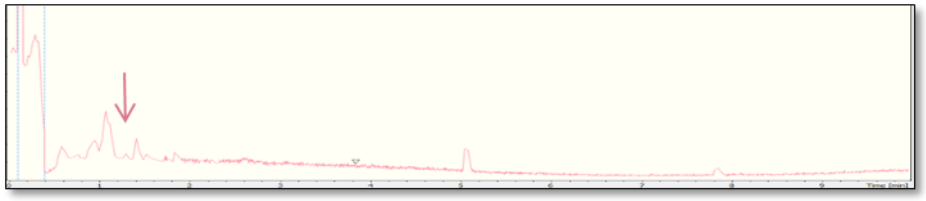
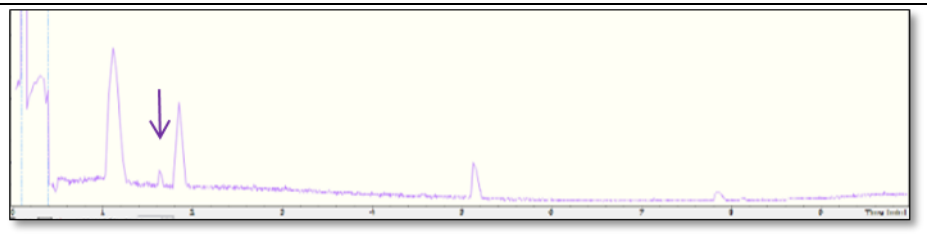
metabolites were only detected in SM3 extracts following growth in SM media at 28ºC degree; whereas, when cultured in the exact same media at 2 degree higher (30°C), these metabolites where not only not detected but another metabolite, not detected in 28°C extract, Puerarin, was detected. Furthermore, this tables highlight the possibility that each strain has the capacity to produces not only metabolites that are already know and which have established bioactivities, but also potentially new metabolites.

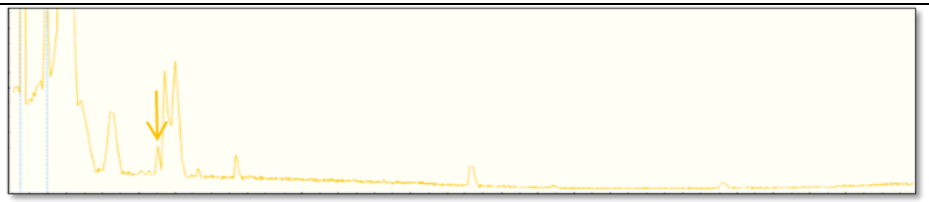
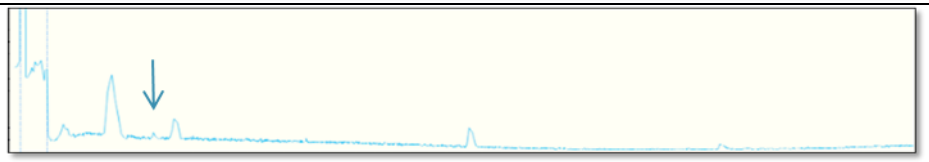
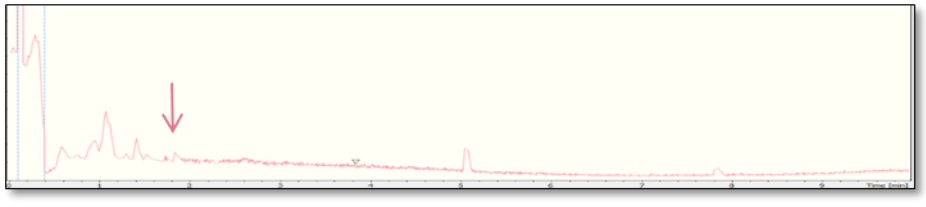
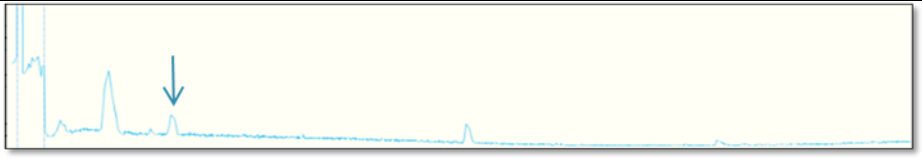
Temperature of growth	Nutrient source media	Extract Phases	Retention time [min]	$[M+H]^+$	Molecular formula	Libraries identification: ChemSpider ID, AntiBase ID, PubChem ID and Reaxys ID	Chromatograms
28ºC	MMM M400 SGG	Organic	1.1	585.3187 585.3190 585.3189	$C_{21}H_{44}N_8O_{11}$	no hit no hit no hit	
28ºC	MMM M400 SGG	organic	1.8	625.3139 625.3137 625.3132	$C_{20}H_{36}N_{18}O_6$	no hit no hit no hit	

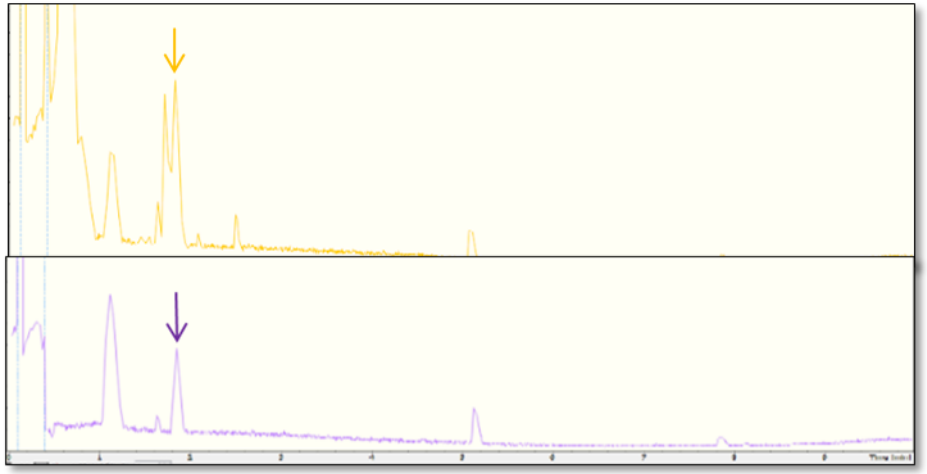
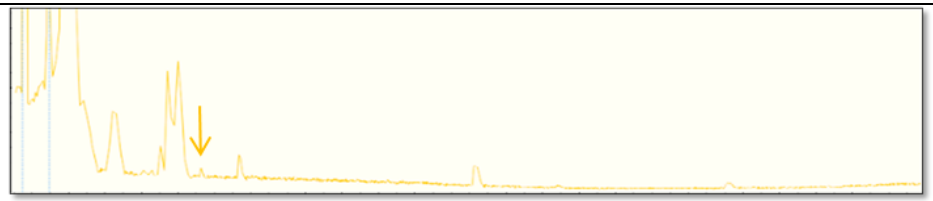
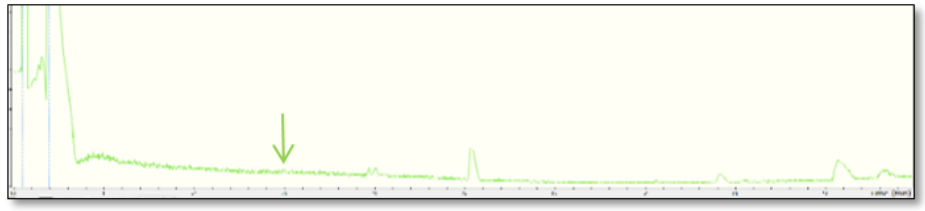
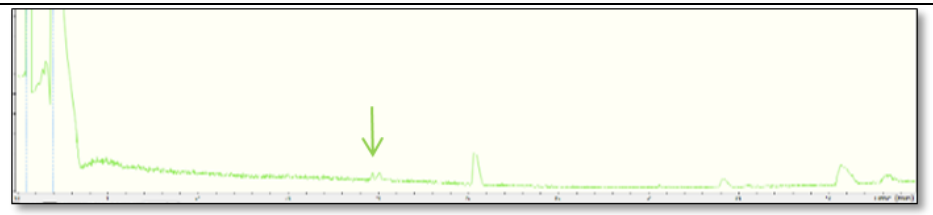
28ºC	M400 SGG	Organic	1.8	625.3137 625.3132	$C_{22}H_{48}N_4O_{16}$ $C_{23}H_{44}N_8O_{12}$	no hit no hit	
	M19	organic	3.7	249.0703	$C_{20}H_8$ $C_5H_4N_{12}O$	no hit	
28ºC	M19	organic	4.1	425.3114	$C_{21}H_{44}O_8$	no hit	
28ºC	SM	organic	4.3	683.4704	$C_{34}H_{62}N_6O_8$	no hit	

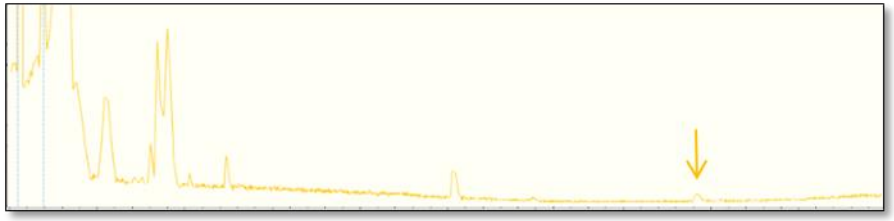
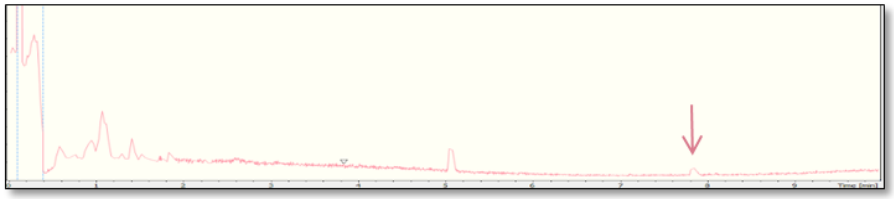
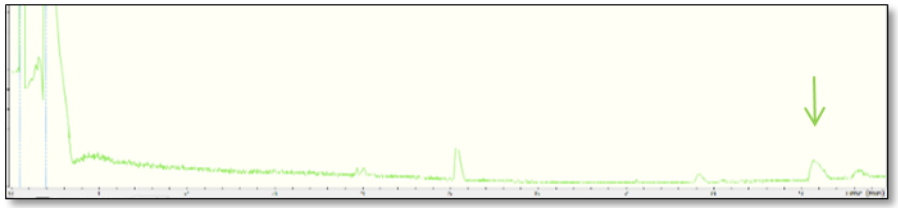
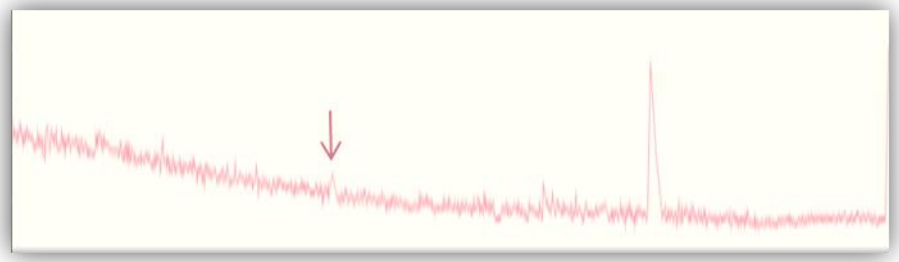
28ºC	M19	organic	4.6	483.3531	$C_{24}H_{50}O_9$ $C_{25}H_{46}N_4O_5$	no hit	
28ºC	M19	organic	5.4	616.4635	$C_{31}H_{61}N_5O_7$ $C_{30}H_{65}NO_{11}$	no hit	
28ºC	M19	organic	5.9	674.5038	$C_{34}H_{67}N_5O_8$ $C_{33}H_{71}NO_{12}$	9171191 no hit	
28ºC	M19	organic	6.2	732.5464	$C_{34}H_{65}N_{15}O_3$ $C_{36}H_{77}NO_{13}$	no hit	
28ºC	M19	organic	6.7	790.5875	$C_{40}H_{79}N_5O_{10}$	no hit	

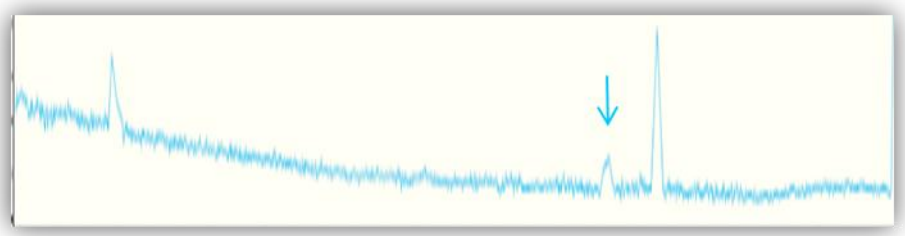
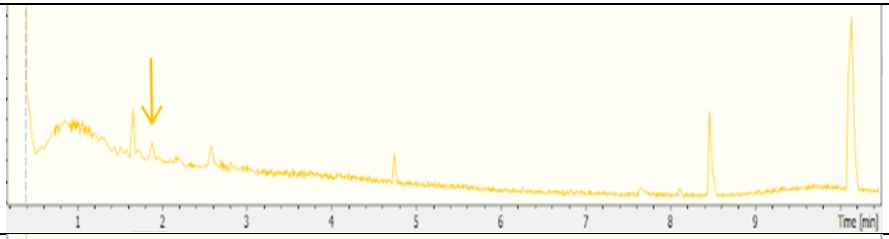
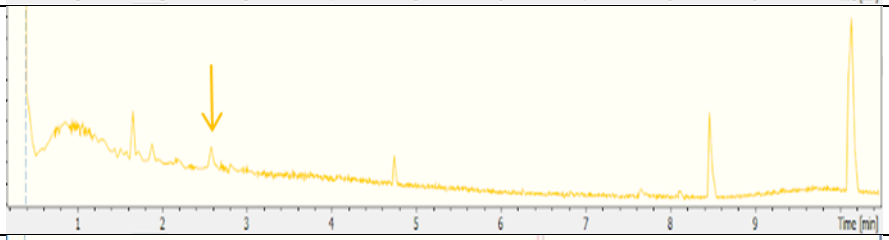
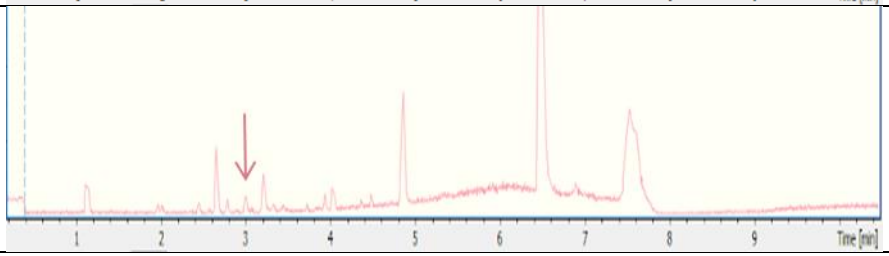
28ºC	M19	organic	7.8	403.2321	$C_{24}H_{38}O_4$ $C_{13}H_{33}N_5O_9$	11575 no hit	
28ºC	M400, SGG	Organic	7.8	403.2336 403.2336	$C_{21}H_{30}N_4O_4$ $C_{24}H_{38}O_4$ $C_{21}H_{30}N_4O_4$	no hit 11575 no hit	
28ºC	M19	water	0.6	625.1592	$C_{21}H_{24}N_{10}O_{13}$	no hit	
28ºC	MMM	water	1.1	585.3186	$C_{21}H_{44}N_8O_{11}$	no hit	

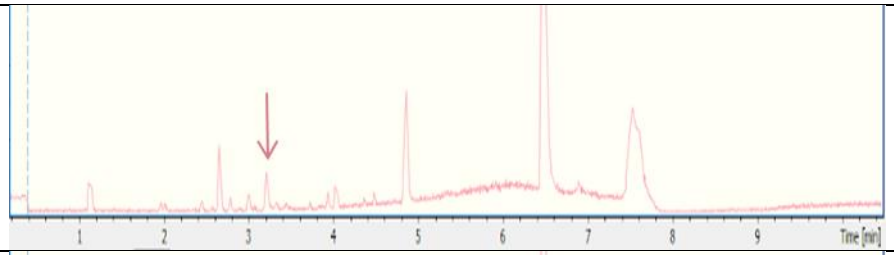
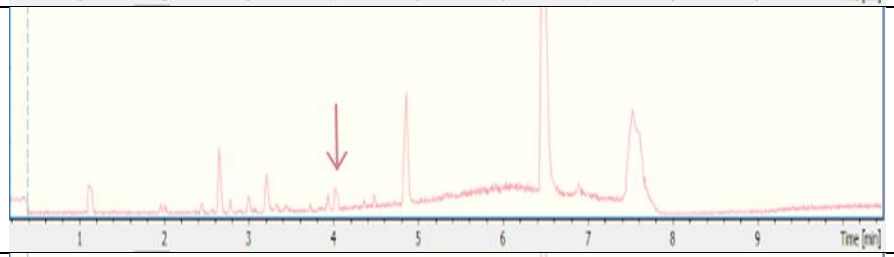
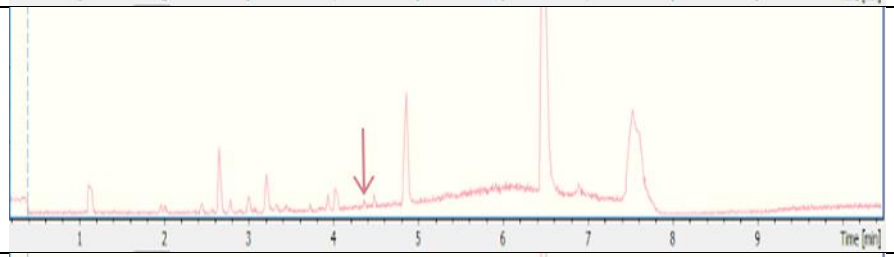
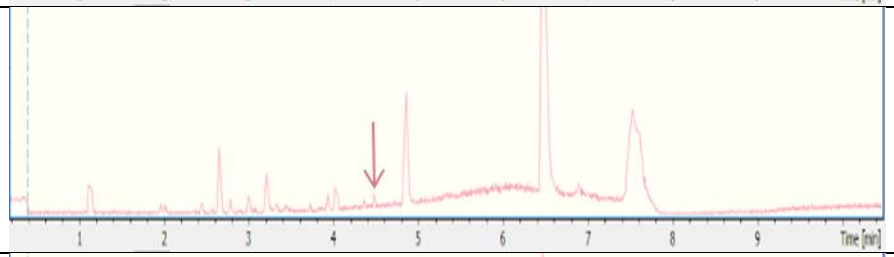
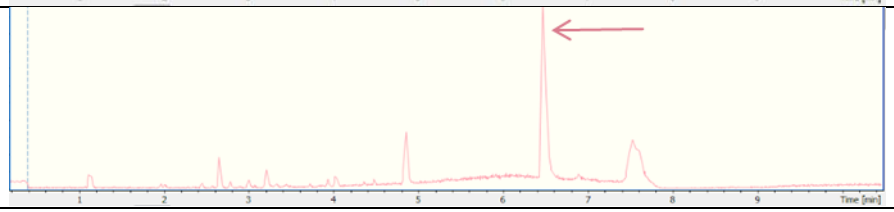
28ºC	M19	water	1.1	585.3181	$C_{20}H_{48}N_4O_{15}$ $C_{20}H_{48}N_4O_{16}$	no hit	
28ºC	M400	water	1.1	585.3158	$C_{20}H_{48}N_4O_{16}$	no hit	
28ºC	SYP	water	1.1	585.2506	$C_{21}H_{28}N_{16}O_5$ $C_{24}H_{36}N_6O_{11}$	no hit	
28ºC	M19	water	1.3	428.1229	$C_{17}H_{35}NO_{11}$	no hit	
28ºC	SGG	water	1.6	611.2975	$C_{17}H_{28}N_4O_7$	no hit	

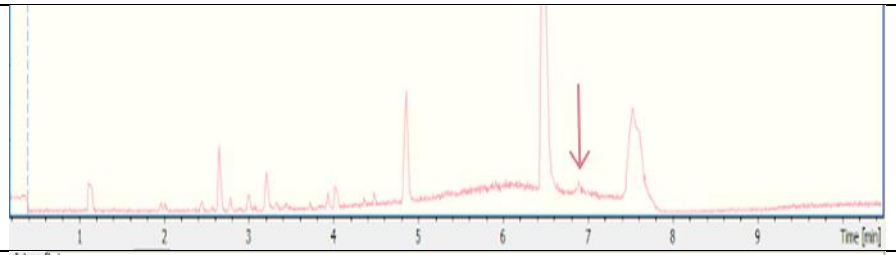
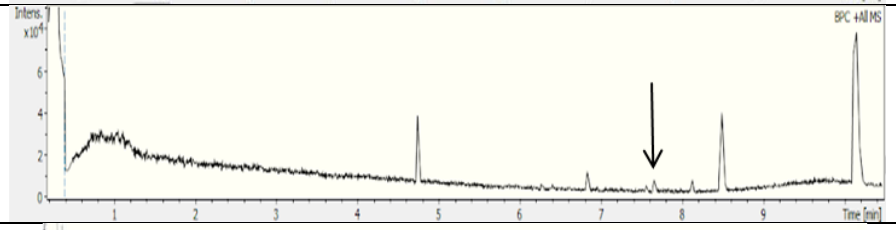
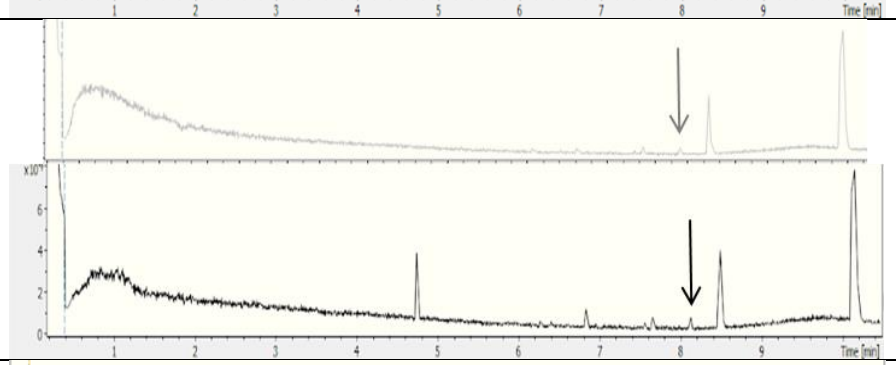
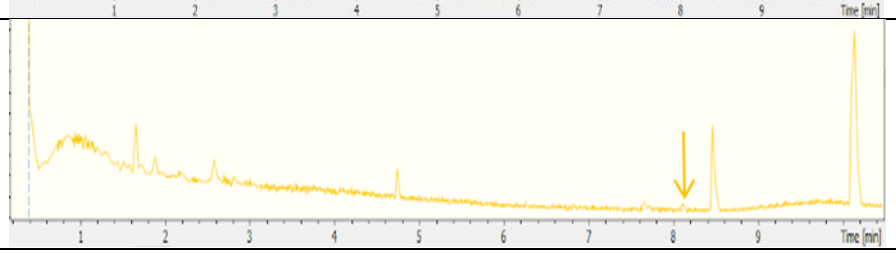
28ºC	MMM	water	1.6	611.2979	$C_{34}H_{38}N_6O_5$ $C_{22}H_{42}N_8O_{12}$	no hit	
28ºC	M400	water	1.6	513.3761	$C_{25}H_{48}N_6O_5$	no hit	
28ºC	M19	water	1.8	627.3282	$C_{22}H_{51}N_4O_{16}$	no hit	
28ºC	M400	water	1.8	625.3142	$C_{14}H_{31}N_{27}O_3$	no hit	

28ºC	MMM SGG	water	1.8 1.9	625.3131 625.3131	$C_{23}H_{44}N_8O_{12}$ $C_{22}H_{48}N_4O_{16}$ $C_{22}H_{48}N_4O_{16}$	no hit no hit no hit	
28ºC	MMM	water	2.1	397.2086	$C_{12}H_{23}N_{13}O_3$	no hit	
28ºC	OM	water	3.0	821.4400	$C_{36}H_{48}N_{22}O_2$	no hit	
28ºC	OM	water	3.9	1225.5876	$C_{59}H_{84}N_8O_{20}$	no hit	

28ºC	MMM	water	7.8	403.2336	$C_{21}H_{30}N_4O_4$	no hit	
28ºC	M19	water	7.9	403.2317	$C_{17}H_{26}N_{10}O_2$	no hit	
28ºC	OM	water	9.1	683.5551	$C_{35}H_{70}N_8O_5$	no hit	
30 ºC	M400	organic	0.8	582.3246	$C_{27}H_{40}N_{11}O_4$	no hit	n/a
30 ºC	SYP	organic	2.8	821.4384	$C_{37}H_{64}N_4O_{16}$	no hit	n/a
30 ºC	M19	organic	6.3	579.2927	$C_{30}H_{38}N_6O_6$	no hit	

30 ºC	M400	organic	8.2	665.5816	$C_{40}H_{76}N_2O_5$	no hit	
30 ºC	MMM	water	1.9	814.4551	$C_{37}H_{63}N_7O_{13}$	no hit	
30 ºC	MMM	water	2.6	920.4609	$C_{43}H_{65}N_7O_{15}$	no hit	
30 ºC	M19	water	3.0	438.3793	$C_{23}H_{51}NO_6$	no hit	

30 ºC	M19	water	3.2	403.2327	$C_{21}H_{30}N_4O_4$	no hit	
30 ºC	M19	water	4.0	384.1947	$C_{23}H_{45}NO_3$ $C_{26}H_{25}NO_2$	no hit no hit	
30 ºC	M19	water	4.3	421.2238	$C_9H_{24}N_{16}O_4$	no hit	
30 ºC	M19	water	4.5	465.2767	$C_{28}H_{36}N_2O_4$	no hit	
30 ºC	M19	water	6.5	663.4521	$C_{24}H_{65}N_5O_{15}$ $C_{44}H_{58}N_2O_3$	no hit no hit	

30 ºC	M19	water	6.9	750.4056	$C_{52}H_{51}N_3O_2$	no hit	
30 ºC	SYP	water	7.6	424.282	$C_{23}H_{39}NO_6$	no hit	
30 ºC	SM SYP	water	8.1	465.2778 465.2778	$C_{33}H_{36}O_2$	no hit	
30 ºC	MMM	water	8.1	465.2757	$C_{20}H_{49}N_7O_5$	no hit	

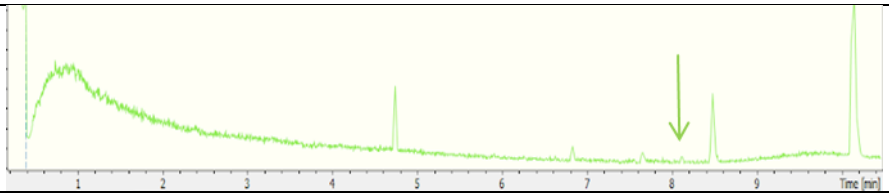
	OM	water	8.1	465.2765	$C_{11}H_{29}N_{17}O_4$	no hit	

Table 5: Potential new compounds detected from LC-MS analysis of the crude extracts samples from SM3

Crude extracts fermented at 28 °C and 30 °C in seven media (M19, M400, MMM, OM, SGG, SM and SYP) from both aqueous and organic phases-. n/a= picture of the chromatogram not available. The metabolites originated from media/temperature controls were excluded from the table to enable visualization of only the metabolites potentially produced by our isolates when cultured in these specific parameters.

The remaining total metabolites from both SM3 and B226SN104 crude extracts, detected in all tested temperatures and media tested, including potential new metabolites are in the supplementary figures 1 to 8 and supplementary tables 1 to 8. It is important to note that the term "potential new metabolite" is only "potential" as the metabolites were compared to a selection of different libraries, only structural characterisation by NMR can ultimately definitely confirm the novelty of those potentially new metabolites.

Overall, a total of 472 metabolites were detected by LC-MS analysis and manual de-replication, including 192 potential new metabolites (Table 4), which allowed us to observe the complexity of chemical diversity in both marine strains. It is well established that the production of secondary metabolites in *Streptomyces* species is regulated at least to some extent by temperature and by nutrient sources including relative concentration of specific nutrient such as a single or combined carbon and phosphorous source or by different carbon and nitrogen sources, (Martin *et al.*, 2011). Thus it is clear that different nutrient sources in both SM3 and B226SN104 affect metabolite production in both SM3 and B226SN104.

This study therefore reveals that the production of secondary metabolites in these two marine *Streptomyces* strains is likely to be complex and that both strains possess the potential to produce both existing and novel bioactive secondary metabolites. In addition to revealing the likely complexity of the regulation of these metabolites in both strains, it is also indicates that OSMAC based approaches can be useful in not only inducing production of metabolites in new isolates but also has merit in being used to re-culture existing isolates that have already been cultures, but not under specific sets of physiological conditions, thereby opening up the possibility that these strains may produce different metabolites and some with potential novelty. This would negate the need to constantly re-collect new strains from new sampling

regimes; suggesting that OSMAC could represent a new way to explore the production of bioactive metabolites. Thus such an approach could help elucidate the full metabolic potential of strains that have already been isolated in many different laboratories, but that are likely to possess the potential to produce more and different types of metabolites, by subjecting these strains to different sets of physiological parameters. Such an approach would of course be both strain dependant, and parameter dependant, as certain changes could result in either a higher or reduced number of metabolites being produced. However not withstanding this, these approaches involving existing strain collections could reduce the need to re-collect new samples, thereby saving time and expenses such as expedition costs, and also reduce barriers that may exist in the future in sample collection in certain oceanic regions. However, it should also be considered that finding new physiological parameters that induce the production of metabolites in specific strains, may also be time consuming.

Conclusion

The growth of our two marine *Streptomyces* strains SM3 and B226SN104 employing an OSMAC based approach, involving seven different culture media and two temperatures conditions, resulted in a complex metabolome being observed in each strain, with a total of 472 metabolites being identified, with 192 of these being potentially new metabolites. Some metabolites were detected in all samples, regardless of the temperature or the media used for the fermentations; while other metabolites was only detected under specific culture conditions. This reveals, the complexity of the chemical diversity of these two marine strains and demonstrates both nutrient source and temperature as important parameters to monitor in studying the production of secondary metabolites and for the discovery of potential new metabolites in these two isolates

The complexity of the metabolomes in each strain and the selective production of some metabolites, suggests the presence of cryptic or silent gene clusters in these strains that are specifically activated under different culture regime-temperature conditions.

Material and Methods

• Origin of strains

The bacterial strains were obtained from the marine sponge *Haliclona simulans* (class Demospongiae, order Haplosclerida, family Chalinidae) for the shallow water strains, at a depth of 15 m in Galway bay (Ireland); and *Inflatella pellicula* (class Demospongiae; order Poecilosclerida; suborder Myxillina; family Coelosphaeridae) for the deep sea sample on the Porcupine bank at a depth of 2900 m (North Atlantic Ocean). These strains were obtained previously, I did not participate their collection.

• DNA extraction

Genomic DNA from both B226SN104 and SM3 was isolated using the phenol-chloroform-isoamyl alcohol extraction method (Wilson, 2001). Isolates were grown overnight (~16 h) in 5 mL liquid cultures. Cells were pelleted by centrifugation (6000 g), supernatants were decanted and discarded. Cell pellets were resuspended in 467 µL TE buffer. 30 µL of 10% SDS and 3 µL of 20 mg/mL Proteinase K (Fermentas, Sankt Leon-Rot, Germany) was added to each tube and incubated for 1 h at 37 °C. An equal volume of phenol/chloroform (phenol-chloroform-isoamyl alcohol mixture ratio 25:24:1, Sigma Aldrich, Arklow, Ireland) was added and mixed well. Tubes were centrifuged, at ~18,042 g for 10 min (Eppendorf Centrifuge, 5417r, Eppendorf UK Ltd., Stevenage, UK). The upper phase of each Eppendorf tube was aspirated to fresh 2 mL Eppendorf tubes, avoiding the interphase. Sodium acetate (NaOAc) 3M (100 µl)

pH 5.2 was added to each tube and mixed well. 600 µL of isopropanol (Sigma Aldrich, Arklow, Ireland) was added, mixed well, and incubated at room temperature for 15 min. Tubes were then centrifuged at ~18,042 g for 20 min. Supernatants were removed and discarded. DNA pellets were washed with cold (4 °C) 70% EtOH. Tubes were centrifuged at ~18,042 g for 10 min. Ethanol was removed and discarded and DNA pellets were allowed to air-dry. DNA was re-suspended in 1 mL TE buffer. 1 µL of RNase A was added to the tubes, which were then incubated at 37 °C for 30 min. DNA was again purified by phenol extraction. DNA was analysed by gel electrophoresis and quantified using a spectrophotometer (NanoDrop ND-1000, Thermo Scientific, Gloucester, UK). The DNA solutions were stored at −20 °C.

- **Whole genome sequencing**

Whole genome sequencing was performed by Macrogen (Seoul, South 1334 Korea), mate pair libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. All libraries were sequenced in 250 bp paired read runs on the Illumina MiSeq platform. Reads were trimmed for quality with Sickle (https://github.com/najoshi/sickle) and Scythe (https://github.com/vsbuffalo/scythe).

- **Genomic and bioinformatics methods**

Genomes were assembled by Dr. Eduardo Almeida using SPAdes, version 3.1.1 and comparative genomic methods were carried out on the following six genomes: B226SN104, SM3, SM6, SM9, SM14 and SM19. Potential bioactive gene clusters were identified by searching the genomes with antiSMASH version 3 and parsed with R methods, R is a free software environment for statistical computing and graphics (https://www.r-project.org/). Identified gene clusters were counted and graphed using Heatmap and R. Values were automatically converted from the normal distribution of the number of gene clusters to the

standard normal distribution by "standardizing" with the use of z-scores. The dendrogram was generated automatically according to the default parameters ("predicted clusters"). The most promising genomes among all the isolates are observed by the blue colours on the map (scale colour from white-pink to blue). Normal distribution values of the total number of gene clusters identified among the six genomes isolates were manually added on the top of the heat map.

- **Production of secondary metabolites by fermentation**

The production of secondary metabolites was performed following the OSMAC approach with fermentation of the strains in different nutrient-sources media and different temperature conditions. The composition of the fermentation media is provided in Table 2. The *Streptomyces* samples were first streaked on a plate and incubated at 28 °C for 3 days to allow for sporulation, then a single colony of each strain was individually selected to be cultivated in 250mL Erlenmeyer flasks, containing 50 mL of rich media M410. They were incubated for 7 days at 28°C, in a rotary shaker incubator at 200 rpm. 1mL of each culture media was inoculated to 250mL flasks containing 49 mL of each respective media (OM/ MMM/ SGG/ M400/ M19/ SYP/ SM) and incubated for 14 days at 28 °C, 200rpm. In addition, 1 mL of the culture M410 media was used for the analysis of anti-microbial activity. The fermented culture samples were centrifuged at 4 °C, max speed for 10-15 min. The pellet was freeze-dried at-80°C and the cell free supernatant (50 mL) stored at -80 °C or used immediately for extraction. Identical methods were employed for growth of the strains at 30 °C. See Table 2 for exact details about the media used for fermentation. Control media of each media were cultured in the same conditions than the isolates. Due to the high number of samples, no replicates were performed at this stage.

- **Antimicrobial activities screening - Well diffusion assays**

The isolated strains were screened for antimicrobial activity using the well diffusion assay method against different test strains such as *Staphylococcus aureus* (NCIMB 9518), *Pseudomonas aeruginosa* (PAO1), *Bacillus subtilis* (IE32), *Escherichia coli* (NCIMB 12210), *Candida glabrata* (CBS138), and *Candida albicans* (Sc5314). YPD (Yeast Extract-Peptone-Dextrose) and LB (Luria-Bertani) agar plates were prepared and streaked with a single strain of each test strains at an OD=0.1, then wells were manually made in each plate by using sterile tips. A volume of 75uL of the cell free supernatant from the M410-strain cultures broth at day 7 of fermentation, were poured into each well. Plates were then incubated overnight at 37 °C for the bacterial plates and 28 °C for the fungal plates. Both anti-bacterial and anti-fungal activities, were deduced by observation of growth inhibition around the well. To confirm the validity of the results, in parallel, negative controls were performed by replacing the volume of 75 µL of culture by 75 µL of sterile water and the results were validated by the absence of growth inhibition surrounding the wells.

- **Liquid Liquid Extraction (LLE)**

Cell-free supernatants (50 mL) were chemically extracted using Liquid Liquid Extraction (LLE) with ethyl-acetate as solvent. The solvent was added to the filtrate in the ratio of 1:1(v/v), shaken vigorously and left to stand for 5 min to obtain the organic and aqueous phases. The organic phase was separated from the aqueous phase using a separating funnel. Both layers were concentrated by evaporating to dryness at 45 °C in a GeneVac to generate dry crude extracts.

- **Liquid Chromatography - Mass Spectrometry (LC-MS)**

LC-MS was conducted on all the crude extracts on a Bruker MAXIS II QTOF. The Bruker MAXIS II QTOF was run in tandem with an Agilent 1260 HPLC using a Phenomenex Kinetex 2.6 mM XB-C18, 100 x 2.1 mm column. All samples were run at a concentration of 0.5 mg/ml using a gradient of 95 % Water / 5 % Acetonitrile – 100 % Acetonitrile.

- **Manual de-replication**

Manual de-replication of the metabolites from the LC-MS data was obtained by deduction of the possible molecular weights and molecular formulas predicted from the Bruker software using Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). For the interpretation of the MS data a number of different libraries were used, such as Chemspider, Antibase, Reaxys and The dictionary of Natural Products. The potential new metabolites were deduced by the absence of any hits following searched on these databases. A second manual de-replication of the identified metabolites was obtained by matching the LC-MS raw data with the spectral data from the Molecular Network.

- **Metabolomic analysis and Visualisation of Molecular Network**

The raw data generated from the LC-MS was converted into mzXLM file format by using the downloaded zip file GNPS_Vendor_Conversion obtained through the DataAnalysis software from the online portal of Global Natural Products Social Molecular Networking GNPS (https://gnps.ucsd.edu). All mzXLM files were larger than 20MB so they were uploaded to ProteoSAFe (http://proteomics.ucsd.edu/ProteoSAFe/) using FileZilla as an ftp client. When all the files were uploaded to ProteoSAFe they became available within the GNPS system. Separately, a metadata table composed of attributes (metadata) describing the samples properties was created by following the GNPS requirements workflow, in a text file format, and then used as an input to the molecular networking of GNPS. Molecular networking was performed using the online data analysis portal of Global Natural Products Social Molecular

Networking. The data were filtered by removing all MS/MS peaks within ±17 Da of the precursor *m/z*. The MS/MS spectra were window filtered by choosing only the top 6 peaks in the ±50 Da window throughout the spectrum. The data was then clustered with MS-Cluster with a parent mass tolerance of 0.02 Da and an MS/MS fragment ion tolerance of 0.02 Da to create consensus spectra. Consensus spectra that contained less than 2 spectra were discarded. A network was then created, where edges were filtered to have a cosine score of above 0.7 and more than 6 matched peaks. Further edges between two nodes were retained in the network only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. All matches that were kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. The molecular network was visualized using Cytoscape version 3.7.1 and displayed using the settings "Organic layout" with "sample 2" style. The node (metabolite) originated from media controls were excluded from the network to enable visualization of only the metabolites potentially produced by our isolates in these specific parameters. To visualize each different media from all samples, regardless of the organics and water phases, color-codes was established manually for each media via the group columns from "pie" in "chart" in the Node window.

- **Colour-code**

A colour code was applied to all the different media used for production of secondary metabolites to facilitate ease of visual representation. The same colour code was used for both Figures 3 and 6 as follows: M19 red, M400 blue, MMM yellow, OM green, SGG purple, SM grey and SYP black.

References

Almeida, E. L., Carrillo Rincon, A. F., Jackson, S. A., & Dobson, A. D. W. (2019). Corrigendum: Comparative Genomics of Marine Sponge-Derived *Streptomyces* spp. Isolates SM17 and SM18 With Their Closest Terrestrial Relatives Provides Novel Insights Into Environmental Niche Adaptations and Secondary Metabolite Biosynthesis Potential. *Front Microbiol, 10*, 2213. doi:10.3389/fmicb.2019.02213

Almeida, E. L., Carrillo Rincon, A. F., Jackson, S. A., & Dobson, A. D. W. (2019). In silico Screening and Heterologous Expression of a Polyethylene Terephthalate Hydrolase (PETase)-Like Enzyme (SM14est) With Polycaprolactone (PCL)-Degrading Activity, From the Marine Sponge-Derived Strain *Streptomyces* sp. SM14. *Front Microbiol, 10*, 2187. doi:10.3389/fmicb.2019.02187

Almeida, E. L., Kaur, N., Jennings, L. K., Carrillo Rincon, A. F., Jackson, S. A., Thomas, O. P., & Dobson, A. D. W. (2019). Genome Mining Coupled with OSMAC-Based Cultivation Reveal Differential Production of Surugamide A by the Marine Sponge Isolate *Streptomyces* sp. SM17 When Compared to Its Terrestrial Relative S. albidoflavus J1074. *Microorganisms, 7*(10). doi:10.3390/microorganisms7100394

Antoraz, S., Santamaria, R. I., Diaz, M., Sanz, D., & Rodriguez, H. (2015). Toward a new focus in antibiotic and drug discovery from the *Streptomyces* arsenal. *Front Microbiol, 6*, 461. doi:10.3389/fmicb.2015.00461

Awakawa, T., Yang, X. L., Wakimoto, T., & Abe, I. (2013). Pyranonigrin E: a PKS-NRPS

hybrid metabolite from Aspergillus niger identified by genome mining. *Chembiochem, 14*(16), 2095-2099. doi:10.1002/cbic.201300430

Bender, T., Schuhmann, T., Magull, J., Grond, S., & von Zezschwitz, P. (2006). Comprehensive study of okaspirodiol: characterization, total synthesis, and biosynthesis of a new metabolite from *Streptomyces*. *J Org Chem, 71*(19), 7125-7132. doi:10.1021/jo060149e

Borchert, E., Jackson, S. A., O'Gara, F., & Dobson, A. D. (2016). Diversity of Natural Product Biosynthetic Genes in the Microbiome of the Deep Sea Sponges *Inflatella pellicula, Poecillastra compressa,* and *Stelletta normani*. *Front Microbiol, 7*, 1027. doi:10.3389/fmicb.2016.01027

Callow, R. K., & Taylor, D. A. H. (1952). 429. The cardio-active glycosides of *Strophanthus sarmentosus* P.DC. "sarmentoside B" and its relation to an original sarmentobioside. *Journal of the Chemical Society (Resumed)*(0), 2299-2304. doi:10.1039/JR9520002299

Cleary, J. L., Kolachina, S., Wolfe, B. E., & Sanchez, L. M. (2018). Coproporphyrin III Produced by the Bacterium *Glutamicibacter arilaitensis* Binds Zinc and Is Upregulated by Fungi in Cheese Rinds. *mSystems, 3*(4). doi:10.1128/mSystems.00036-18

Codd, R., Soe, C. Z., Pakchung, A. A. H., Sresutharsan, A., Brown, C. J. M., & Tieu, W. (2018). The chemical biology and coordination chemistry of putrebactin, avaroferrin,

bisucaberin, and alcaligin. *J Biol Inorg Chem, 23*(7), 969-982. doi:10.1007/s00775-018-1585-1

Darabpour, E., Ardakani, M. R., Motamedi, H., Ronagh, M. T., & Najafzadeh, H. (2012). Purification and optimization of production conditions of a marine-derived antibiotic and ultra-structural study on the effect of this antibiotic against MRSA. *European Review for Medical and Pharmacological Sciences, 16*(2), 157-165.

Ding, Y., de Wet, J. R., Cavalcoli, J., Li, S., Greshock, T. J., Miller, K. A., Finefield, J. M., Sunderhaus, J. D., McAfoos, T. J., Tsukamoto, S., Williams, R. M., Sherman, D. H. (2010). Genome-based characterization of two prenylation steps in the assembly of the stephacidin and notoamide anticancer agents in a marine-derived Aspergillus sp. *J Am Chem Soc, 132*(36), 12733-12740. doi:10.1021/ja1049302

Doull, J. L., Singh, A. K., Hoare, M., & Ayer, S. W. (1994). Conditions for the production of jadomycin B by *Streptomyces venezuelae* ISP5230: effects of heat shock, ethanol treatment and phage infection. *J Ind Microbiol, 13*(2), 120-125. doi:10.1007/BF01584109

Fan, B., Parrot, D., Blumel, M., Labes, A., & Tasdemir, D. (2019). Influence of OSMAC-Based Cultivation in Metabolome and Anticancer Activity of Fungi Associated with the Brown Alga *Fucus vesiculosus*. *Marine Drugs, 17*(1). doi:10.3390/md17010067

Fdhila, F., Vazquez, V., Sanchez, J. L., & Riguera, R. (2003). dd-diketopiperazines: antibiotics active against *Vibrio anguillarum* isolated from marine bacteria associated

with cultures of *Pecten maximus*. *J Nat Prod, 66*(10), 1299-1301. doi:10.1021/np030233e

Finefield, J. M., & Williams, R. M. (2010). Synthesis of Notoamide J: A Potentially Pivotal Intermediate in the Biosynthesis of Several Prenylated Indole Alkaloids. *Journal of Organic Chemistry, 75*(9), 2785-2789. doi:10.1021/jo100332c

Goodfellow, M., & Fiedler, H. P. (2010). A guide to successful bioprospecting: informed by actinobacterial systematics. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology, 98*(2), 119-142. doi:10.1007/s10482-010-9460-2

Gutierrez, M., Pereira, A. R., Debonsi, H. M., Ligresti, A., Di Marzo, V., & Gerwick, W. H. (2011). Cannabinomimetic lipid from a marine cyanobacterium. *J Nat Prod, 74*(10), 2313-2317. doi:10.1021/np200610t

Hagiwara, D., Sakai, K., Suzuki, S., Umemura, M., Nogawa, T., Kato, N., Osada, H., Watanabe, A., Kawamoto, S., Gonoi, T., Kamei, K. (2017). Temperature during conidiation affects stress tolerance, pigmentation, and trypacidin accumulation in the conidia of the airborne pathogen Aspergillus fumigatus. *PLoS One, 12*(5), e0177050. doi:10.1371/journal.pone.0177050

Hayashi, K., Inoguchi, M., Kondo, H., & Nozaki, H. (2000). Gibbestatin B inhibits the GA-induced expression of alpha-amylase expression in cereal seeds. *Phytochemistry, 55*(1), 1-9.

He, M., Min, J. W., Kong, W. L., He, X. H., Li, J. X., & Peng, B. W. (2016). A review on the pharmacological effects of vitexin and isovitexin. *Fitoterapia, 115*, 74-85. doi:10.1016/j.fitote.2016.09.011

Ishibashi, M., Funayama, S., Anraku, Y., Komiyama, K., & Omura, S. (1991). Novel antibiotics, furaquinocins C, D, E, F, G and H. *J Antibiot (Tokyo), 44*(4), 390-395.

Jackson, S. A., Crossman, L., Almeida, E. L., Margassery, L. M., Kennedy, J., & Dobson, A. D. (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene Clusters in the Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates. *Marine Drugs, 16*(2). doi:ARTN 6710.3390/md16020067

Jang, J. P., Jang, J. H., Soung, N. K., Kim, H. M., Jeong, S. J., Asami, Y., . . . Ahn, J. S. (2012). Benzomalvin E, an indoleamine 2,3-dioxygenase inhibitor isolated from Penicillium sp. FN070315. *J Antibiot (Tokyo), 65*(4), 215-217. doi:10.1038/ja.2011.141

Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., Bourke, M. B., Rea, M. C., O'Connor, P. M., Ross, R. P., Hill, C., O'Gara, F., Marchesi, J. R., Dobson, A. D. (2009). Isolation and analysis of bacteria with antimicrobial activities from the marine sponge *Haliclona simulans* collected from Irish waters. *Mar Biotechnol (NY), 11*(3), 384-396. doi:10.1007/s10126-008-9154-1

Kumano, T., Tomita, T., Nishiyama, M., & Kuzuyama, T. (2010). Functional characterization

of the promiscuous prenyltransferase responsible for furaquinocin biosynthesis: identification of a physiological polyketide substrate and its prenylated reaction products. *J Biol Chem, 285*(51), 39663-39671. doi:10.1074/jbc.M110.153957

Leach, J. L., Garber, S. A., Marcon, A. A., & Prieto, P. A. (2005). In vitro and *in vivo* effects of soluble, monovalent globotriose on bacterial attachment and colonization. *Antimicrob Agents Chemother, 49*(9), 3842-3846. doi:10.1128/AAC.49.9.3842-3846.2005

Ledyard, K. M., & Butler, A. (1997). Structure of putrebactin, a new dihydroxamate siderophore produced by Shewanella putrefaciens. *Journal of Biological Inorganic Chemistry, 2*(1), 93-97. doi:DOI 10.1007/s007750050110

Lee, M. H., Kataoka, T., Honjo, N., Magae, J., & Nagai, K. (2000). *In vivo* rapid reduction of alloantigen-activated CD8+ mature cytotoxic T cells by inhibitors of acidification of intracellular organelles, prodigiosin 25-C and concanamycin B. *Immunology, 99*(2), 243-248. doi:10.1046/j.1365-2567.2000.00961.x

Lind, A. L., Smith, T. D., Saterlee, T., Calvo, A. M., & Rokas, A. (2016). Regulation of Secondary Metabolism by the Velvet Complex Is Temperature-Responsive in Aspergillus. *G3 (Bethesda), 6*(12), 4023-4033. doi:10.1534/g3.116.033084

Long, S., Resende, D., Kijjoa, A., Silva, A. M. S., Fernandes, R., Xavier, C. P. R. Vasconcelos, M. H., Sousa, E., Pinto, M. M. M. (2019). Synthesis of New Proteomimetic Quinazolinone Alkaloids and Evaluation of Their Neuroprotective and

Antitumor Effects. *Molecules, 24*(3). doi:10.3390/molecules24030534

Managamuri, U., Vijayalakshmi, M., Ganduri, V., Rajulapati, S. B., Bonigala, B., Kalyani, B. S., & Poda, S. (2017). Isolation, identification, optimization, and metabolite profiling of *Streptomyces* sparsus VSM-30. *3 Biotech, 7*(3), 217. doi:10.1007/s13205-017-0835-1

Martin, J. F., Sola-Landa, A., Santos-Beneit, F., Fernandez-Martinez, L. T., Prieto, C., & Rodriguez-Garcia, A. (2011). Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in *Streptomyces*. *Microb Biotechnol, 4*(2), 165-174. doi:10.1111/j.1751-7915.2010.00235.x

Milne, P. J., Hunt, A. L., Rostoll, K., Van Der Walt, J. J., & Graz, C. J. (1998). The biological activity of selected cyclic dipeptides. *J Pharm Pharmacol, 50*(12), 1331-1337. doi:10.1111/j.2042-7158.1998.tb03355.x

Nazari, L., Manstretta, V., & Rossi, V. (2016). A non-linear model for temperature-dependent sporulation and T-2 and HT-2 production of *Fusarium langsethiae* and *Fusarium sporotrichioides*. *Fungal Biol, 120*(4), 562-571. doi:10.1016/j.funbio.2016.01.010

Niu, S., Li, S., Chen, Y., Tian, X., Zhang, H., Zhang, G., Zhang, W., Yang, X., Zhang, S., Ju, J., Zhang, C. (2011). Lobophorins E and F, new spirotetronate antibiotics from a South China Sea-derived *Streptomyces* sp. SCSIO 01127. *J Antibiot (Tokyo), 64*(11), 711-716. doi:10.1038/ja.2011.78

Oppong-Danquah, E., Parrot, D., Blumel, M., Labes, A., & Tasdemir, D. (2018). Molecular Networking-Based Metabolome and Bioactivity Analyses of Marine-Adapted Fungi Co-cultivated With Phytopathogens. *Front Microbiol, 9*, 2072. doi:10.3389/fmicb.2018.02072

Pan, R., Bai, X., Chen, J., Zhang, H., & Wang, H. (2019). Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review. *Front Microbiol, 10*, 294. doi:10.3389/fmicb.2019.00294

Pan, Y., & Jackson, R. T. (2009). Dietary phylloquinone intakes and metabolic syndrome in US young adults. *J Am Coll Nutr*, *28*(4), 369-379. doi:10.1080/07315724.2009.10718099

Petersen, L.E., Kellermann, M..Y., Schupp, P.J. (2020) Secondary Metabolites of Marine Microbes: From Natural Products Chemistry to Chemical Ecology. In: Jungblut, S., Liebich, V., Bode-Dalby, M. (eds) YOUMARES 9 - The Oceans: Our Research, Our Future. Springer, Cham. https://doi.org/10.1007/978-3-030-20389-4_8.

Reen, F. J., Gutierrez-Barranquero, J. A., Dobson, A. D., Adams, C., & O'Gara, F. (2015). Emerging concepts promising new horizons for marine biodiscovery and synthetic biology. *Marine Drugs, 13*(5), 2924-2954. doi:10.3390/md13052924

Rhee, K. H. (2002). Isolation and characterization of *Streptomyces* sp KH-614 producing anti-VRE (vancomycin-resistant enterococci) antibiotics. *J Gen Appl Microbiol,*

48(6), 321-327.

Romano, S., Jackson, S. A., Patry, S., & Dobson, A. D. W. (2018). Extending the "One Strain Many Compounds" (OSMAC) Principle to Marine Microorganisms. *Marine Drugs, 16*(7). doi:10.3390/md16070244

Saha, M., Ghosh, D., Jr., Ghosh, D., Garai, D., Jaisankar, P., Sarkar, K. K., Dutta, P. K., Das, S., Jha, T., Mukherjee, J. (2005). Studies on the production and purification of an antimicrobial compound and taxonomy of the producer isolated from the marine environment of the Sundarbans. *Appl Microbiol Biotechnol, 66*(5), 497-505. doi:10.1007/s00253-004-1706-3

Schneemann, I., Nagel, K., Kajahn, I., Labes, A., Wiese, J., & Imhoff, J. F. (2010). Comprehensive investigation of marine Actinobacteria associated with the sponge Halichondria panicea. *Appl Environ Microbiol, 76*(11), 3702-3714. doi:10.1128/AEM.00780-10

Sharma, R., Jamwal, V., Singh, V. P., Wazir, P., Awasthi, P., Singh, D., Vishwakarma, R. A., Gandhi, S. G., Chaubey, A. (2017). Revelation and cloning of valinomycin synthetase genes in *Streptomyces lavendulae* ACR-DA1 and their expression analysis under different fermentation and elicitation conditions. *J Biotechnol, 253*, 40-47. doi:10.1016/j.jbiotec.2017.05.008

Song, Y., Li, Q., Liu, X., Chen, Y., Zhang, Y., Sun, A., Zhang, W., Zhang, J., Ju, J. (2014). Cyclic Hexapeptides from the Deep South China Sea-Derived *Streptomyces*

scopuliridis SCSIO ZJ46 Active Against Pathogenic Gram-Positive Bacteria. *J Nat Prod, 77*(8), 1937-1941. doi:10.1021/np500399v

Sproule, A., Correa, H., Decken, A., Haltli, B., Berrue, F., Overy, D. P., & Kerr, R. G. (2019). Terrosamycins A and B, Bioactive Polyether Ionophores from *Streptomyces* sp. RKND004 from Prince Edward Island Sediment. *Marine Drugs, 17*(6). doi:10.3390/md17060347

Sujatha, P., Bapi Raju, K. V., & Ramana, T. (2005). Studies on a new marine streptomycete BT-408 producing polyketide antibiotic SBR-22 effective against methicillin resistant *Staphylococcus aureus*. *Microbiol Res, 160*(2), 119-126. doi:10.1016/j.micres.2004.10.006

Tani, H., Koshino, H., Sakuno, E., Cutler, H. G., & Nakajima, H. (2006). Botcinins E and F and Botcinolide from Botrytis cinerea and structural revision of botcinolides. *J Nat Prod, 69*(4), 722-725. doi:10.1021/np060071x

Tapondjou, L. A., Ponou, K. B., Teponno, R. B., Mbiantcha, M., Djoukeng, J. D., Nguelefack, T. B., Watcho, P., Cadenas, A. G., Park, H. J. (2008). *In vivo* anti-inflammatory effect of a new steroidal saponin, mannioside A, and its derivatives isolated from *Dracaena mannii*. *Arch Pharm Res, 31*(5), 653-658. doi:10.1007/s12272-001-1208-3

Thomson, N. R., Crow, M. A., McGowan, S. J., Cox, A., & Salmond, G. P. (2000). Biosynthesis of carbapenem antibiotic and prodigiosin pigment in Serratia is under

quorum sensing control. *Mol Microbiol, 36*(3), 539-556. doi:10.1046/j.1365-2958.2000.01872.x

Tian, Z. H., Cheng, Q., Yoshimoto, F. K., Lei, L., Lamb, D. C., & Guengerich, F. P. (2013). Cytochrome P450 107U1 is required for sporulation and antibiotic production in *Streptomyces coelicolor*. *Archives of Biochemistry and Biophysics, 530*(2), 101-107. doi:10.1016/j.abb.2013.01.001

Tsukamoto, S., Kato, H., Samizo, M., Nojiri, Y., Onuki, H., Hirota, H., & Ohta, T. (2008). Notoamides F-K, prenylated indole alkaloids isolated from a marine-derived Aspergillus sp. *J Nat Prod, 71*(12), 2064-2067. doi:10.1021/np800471y

Wang, W. J., Li, D. Y., Li, Y. C., Hua, H. M., Ma, E. L., & Li, Z. L. (2014). Caryophyllene sesquiterpenes from the marine-derived fungus Ascotricha sp. ZJ-M-5 by the one strain-many compounds strategy. *J Nat Prod, 77*(6), 1367-1371. doi:10.1021/np500110z

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crusemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderon, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A.

R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., P, C. A. B., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodriguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. F., Alexandrov, T., Litaudon, M., Wolfender, J. L. Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Muller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B. O., Pogliano, K., Linington, R. G., Gutierrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C., Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol, 34*(8), 828-837. doi:10.1038/nbt.3597

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. (2012). Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A, 109*(26), E1743-1752. doi:10.1073/pnas.1203689109

Weller, M. G. (2012). A unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques. *Sensors (Basel), 12*(7), 9181-9209.

doi:10.3390/s120709181

Wietz, M., Mansson, M., & Gram, L. (2011). Chitin stimulates production of the antibiotic andrimid in a *Vibrio coralliilyticus* strain. *Environ Microbiol Rep, 3*(5), 559-564. doi:10.1111/j.1758-2229.2011.00259.x

Wilson K. (2001). Preparation of genomic DNA from bacteria. *Current protocols in molecular biology*, *Chapter 2*, . https://doi.org/10.1002/0471142727.mb0204s56

Wong, S. M., Musza, L. L., Kydd, G. C., Kullnig, R., Gillum, A. M., & Cooper, R. (1993). Fiscalins: new substance P inhibitors produced by the fungus *Neosartorya fischeri.* Taxonomy, fermentation, structures, and biological properties. *J Antibiot (Tokyo), 46*(4), 545-553.

Woo, J. T., Ohba, Y., Tagami, K., Sumitani, K., Yamaguchi, K., & Tsuji, T. (1996). Concanamycin B, a vacuolar H(+)-ATPase specific inhibitor suppresses bone resorption in vitro. *Biol Pharm Bull, 19*(2), 297-299.

Wu, C. Y., Hsu, C. C., Chen, S. T., & Tsai, Y. C. (2001). Soyasaponin I, a potent and specific sialyltransferase inhibitor. *Biochem Biophys Res Commun, 284*(2), 466-469. doi:10.1006/bbrc.2001.5002

Wu, H., Qu, S., Lu, C., Zheng, H., Zhou, X., Bai, L., & Deng, Z. (2012). Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in *Streptomyces hygroscopicus* 5008. *BMC Genomics, 13*, 337. doi:10.1186/1471-2164-

13-337

Yan, P. S., Song, Y., Sakuno, E., Nakajima, H., Nakagawa, H., & Yabe, K. (2004). Cyclo(L-leucyl-L-prolyl) produced by Achromobacter xylosoxidans inhibits aflatoxin production by *Aspergillus parasiticus*. *Appl Environ Microbiol, 70*(12), 7466-7473. doi:10.1128/AEM.70.12.7466-7473.2004

Ying, Y. M., Huang, L., Tian, T., Li, C. Y., Wang, S. L., Ma, L. F., Shan, W. G., Wang, J. W., Zhan, Z. J. (2018). Studies on the Chemical Diversities of Secondary Metabolites Produced by *Neosartorya fischeri* via the OSMAC Method. *Molecules, 23*(11). doi:10.3390/molecules23112772

Zhou, Y. X., Zhang, H., & Peng, C. (2014). Puerarin: a review of pharmacological effects. *Phytother Res, 28*(7), 961-975. doi:10.1002/ptr.5083

Zotchev, S. B. (2012). Marine actinomycetes as an emerging resource for the drug development pipelines. *J Biotechnol, 158*(4), 168-175. doi:10.1016/j.jbiotec.2011.06.002

3 Chapter 3: Genome mining, together with metabolomic and molecular networking approaches to elucidate metabolite production in the marine sponge isolate *Streptomyces* SM9.

Abstract

Microbes associated with marine sponges are exposed to highly competitive environments both physiologically and nutritionally which is likely to promote the production of novel secondary metabolites. This study focuses on the detection and identification of potential bioactive compounds from the *Streptomyces* isolate SM9, which following functional screening for anti-bacterial and anti-fungal activities displayed bioactivity against *Candida glabrata* (CBS138) and *Candida albicans* (Sc5314). Following genome scanning of SM9, approximately 56 potential biosynthetic gene clusters were identified. In addition the OSMAC (One Strain MAny Compounds) approach was employed on the strain to determine the optimal media for the production of secondary metabolites including production of potential new compounds. Cell free supernatants were chemically extracted and crude extracts containing potential secondary metabolites from both aqueous and organic phases were analyzed for their metabolome content. LC-MS, GNPS and Cytoscape were employed in the chemical prediction of the molecules involved, the molecular network and the visualization of the secondary metabolites within the chemical diversity of the strain respectively. This resulted in the identification of 140 metabolites including a number of interesting compounds such as Bisucaberin B, Maculosin, Bpagaome, glu-ala-pro-pro, Desferrioxamine B, pro-thr-ala-ile, Desferrioxamine E, Concanamycin B, Lipoamicoumacin B, Probestin, Bhimamycin B, Surugamide A, Fiscalin B, Gibbestatin B, Antimycin A_1 amongst others; together with a potential 60 new metabolites.

The gene clusters encoding some of the metabolites such as Deferrioxamine B, Deferrioxamine E, Antimycin A_1, Surugamide A were located on the SM9 genome by linking metabolomics analysis and genome mining. In addition siderophore and NRPS sub-networks were also identified. Taken together, these results indicate that SM9 is a good producer of potential new bioactive metabolite, the identity of which will require further characterisation using NMR.

Introduction

In marine environments, bacteria are exposed to quite unique environmental conditions such as high pressure, low temperature, dissolved oxygen and nutrient availability, which often result in the production of structurally diverse and biologically important metabolites (Petersen *et al*., 2020). Marine *Streptomyces* species have in particular being shown to produce a wide variety of different bioactive metabolites, which are rich in biological activity. For example marine *Streptomyces* derived metabolites have been reported to exhibit anticancer, antifungal, cytotoxic and antibacterial activity (Niu *et al*., 2011; Zotchev, 2012; Song *et al*., 2014; Jackson *et al*., 2018).

The One Strain MAny Metabolites (OSMAC) principal involves growing one strain under a variety of different environmental conditions, which typically results in the production of many different types of metabolites. This may ultimately lead to the potential discovery of new bioactive metabolites such as for example antibiotics (Antoraz *et al*. 2015; Romano *et al*., 2018; Pan *et al*., 2019). Parameters such as changes in nutrient sources, involving the modification of both carbon and nitrogen sources in the culture media are known to be particularly useful in altering the metabolite production profile in bacterial strains and in promoting or stimulating the production of secondary metabolites (*Sproule et al., 2019; Ying et al., 2018;* Reen *et al.,*

2015; Pan *et al*., 2019). In addition, nutrient sources are known to regulate secondary metabolism in *Streptomyces* strains (Martin *et al*., 2011) with for example antibiotic SBR-22 production in the marine sediment isolate *Streptomyces psammoticus* BT-408 being reported to be regulated by different nutrient sources.

In the previous chapter, the One Strain MAny Metabolites (OSMAC) approach was employed on the two isolates *Streptomyces* sp. SM3 and *Streptomyces* sp. B226SN104, by monitoring changes in both temperatures and nutrient sources parameters. In this study, a similar OSMAC approach from the previous experiment was employed to investigate the effect the same seven different culture conditions with differing carbon and nitrogen nutrient sources, namely M19, M400, MMM, OM, SGG, SM and SYP; on secondary metabolite production in the marine *Streptomyces* sp. SM9 which had previously been shown to possess anti-fungal activities, and which had been isolated from the sponge *Haliclona simulans* (Kennedy *et al.*, 2009).

The effect of culturing SM9 in fermentations involving the aforementioned different nutrient media sources was investigated, for the production of bioactive metabolites from a microbiology, bioinformatics, genomics, metabolomics and chemical perspective. The overall aim was to link specific fermentation conditions to the production of specific metabolites; to link metabolites to Biosynthetic gene clusters (BGCs) on the SM9 genome; and to therefore link potential metabolite production to these clusters, and to culture regimes.

Results and Discussion

We had previously reported on the isolation and characterization of 18 marine sponge derived *Streptomyces* strains, which displayed a range of antimicrobial and antifungal activities against a number of clinical pathogens (Kennedy *et al.*, 2009; Jackson *et al.*, 2018) as well as a large scale study employing an OSMAC approach on two of these marine *Streptomyces* strains, namely B226SN104 and SM3 detailing the optimal fermentation conditions for production of

metabolites and potential new compounds which was reported in Chapter 1. In this study we reproduced a similar set of fermentation conditions to those used in Chapter 1 by growing SM9 at 28°C on the seven nutrient source media (M19, M400, MMM, OM, SGG, SM and SYP) previously used for the production of metabolites in *Streptomyces* SM3 and *Streptomyces* B226SN104 [Chapter 1] .

3.1 Bioactivity and detection of metabolites

3.1.1 Bioactivity screening

During this fermentation process, to confirm that SM9 isolate was producing bioactivity when cultured under these growth conditions, the isolate was screened for antimicrobial activity against strains of relevant clinical pathogens, when grown in M410 media both in the presence and absence of synthetic sea salt (Instant Ocean) (+/- sw). When grown on this medium the shallow water isolate SM9 displayed bioactivity against both *C. glabrata* (CBS138) and *C. albicans* (Sc5314), in well diffusion assays (Table 6). No difference in bioactivity was observed when SM9 was grown in M410 media with or without salt presence when tested against *C. glabrata* (CBS138), but interestingly, when tested against *C. albicans* (Sc5314), there was a visible increase in bioactivity when SM9 was grown in M410 without salt presence compared to when SM9 was grown in M410 in the presence of salt (Table1). This could be due to the fact that SM9 was originally isolated from a marine sponge in an environment with natural salt, and that the removal of salt from the culture media, may have induced osmotic stress, resulting in an increase in bioactivity. Although this increase was only visible when SM9 was tested against *C. albicans* (Sc5314).

Isolate ID	Bioactivities	Against relevant pathogenic species	Picture
SM9	Anti-fungal activities	*C. glabrata (CBS138).* *C. albicans* (Sc5314).	
Controls	No anti-fungal activity	*C. glabrata (CBS138).* *C. albicans* (Sc5314).	

Table 6: Bioactivity of the shallow water isolate SM9 tested by well diffusion assays against the relevant pathogenic species: *Staphylococcus aureus* (NCIMB 9518), *Pseudomonas aeruginosa* (PAO1), *Bacillus subtilis* (IE32), *Escherichia coli* (NCIMB 12210), *Candida glabrata* (CBS138), and *Candida albicans* (Sc5314).

+sw = with synthetic sea salt(Instant Ocean); sw = without artificial salt. Controls were fermented in the same condition as for the SM9 assay, with the exception of no inoculum was added to the M410 media.

3.1.2 Detection of metabolites – chemical results

Figure 16 represents the total manually de-replicated metabolites detected from all the organic and aqueous phase extracts following analysis by LC-MS and manual de-replication, when SM9 was cultured at 28°C, in seven different growth media, namely M19, M400, MMM, OM, SGG, SM and SYP.

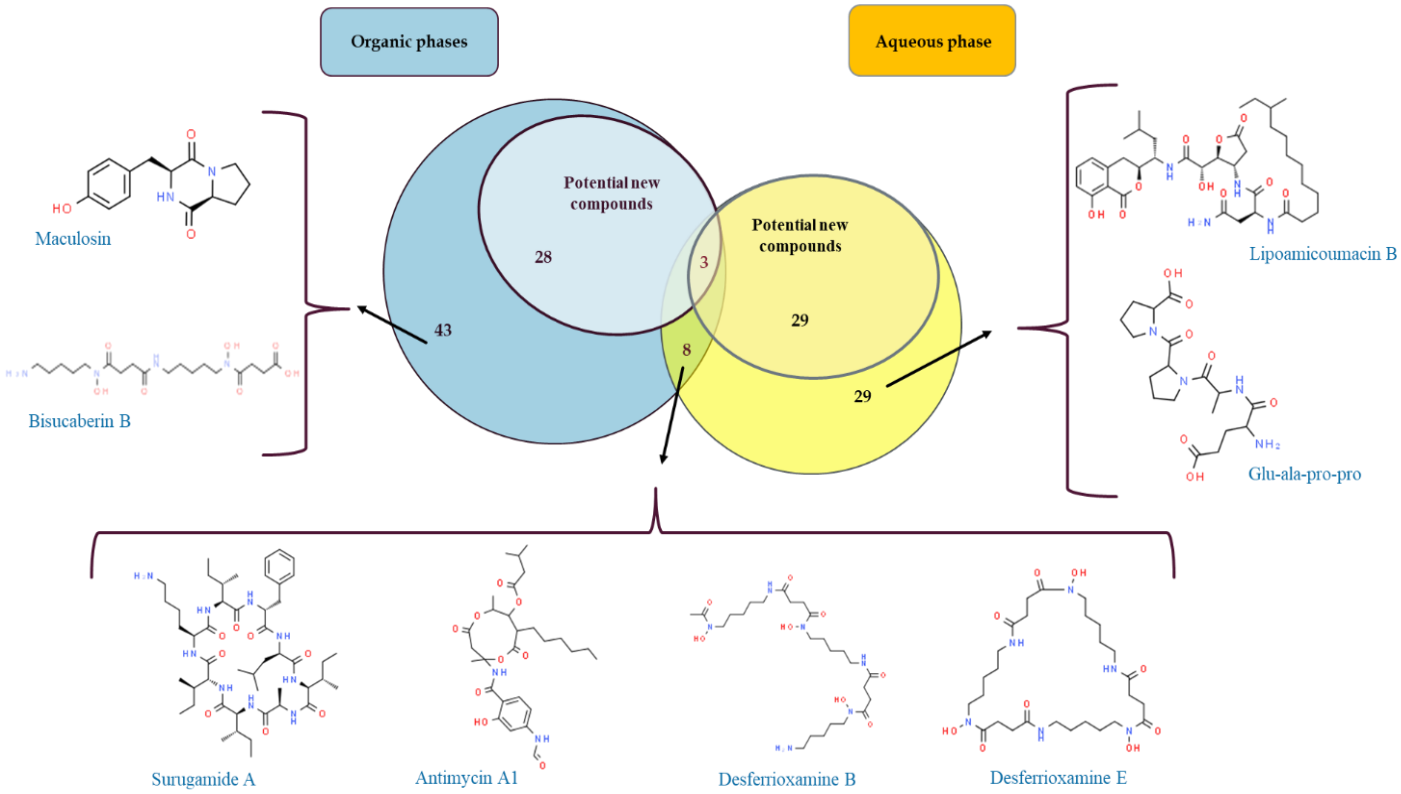


Figure 16: Venn diagram of the specific and shared metabolites, including potential new metabolites, produced by SM9 strain, fermented in 7 media at 28°C, detected chemically in crude extracts from both organic and aqueous phase, by LC-MS and manual de-replication.

Picture of the metabolites were generated with Chemspider.

Following growth liquid liquid extractions (LLE) were then performed to obtain both aqueous and organic phase extracts from the SM9 cultures, and LC-MS analysis was then performed on all the crude extracts followed by manual de-replications strategies, using the following libraries: Chemspider, Antibase, Reaxys and The Dictionary of natural products, in an attempt to identity the detected metabolites.

When comparing the two phase extracts quantitatively, the results indicated that from all the detected metabolites from the SM9 crude extracts, only a small portion (11 of 140 compounds) were detected in both organic and aqueous phase extracts, corresponding to 7.86% of the total

detected metabolites. These included the following metabolites: Surugamide A, Antimycin A_1, Desferrioxamine B and Desferrioxamine E (Figure 16). While in contrast, 50.71 % and 41.43% of the total detected metabolites were present in only either the organic phase or the aqueous phases respectively. A total of 71 compounds were detected only in the organic phase extracts, representing 50.71 % of the total metabolites that were detected, and included Maculosin and Bisucaberin B (Figure 16). In addition 58 compounds were only detected in the aqueous phase extracts, representing 41.43% of the total detected metabolites, and included the metabolites Lipoamicoumacin B and glu-ala-pro-pro (Figure 16). Thus more metabolite were detected from the organic phase extracts than from the aqueous phases extracts. But nevertheless, the aqueous phase extracts also contain interesting metabolites.

Furthermore, when focusing specifically on the production of potential new metabolites: 28 potential new metabolites were detected only in the organic phase crude extracts when compared to 29 being detected only in the aqueous phase extracts; while 3 were detected in both of the two phase extracts. This highlights the fact that despite the difficulties in analysing and detecting metabolites from the aqueous phases extracts, as these aqueous phases extract also typically contain impurities and ionic analytes (Sargent, 2013), there is still a good possibility of identifying potential new metabolites from aqueous phase extracts.

From these results, it is clear that the organic phase extracts contained the highest number of total detected metabolites (Figure 16) so our attention subsequently focused on the organic phase crude extracts for a more in-depth analysis of the metabolites produced in the different nutrient sources media. Figure 17a shows a Venn diagram representing both the specific and shared compounds produced by the SM9 isolate when it was fermented in the 6 media at 28°C]. The six media chosen from the 7 originally tested for this analysis, were chosen based on their highest number of detected metabolites by manual dereplication and on their highest bioactivity (Table 9).

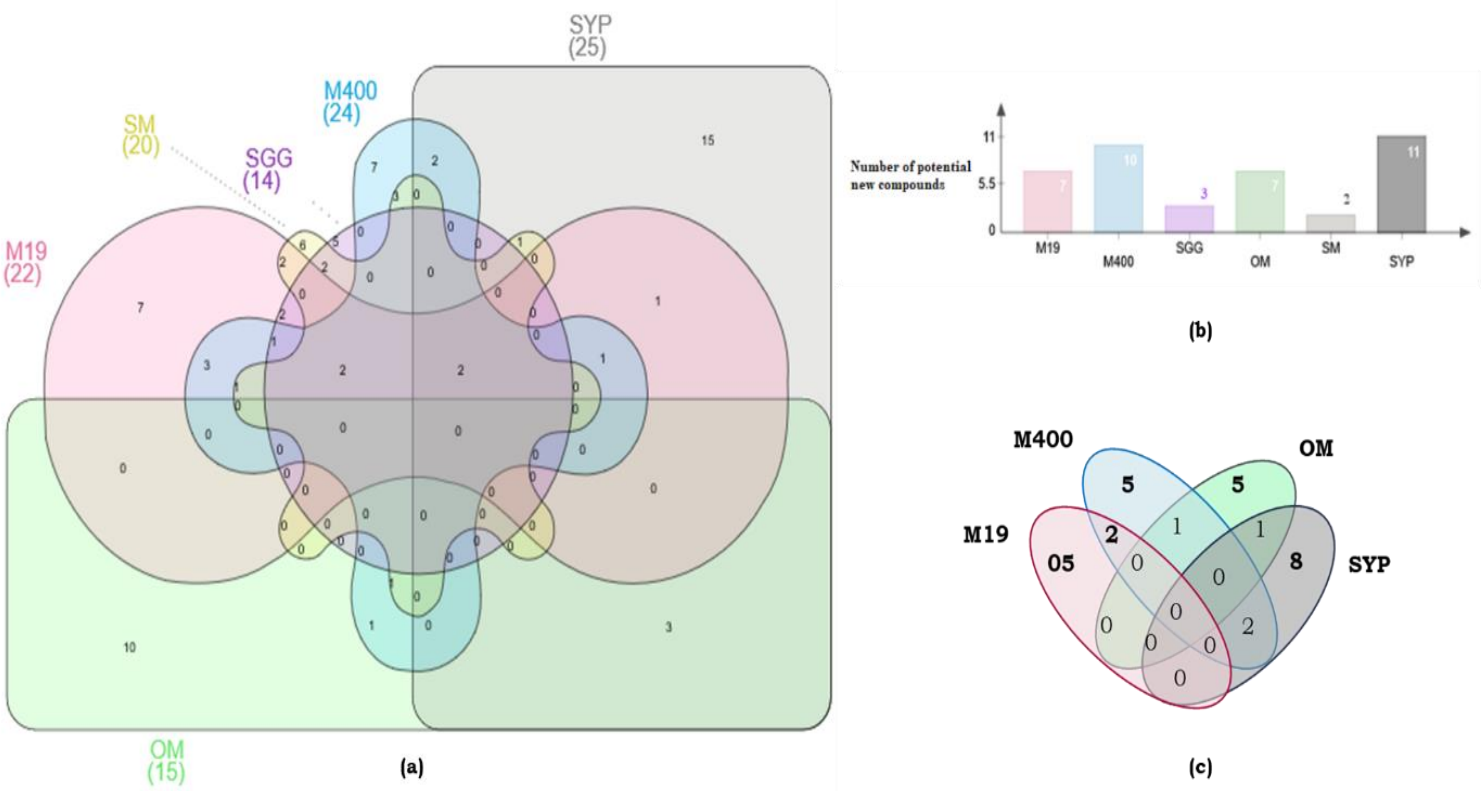


Figure 17: Venn diagram of the specific and shared metabolites produced by SM9 strain in six fermentation media from the organic phase crude extract samples (a), including their potential new metabolites (b). Venn diagram of the specific and shared potential new metabolites produced by SM9 strain, detected in the four most promising media extracts from the organic phase crude extracts (c).

The Venn diagram was produced by using the website from http://www.interactivenn.net/ - (Heberle et al. 2015). The histogram was produced from the Jvenn website (http://jvenn.toulouse.inra.fr/app/example.html) - (Bardou et al. 2014). The four media (Medium 19; Medium 400; Oatmeal and Media SYP) were selected from the results of the histogram in Figure 17b as the more promising media for the production of potential new metabolites when SM9 is fermented at 28°C.

Extracts from SM9 grown in SYP resulted in the production of the highest number of secondary metabolites, followed by M400, M19 and SM when comparing production from all six growth media. (Figure 17a). In all 25 metabolites were detected in SYP extracts, with 24 and 22 metabolites being detected in M400 and M19 extracts respectively; while 20, 14 and 15

metabolites were detected in SM, SGG and OM extracts respectively (Figure 17a). This suggests that the nutrient regimes in the SYP, M400 and M19growth media were optimal for metabolite production in SM9, further demonstrating the importance of nutrient source combination in regulating metabolite production in *Streptomyces* species. These results are in accordance with other studies focusing on the effect of nutrients on the production of secondary metabolites in marine bacteria (Romano et al., 2018)), including the earliest study of Okazaki, et al. (1975) who reported the production of an antibiotic produced by the marine actinomycete *Chainia purpurogena*, where production was dependent on the addition of an extract from the seaweed *Laminaria* (Okazaki et al., 1975). In addition, Sujathat et al. (2005) assessed the effect of 6 inorganic single nitrogen sources and 13 single amino acids, as single nitrogen sources, on the production of the antibiotic SBR-22 in the marine *Streptomyces psammoticus* BT-408 strain (Sujathat et al., 2005; Romano et al., 2018). They reported that optimization of medium and cultural conditions resulted in 1.82-fold increase in antibiotic yield, as well as, that Glucose and ammonium nitrate were found to be best carbon and nitrogen sources respectively for antibiotic production (Sujathat et al., 2005). Fan and co workers also investigated the impact of culture regime on chemical diversity by comparison of liquid and solid culture-derived fungal extract to confirm that the chemical composition and other characteristics including nutrient source (carbon source) of the growth media had significant effect on metabolite production of the fungi derived from *F. vesiculosus* (Fan et al., 2019).

Furthermore, from figure 17a, it is also clear that OM and SYP extracts followed by M19 and M400 extracts were those in which most of the specific metabolite produced by SM9 isolate were detected, representing 15, 10 followed by 7 and 7 specific unique metabolites respectively. For example in OM extract, 15 metabolites were detected. Among those 15 metabolites, 3 metabolites were also detected in SYP extract, 1 and 1 of them was also detected in M400 extract and in SM extract - in addition of OM extract, which make them not unique

metabolites as they are detected in several extracts from different culturing parameters; while the remaining 10 of the 15 metabolites produced were only detected in OM extract specifically, following specific OM culturing parameters making those 10 metabolites unique to OM culturing condition. Therefore, culturing SM9 in the OM media promoted the production of 15 metabolites, including 10 which were specifically dependent on the OM culturing parameters when compared to the others media tested and analyzed. Similarly, 15 of 25 metabolites that were produced were dependant on the SYP culturing parameters; while 7 of the 22 and 7 of the 24 metabolites produced were dependant on the M19 and M400 culturing parameters respectively (Figure 17a). It is clear that SYP extracts contained the highest number of both detected metabolites and potentially specific metabolites, with 25 metabolites being identified following growth in SYP, including 15 specific metabolites which appeared to only be produced in this growth media containing starch -yeast extract-peptone (SYP). Although, growth in OM, M19 and M400 did also result in the production of specific metabolites; with SM9 producing 10, 7 and 7 specific metabolites respectively, following growth in these media (Figure 17a). The use of an OSMAC based approach is regarded as an essential method for activating silent BGCs and to enhance the overall chemical diversity of microorganisms (Fan et al,. 2019). The presence of these specific 15,10,7 and 7 metabolites could suggests the potential activation of silent BCGs encoding these metabolites; due to varying culture parameters, although this would require further investigation to be confirmed.

When focusing our attention on the production of potential new metabolites, a total of 11, 10, 7 and 7 potential new compounds were detected in SYP, M400, M19 and OM extracts respectively; while only 3 and 2 potential new compounds were detected in SGG and SM media extracts (Figure 17b). This difference indicates that SYP, M400, M19 and OM appear to be the four most promising media for production of potential new metabolites by SM9 (Figure 17b). Furthermore, when looking at the shared and specific potential new metabolites detected

among these 4 most promising extracts, namely SYP, M400, M19 and OM we can see that for each extract, some of the potential new metabolites are specific to the culturing parameters while other potential new metabolites are shared between the extracts and therefore not specific to the culturing parameters (Figure 17c). For example, in the SYP extract: 8 of the 11 potential new metabolites were specifically detected only from that extract, while the remaining 3 potential new compounds were detected not only in SYP but also in M400 (for 2 of them) and in OM extracts (for 1 of them) (Figure 17c). The same was true for the OM, M400 and M19 extracts, where, only 5 of 7, 5 of 10 and 5 of 7 potential new metabolites were specifically detected in these specific culturing parameters respectively (Figure 17c).

Overall, when looking at the production of potential new metabolite detected in specific SM9 extracts, the most promising nutrient source was SYP, followed by M19 and OM which also gave good overall levels of metabolite production.

3.1.3 Detection of metabolites – metabolomic results

We then employed metabolomics analysis through molecular networking on the crude organic and aqueous extracts of the SM9 cultures that had been fermented in the six different culture media namely- M19, M400, OM, SGG, SM and SYP at 28°C; to examine the SM9 metabolome and to visualize the chemical diversity of the metabolites produced by the strain growing under these different conditions, by comparing the mass spectrometry profiles present in the extracts (Figure 18a). As with our previous study (Chapter 1) the Molecular networks analysis was performed using the MS/MS based Global Natural Product Molecular Networking GNPS online platform, which employs rapid and automated mass spectral mining of large number of samples by using high-resolution mass spectrometry parent ion fragmentation data (MS/MS);

with parent ions being connected based on the chemical architecture of the metabolite (Watrous *et al*., 2012; Fan *et al*., 2019; Oppong-Danquah *et al.*, 2018). The six media for the network analysis were selected based on the most promising number of potential new metabolites obtained following manual de-replication. The metabolome of SM9 is presented in Figure 18a, with each single coloured node corresponding to a metabolite detected in one specific media extract, while multi-coloured nodes correspond to metabolites detected in several media extracts. This allows us to observe which of the metabolites is present only in one media extract (specific) and which of the metabolite are present in a number media extract (shared). The edges between nodes correspond to the relatedness of the individual nodes within a cluster based on the cosine score. The cosine score determines the relationship between individual nodes to each other which is indicated by the visual thickness of the edge between related nodes in the network. The higher the cosine score, the thicker the edge will be and therefore the more related the individual nodes are to each other (Wang *et al*., 2016).

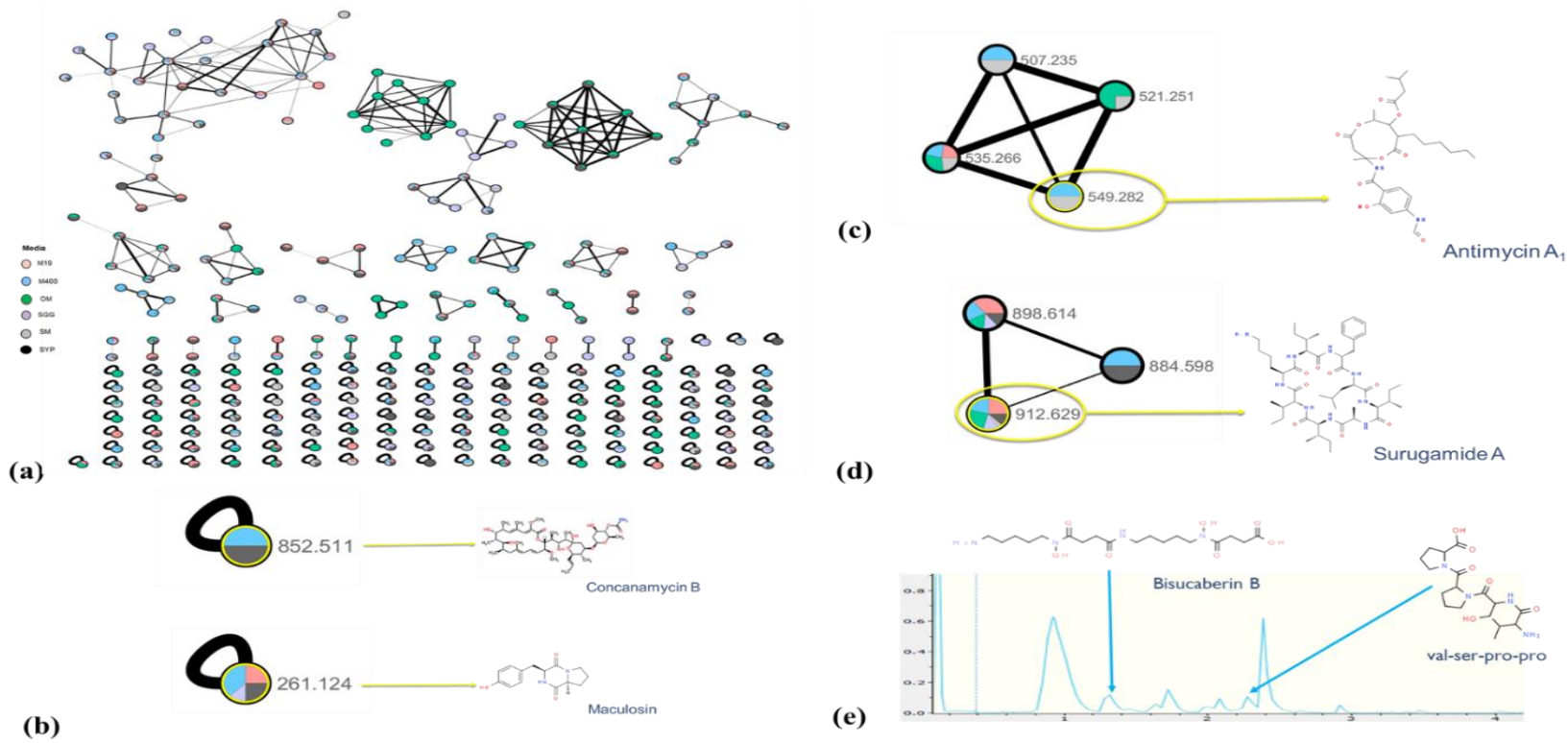


Figure 18: Molecular Network of SM9 fermented in 6 media (a). Enlarged view of the isolated metabolites Concanamycin B and Maculosin in the network (b). Enlarged sub-network view of the metabolites Antimycin A1 and Surugamide A in their respective clusters (c) and (d) respectively. Chemical structure of manually de-replicated compounds (e).

Results obtained using Chemspider, Antibase, GNPS and Cytoscape. Color code per extracts: pink= M19, blue=M400, purple= SGG, green= OM, grey= SM and black=SYP extracts. Picture of metabolites obtained from Chemspider. The nodes corresponds to the compounds detected by GNPS and Cytoscape. The edges corresponds to the individual node relatedness within a cluster based on the cosine score. This value determines the relationships between individual nodes to each other which is indicated by the visual thickness of the edge between related nodes in the network.

Globally, 36 metabolite clusters, from different size (from a cluster of 2 metabolites to cluster of 32 metabolites) were detected through molecular networking (Figure 18a). In addition, we can see that some metabolites produced by SM9 in only one specific media interact with other metabolites that can be produced following growth in a variety of different media. This is illustrated in Figure 18a, where within some clusters there are nodes highlighted by one colour only linked to other multi-coloured nodes, suggesting that those specific metabolites have the ability to interact with the shared metabolites. It is clear from Figure 18 that nutrient source is an important parameter in the culture conditions during fermentation and that the production of secondary metabolites is regulated at least to some extent by the nutrient sources present in the fermentation media. Figure 18 also highlights the complexity of this nutrient source regulations and suggest activation of different BGCs, encoding the metabolites that were observed under specific culture regime, as opposed to metabolites observed under several different culture regimes. This result is in accordance with Fan et al. (2019) study, which also suggested BGCs activation through metabolomics analysis, using molecular network and an OSMAC approach, when comparing liquid and solid culture regimes of fungi associated with the brown alga *Fucus vesiculosus* (Fan et al,. 2019).

In the SM9 metabolome, some metabolites were isolated that had been identified in the metabolite clusters present in the molecular network. Figure 18b shows the identification of Concanamycin B and Maculosin isolated from the metabolite clusters in the molecular network and which were also identified following the manual de-replication using ChemSpider and Antibase libraries. Concanamycin B was detected in extracts from SYP, and M400 by metabolomics analysis via GNPS and Cytoscape (Fig 18a) while manually detected only in SYP media extracts by chemical analysis via LC-MS and manual de-replication (Table 7). Concanamycin B is a polyketide macrolide (Sharom *et al*., 1995; Kinashi *et al*., 1984) which is an antibiotic and an inhibitor of vacuolar-type H+-ATPase, which suppress bone resorption

in vitro, and is also known to display antifungal and larvicidal properties (Woo *et al*., 1996; Kinashi *et al*., 1984; Sharom *et al*., 1995; Yilla *et al*., 1993) . It has also been reported to possess immunosuppressant properties in the induction of cytotoxic T cells (Lee *et al*., 2000). In parallel, Maculosin was detected in 4 media following metabolomics analysis (e.g. M400, M19, SYP and SGG) (Figure 18a) and detected in 4 media (e.g. M19, M400, MMM and SYP) by chemical analysis and manual de-replication (Table 7). MMM extracts were not analysed by metabolomics analysis. Maculosin [a diketopiperazine cyclo(-L-Pro-L-Tyr-)], was previously isolated from the fungus *Alternaria alternata* and has been reported to possess phytotoxicity activity against spotted knapweed (Stierle *et al*., 1988).

In the metabolome of SM9, some metabolites were isolated that had been identified in the metabolite clusters present in the molecular network. Figure 18b show the identification of Concanamycin B and Maculosin isolated from the metabolite clusters in the molecular network and which were also identified following the manual de-replication using ChemSpider and Antibase libraries. Concanamycin B was detected in extracts from SYP, and M400) by metabolomics analysis via GNPS and Cytoscape (Fig 18a) while manually detected only in SYP media extracts by chemical analysis via LC-MS and manual de-replication (Table 7). Concanamycin B is a polyketide macrolide (Sharom *et al*., 1995; Kinashi *et al*., 1984) which is an antibiotic and an inhibitor of vacuolar-type H+-ATPase, which suppress bone resorption in vitro, and is also known to display antifungal and larvicidal properties (Woo *et al*., 1996; Kinashi *et al*., 1984; Sharom *et al*., 1995; Yilla *et al*., 1993) . It has also been reported to possess immunosuppressant properties in the induction of cytotoxic T cells (Lee *et al*., 2000). In parallel, Maculosin was detected in 4 media following metabolomics analysis (e.g. M400, M19, SYP and SGG) (Figure 18a) and detected in 4 media (e.g. M19, M400, MMM and SYP) by chemical analysis and manual de-replication (Table 7). MMM extracts were not analysed by metabolomics analysis. Maculosin [a diketopiperazine cyclo (-L-Pro-L-Tyr-)], was previously

isolated from the fungus *Alternaria alternata* and which possess phytotoxicity activity against spotted knapweed (Stierle *et al*., 1988).

Nutrient source media	Extract Phases	RT [min]	$[M+H]^+$	Molecular formula	Compound name	Reported Biological activity	Additional reported information (Synthesis route, chemical group ect.)
M19, M400, SGG	Organic	1.4	419.2500	$C_{18}H_{34}N_4O_7$	Bisucaberin B	Ferric ion chelating activity (Fujita et al., 2013).	Siderophore (Fujita et al., 2013).
M19, M400, MMM, SYP	Organic	1.5	261.1238	$C_{14}H_{16}N_2O_3$	Maculosin	Phytotoxicity activity (knapweed and hypocotyls) (Stierle et al., 1998).	Cyclic peptide (Meenaa et al., 2019).
SYP	Organic	1.6	455.2134	$C_{20}H_{30}N_4O_8$	Bpagaome	n/a	n/a
M19	Water	1.7	413.2032	$C_{18}H_{28}N_4O_7$	glu-ala-pro-pro	n/a	n/a
SM SGG	Organic & Water Organic & Water	1.8	561.3611 561.3610	$C_{25}H_{48}N_6O_8$	Desferrioxamine B (=Deferoxamine)	Known to binds iron and aluminium. (Used in iron overdose, hemochromatosis). (Moeschlin et al.1963; Mobarra et al. 2016).	Siderophore (Neubauer et al., 2000).
SM M400	Organic Water	2.0	401.2391	$C_{18}H_{32}N_4O_6$	pro-thr-ala-ile	n/a	n/a
SM, SGG SM	Organic Water	2.5	601.3546 601.3548	$C_{27}H_{48}N_6O_9$	Desferrioxamine E (=Nocardamin)	Antibiotic agent, antibacterial activity against mycobacteria, weak antimicrobial against Gram+, anti-fungal activity, anti-tumor activity (Kalinovskaya et al. 2011; Ishida et al,.2011; Li et al,.2018).	Siderophore (Kalinovskaya, et al. 2011; Ishida et al,. 2011; Li et al,. 2018).

SYP	Organic	3.3	852.5073	$C_{45}H_{73}NO_{14}$	Concanamycin B	Antibiotic and inhibitor of vacuolar-type H+-ATPase, suppress bone resorption in vitro, shows antifungal and larvicidal properties (Woo et al. 1996, 1982; Kinashi ., et al. 1984; Sharom et al. 1995; Yilla M., et al. 1993). Immunosuppressants that inhibit the induction of cytotoxic T cells (Lee ., et al.2000). Inhibits the expression of MHC class II molecules on Colo 205 cells, abrogates the enhancement of the MHC class II expression and suppresses the antigen presentation by MHC class II molecules (Ito ., et al. 1995).	Polyketide Macrolide or Macrolide antibiotic (Sharom et al. 1995; Kinashi ., et al. 1984).
SYP	Water	3.3	716.4078	$C_{37}H_{56}N_4O_{10}$	Lipoamicoumacin B	n/a	NRPS-PKS hybrid (Li, et al. 2015).
M19,SGG	Organic	3.4	503.2844	$C_{26}H_{38}N_4O_6$	Probestin	Aminopeptidase N (APN) inhibitors, inhibits the angiogenic activity and tumor growth (Pathuri et al. 2013).	n/a
SM	Organic	3.7	271.0601	$C_{15}H_{10}O_5$	Bhimamycin B	Moderate activity against *Bacillus subtilis* and *Escherichia coli* (Fotso ., et al. 2003). Pocess the rare naphtho[2,3-c]furan-4,9-dione chromophore responsible of antioxidant and weak antiplasmodial activities (Bezabih et al. 2001; Fotso ., et al. 2003).	Isofuranonaphthoquinone (Fotso., et al. 2003).

SGG OM, SYP	Organic Water	4.9	912.6273	$C_{48}H_{81}N_9O_8$	Surugamide A	cathepsin B inhibitory cyclic peptides Inhibitory activity against bovine cathepsin B, a cysteine protease implicated in invasion of metastatic tumor cells (Ninomiya et al. 2016).	Synthesis route: NRPS (Ninomiya et al.2016).
M400, SGG, SYP	Organic	5.0	387.1811	$C_{23}H_{22}N_4O_2$	Fiscalin-B	Substance P inhibitor, neuroprotective and antitumor effects (Long et al.2018 ; Wong et al. 1993).	quinazolinone derivatives of the marine-derived alkaloids (Long et al.2018).
M19, MMM, SM	Organic		387.1803	$C_{22}H_{26}O_6$	Gibbestatin B	Inhibits the GA-induced expression of alpha-amylase expression (Hayashi et al. 2000).	n/a
SM	Organic & Water	6.8	549.282	$C_{28}H_{40}N_2O_9$	Antimycin A_1	Potential anti-cancer agent (Liu et al. 2016).	Synthesis route: Hybrid PKs-NRPS (Liu et al. 2016; Seipke et al. 2011).

Table 7: Some interesting identified metabolites produced by SM9 in different media conditions following the OSMAC approach

All compounds were identified by LC-MS, and manual de-replication, using Chemspider, Antibase, Reaxys, The Dictionary of Natural Products libraries. For some samples, different $[M+H]^+$ for different media extracts and or phases. n/a= not reported.

Some of the other metabolites observed in the molecular network could be grouped into a metabolite cluster, and form a sub-network, this can be seen in Figure 18c and 18d, which represent the sub-network of the Antimycin A_1 and Surugamide A metabolites that were detected in the crude extracts following both chemical and metabolomics analysis. Antimycin A_1, which is part of a cluster of 4 metabolites, was detected in both SM and M400 extracts following metabolomics analysis (figure 18c) and was only detected in SM extract from chemical analysis via LC-MS and manual de-replication (table 7). It is a potential anti-cancer agent (Liu *et al*., 2016) which is synthetized by PKS-NRPS hybrid route (Liu *et al*., 2016; Seipke *et al*., 2011). Surugamide A was detected in M19, M400, OM, SGG and SYP extracts following metabolomics analysis using GNPS and Cytoscape, and was also detected in SGG, OM and SYP extracts following chemical analysis using LC-MS and manual de-replication (Table 7). Surugamide A, was detected in a cluster of 3 metabolites, and has been previously reported to be produced by other marine sponge derived *Streptomyces* species such as *Streptomyces* sp. SM17 [Almeida et al., 2019]. The surugamides and their derivatives have been shown to possess a number of bioactivities, with the surugamides A–E and the surugamides G–J being shown to possess anticancer activity by inhibiting bovine cathepsin B, a cysteine protease reported to be involved in the invasion of metastatic tumour cells (Ninomiya *et al*., 2016); while another derivative, namely acyl-surugamide A, has been shown to possess anti-fungal activity. Surugamide A production in *Streptomyce*s SM17 has also previously been reported to be regulated by different carbon sources being present in the growth medium. High levels were reported following growth in YD media containing yeast extract, malt extract and dextrin sugars) with surugamide A production being repressed in TSB media, suggesting that carbon catabolite repression (CCR) may influence the production of surugamides in SM17 (Almeida et al., 2019). The high cosine score which is visible by the high thickness of the edges (Figure 18c) suggests that the metabolites in this cluster may be

synthetized by the same biosynthetic route. As this cluster contains one module which has been identified as a PKS-NRPS Hybrid involved in: Antimycin A_1 synthesis, this suggest that this metabolite cluster may be related to a PKS-NRPS hybrid region cluster. Also, the metabolite cluster presented in figure 18d containing 3 metabolites may also be related to a NRPS cluster as it also contain a metabolite which has been identified as likely to be synthetized from a NRPS route (e.g: Surugamide A - 912.629). On the other hand, compared to the cluster in figure 18c, the cluster in figure 18d has a low cosine score and therefore may not contain only NRPS metabolites , which mean that in addition to contain NRPS metabolite, it may also contain other metabolite which are not all synthetized via NRPS route). Indeed, the edges between metabolites 912.629 and 884.598, as well as between metabolites 884.598 and 898.614 are not very thick, representing a low cosine score and therefore could also suggest a different biosynthetic route. Thus a more thorough investigation on the 884.598 and or 898.614 metabolites would need to be undertaken to confirm the exact biosynthetic route used to synthetized of this cluster.

During the manual de-replication, the chemical tools chemspider, antibase, reaxys, and the dictionary of natural product, were employed to identify the metabolites that were produced. Then following bioinformatic and metabolomic analysis, we attempted to link those metabolites that had been identified to the molecular network. However, a number of metabolites which were identified, that were present in the chemical extracts could not be not linked to the metabolomic network (Figure 18e). Of these Bisucaberin B, is a siderophore with a ferric ion chelating activity (Fujita *et al.*, 2013); also known to be produced by marine microorganisms such as by the marine bacterium *Tenacibaculum mesophilum)* (Fujita *et al.*, 2013; Chen et al., 2019); was only detected when SM9 was grown on the M19, M400 and SGG media, while the val-ser-pro-pro metabolite was only detected when SM9 was grown on the M400 media (Figure 18e).

The largest metabolite cluster of the SM9 metabolome is presented in Figure 19. It represents the sub-network of the following metabolites Desferioxamine B, Desferioxamine E, pro-thr-ala-ile and glu-ala-pro, that were detected and identified following both chemical and metabolomics analysis, following growth of SM9 in a number of different culture media. Desferrioxamine B and Desferrioxamine E were both detected in 2 extracts, namely SM and SGG) by manual de-replication (Table 7) and subsequently detected in additional extracts by metabolomics analysis using GNPS and Cytoscape tools, namely M19, M400, SGG, SM and M19, M400, SGG, SM and SYP extracts respectively. The related chemical chromatograms of Desferrioxamine B and Desferrioxamine E in SGG extracts can be seen in Figure 19. Glu-ala-pro-pro and Pro-thr-ala-ile were also detected and identified by both chemical and metabolomics analysis, and their relative chromatograms can also be seen of Figure 19.

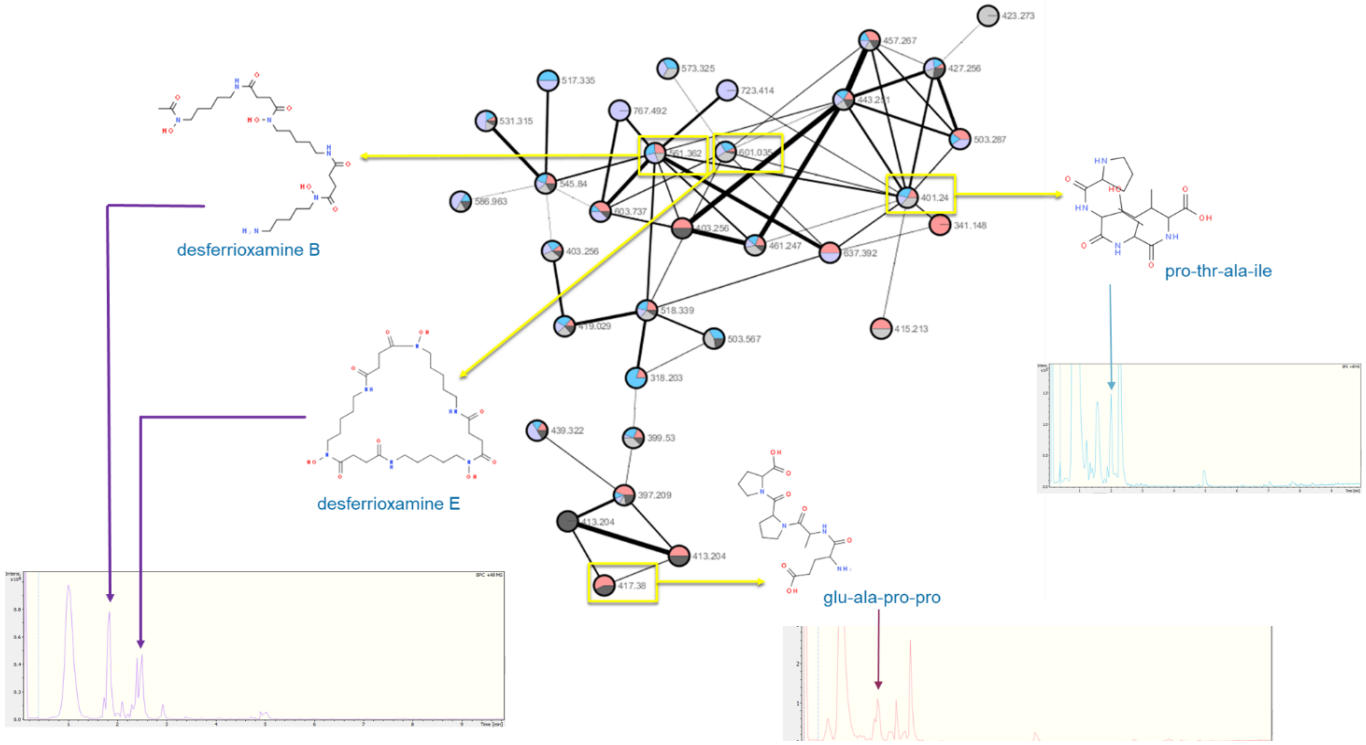


Figure 19: Enlarged sub-network view of the Desferrioxamine B, Desferrioxamine E, glu-ala-pro-pro and pro-thr-ala-ile metabolites in the metabolite cluster, all together with their relative chromatograms from the LC-MS and manual de-replication analysis.

Results obtained using Chemspider, Antibase, GNPS and Cytoscape. Metabolite picture obtained from Chemspider. Color code per extracts: pink= M19, blue=M400, purple= SGG extracts, grey= SM and black=SYP extracts.

Desferrioxamine B is a siderophore which is known to bind iron and aluminium (Sven Moeschlin *et al*., 1964; Mobarra *et al*., 2016; Neubauer *et al*., 2000), while Desferrioxamine E is also a siderophore, that is known to be an antibiotic agent and to possess anti-bacterial, anti-fungal and anti-tumor activity (Kalinovskaya *et al*., 2011; Ishida *et al*., 2011; Yan Li *et al*., 2018). In addition, both siderophores have previously been reported to be produced by marine microorganisms (Chen et al., 2019) including the very well studied marine actinomycete *Salinispora tropica* CNB-440 (Ejje N et al., 2013; Chen et al., 2019). Desferrioxamine B has also been reported to be produced by *Streptomyces pilosus* (Schupp *et al*., 1988) and by *Streptomyces coelicolor* M145 (Barona-Gomez et al., 2004).

3.2 Identification of the secondary metabolite regions responsible of the biosynthesis of the Antimycin A_1, Desferrioxamine B, Desferrioxamine E and Surugamide A detected metabolites from SM9 extracts.

Genome sequencing was performed by Macrogen (Seoul, South Korea), using the PacBio RSII sequencing platform. The PacBio raw reads were processed and quality filtered using the BamTools toolkit v2.4.1 (subread length >1000, subread quality >0.75) (Barnett et al., 2011). The genome assemblies were performed using the Canu v1.7 software (Koren et al., 2017), followed by assembly polishing using Quiver v2.1.0 (Pacific Biosciences Inc). The assembly coverage check was performed using the BBMap program v37.90 (Almeida et al., 2019). Genome assembly statistics were calculated using the QUAST v4.6.3 program (Gurevich et al., 2013). Genome annotation was performed using the Prokka v1.12 program for this study's analysis (Seemann, 2014), and with the NCBI Prokaryotic Genome Annotation Pipeline for data submission to the GenBank database (Tatusova et al., 2016; Benson et al., 2018).

Following assembly using APAdes, the SM9 genome was initially analyzed using AntiSMASH version 3.0.5 and following the subsequent update of the software was then re-analyzed by

AntiSMASH version 5.0.0. The updated version of AntiSMASH identified 56 potential biosynthetic gene clusters (BGCs) compared to 16 potential BGCs identified by the previous AntiSMASH version (data not shown). This could be explained by the fact that the new updated version of AntiSMASH offers more detailed predictions when compared to the previous version, including for example, new gene cluster classes; new refinement of cluster detection rules; improved type II PKS predictions; a new *'region'* concept instead of the original *'cluster'* previous concept which better distinguish the different biological options that lead to BGCs, together with many other features and updates (Blin et al., 2019).

The 56 potential BGCs that were identified included 19 NRPS, 21 PKS, 9 terpene, 3 siderophore, 2 bacteriocin, 1 ectoine and 1 LAP (Linear azol(in)e-containing peptides (LAPs)). In addition, Ectoine, Carotenoid, Hopene, Vazabitide A, Desferrioxamine B, Geosmin, Rhizomide A / B/ C, Surugamide A/D, Albaflavenone, Micromonolactam, Antimycin, and Frontalamides gene regions were identified based on the similarities with most similar know cluster analyzed with AntiSMASH (Table 8). This high number of BCGs and these region clusters highlight the high potential of SM9 to be a big producer of secondary metabolites. Also, the identification of these regions suggest that some of our metabolites detected by LC-MS could be synthetized via the activation of these regions clusters, such as antimycin A_1, desferrioxamine B and surugamide A.

Region	Type	From	To	Most similar known cluster		Similarities
25.1	T1PKS	1	5,481	Frontalamides	nrps-t1pks	28%
103.1	ectoine	1	6,485	Ectoine	other	100
110.1	terpene	1	2,296	Carotenoid	terpene	18
127.1	terpene	2,425	13,344	Geosmin	terpene	100
151.1	bacteriocin	1	3,324	-		
182.1	siderophore	1	1,726	-		
193.1	terpene	1	2,261	-		
227.1	terpene	1	1,156	Hopene	terpene	15
241.1	T1PKS	1	2,949	-		
349.1	T1PKS	1	6,092	-		
360.1	NRPS	1	2,723	Vazabitide A	NRPS	4
369.1	terpene	1	2,923	Hopene	terpene	15
478.1	NRPS	1	7,124	-		
507.1	terpene	1	1,185	-		
517.1	terpene	1	7,758	Phosphoglycans	saccharide	6
541.1	terpene	1	1,213	Geosmin	terpene	100
630.1	T1PKS	1	4,508	Micromonolactam	t1pks	100
637.1	T3PKS	1	13,805	Herboxidiene	t1pks+t3pks	9
664.1	T1PKS	1	3,458	-		
684.1	T1PKS	1	3,301	Micromonolactam	t1pks	100
715.1	NRPS	1	2,118	Lipopeptide 8D1-1/ 8D1-2	NRPS	6
736.1	T1PKS	1	6,136	Micromonolactam	t1pks	100
741.1	NRPS	1	6,637	-		
744.1	bacteriocin	1	4,012	-		
759.1	siderophore	14,655	22,222	Desferrioxamine B	other	100
790.1	NRPS	1	12,075	Dechlorocuracomycin	NRPS	8
799.1	terpene	3,815	23,987	Albaflavenone	terpene	100
860.1	NRPS	1	6,866	-		
882.1	LAP	1	2,741	-		
909.1	T1PKS	1	4,877	Micromonolactam	t1pks	100
924.1	T1PKS	1	3,213	Micromonolactam	t1pks	100
949.1	T1PKS	1	5,842	Stambomycin	t1pks-saccharide	36
965.1	NRPS	1	3,167	Pyridomycin	nrps-t1pks	
967.1	T1PKS	1	3,280	Micromonolactam	t1pks	100
976.1	NRPS	1	3,764	-		
1014.1	T2PKS	1	8,750	Fredericamycin	t2pks	36
1023.1	T1PKS	1	5,654	JBIR-100	t1pks	33
1026.1	NRPS	1	8,054	Surugamide A / D	NRPS	33
1039.1	NRPS	1	1,968	Rhizomide A / B/ C	NRPS	100
1041.1	NRPS	1	7,380	Surugamide A / D	NRPS	19
1042.1	T1PKS	1	3,596	Micromonolactam	t1pks	100
1044.1	NRPS	1	2,805	-		
1045.1	NRPS	1	8,504	-		
1047.1	T1PKS	1	2,259	Micromonolactam	t1pks	100
1049.1	NRPS	1	2,519	-		
1050.1	NRPS	1	2,503	-		
1059.1	T1PKS	1	3,830	Micromonolactam	t1pks	100

1072.1	NRPS	1	9,945	Decholorocuracomycin	NRPS	8
1076.1	T1PKS	1	7,145	Cremimycin	t1pks	17
1081.1	NRPS	1	4,036	-		
1083.1	T1PKS	1	4,482	Micromonolactam	t1pks	100
1084.1	T1PKS	1	5,032	Micromonolactam	t1pks	100
1085.1	T1PKS	1	9,695	Micromonolactam	t1pks	100
1086.1	NRPS	1	10,354	-		
1137.1	siderophore	1	7,144	-		
1193.1	NRPS	1	11,400	Antimycin	nrps-t1pks	20

Table 8: Secondary metabolite region clusters of SM9, identified by AntiSMASH version 5.0.0. Region: the region number; Type: the product types as detected by antiSMASH; From, To: the location of the region (in nucleotides); Most similar known cluster: the closest compound from the MiBIG database, along with its type (e.g. t1pks+t3pks); Similarity: a percentage of genes within the closest known compound that have a significant BLAST hit to genes within the current region (Blin *et al.*, 2019).

(Result were reloaded from AntiSMASH- version 3.0.5 to version 5.0.0. from the assesion number: NEUA01000302.1).

3.2.1 Genomic results – Identification of the secondary metabolite regions responsible of the biosynthesis of the Antimycin A_1

Using the KnowClusterBlast from AntiSMASH, it is clear that the SM9 genome contains a NRPS region [1193.1] which has 20% similarities with the antimycin gene cluster at the nucleotide level (Figure 20). More precisely, "similarity" on AntiSMASH documentations are visible in this link Understanding the output - antiSMASH Documentation (secondarymetabolites.org) and are referred as "a percentage of genes within the closest known compound that have a significant BLAST hit to genes within the current region".

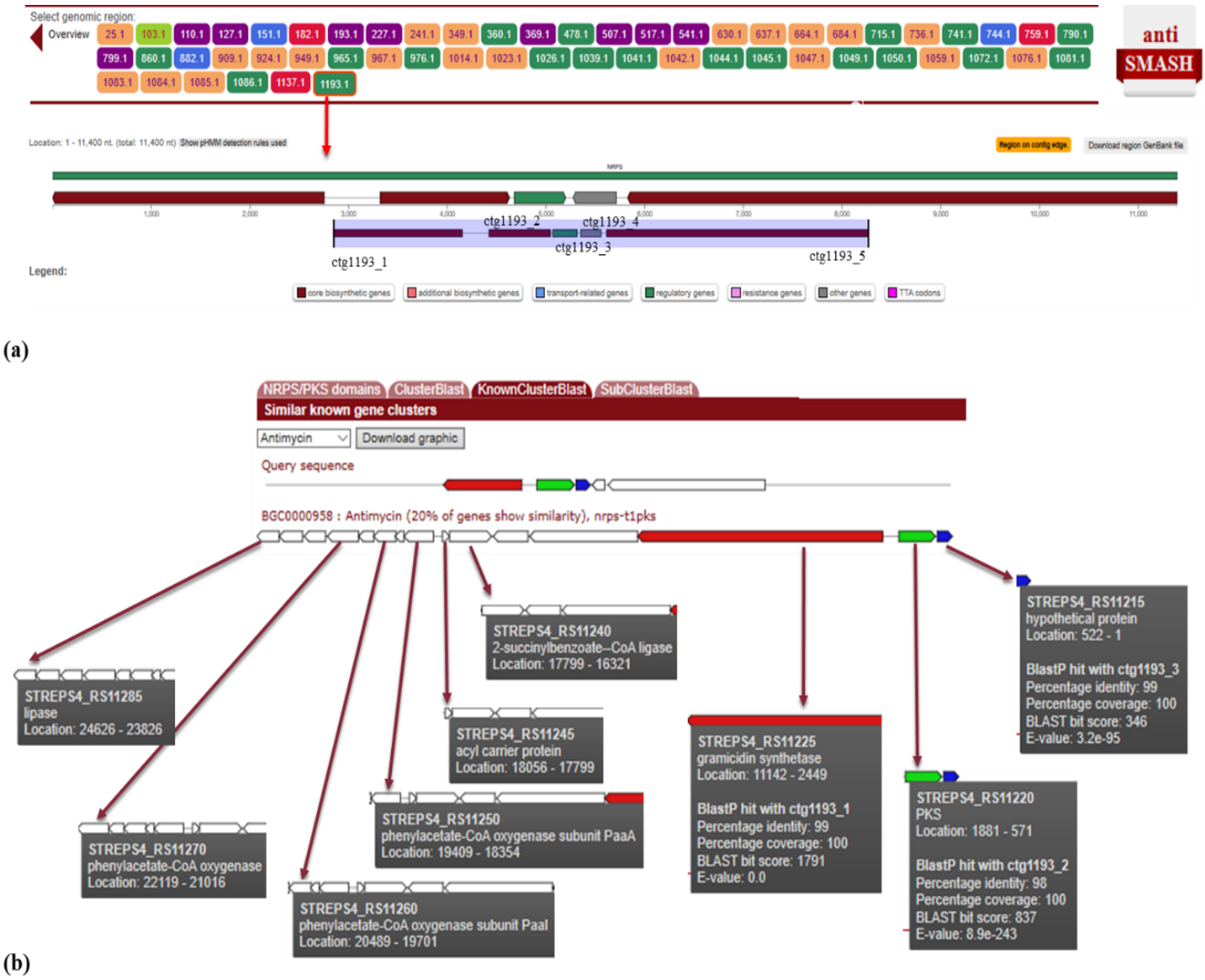


Figure 20: Detail of NRPS region 1193.1 identified by AntiSMASH version 5.0.0 (a). KnownClusterBlast of the region cluster showing similarities and differences with the referenced BGC0000958 antimycin A_1 gene cluster (b).

From Figures 20a and 20b, all the genes required to synthetize an Antimycin metabolite are present in this region cluster, such as core biosynthetic genes and regulatory gene (Figure 20a). Furthermore, the Antimycin metabolite that SM9 isolate is likely to produce from this region cluster is one Antimycin without Phenylacetate-CoA oxygenase, Phenylacetate-CoA oxygenase subunit PaaI and PaaA and a CoA-ligase compared to the Antimycin produced by

the reference BGC0000958 (from *Streptomyces sp.* S4) Antimycin gene cluster (Figure 20b). (The missing genes are not considered as necessary to Antimycin biosynthesis). This corroborates the chemical results previously obtained, where the Antimycin A_1 metabolite was detected following LC-MS analysis in SM9 extracts.

This was also confirmed following BlastP analysis on this NRPS region, where we could see that both the ctg1193_1 and ctg1193_2 genes share 100% of query cover and 99% identities with the AntC and AntB genes respectively, while the ctg1193_3 gene shares 100% query cover, and appears to encode an RNA polymerase sigma factor (Figure 21). Thus this results suggest that the ctg1193_1 gene is likely to encode for antC, that the ctg1193_2 gene encodes for antB and that the ctg1193_3 gene encodes for the σ^{AntA} sigma factor. It has previously been reported that σ^{AntA} may represent a new sub-family of ECF σfactors that to date have only been found in the *ant* gene clusters of *Streptomyces* species (Seipke et al., 2014).

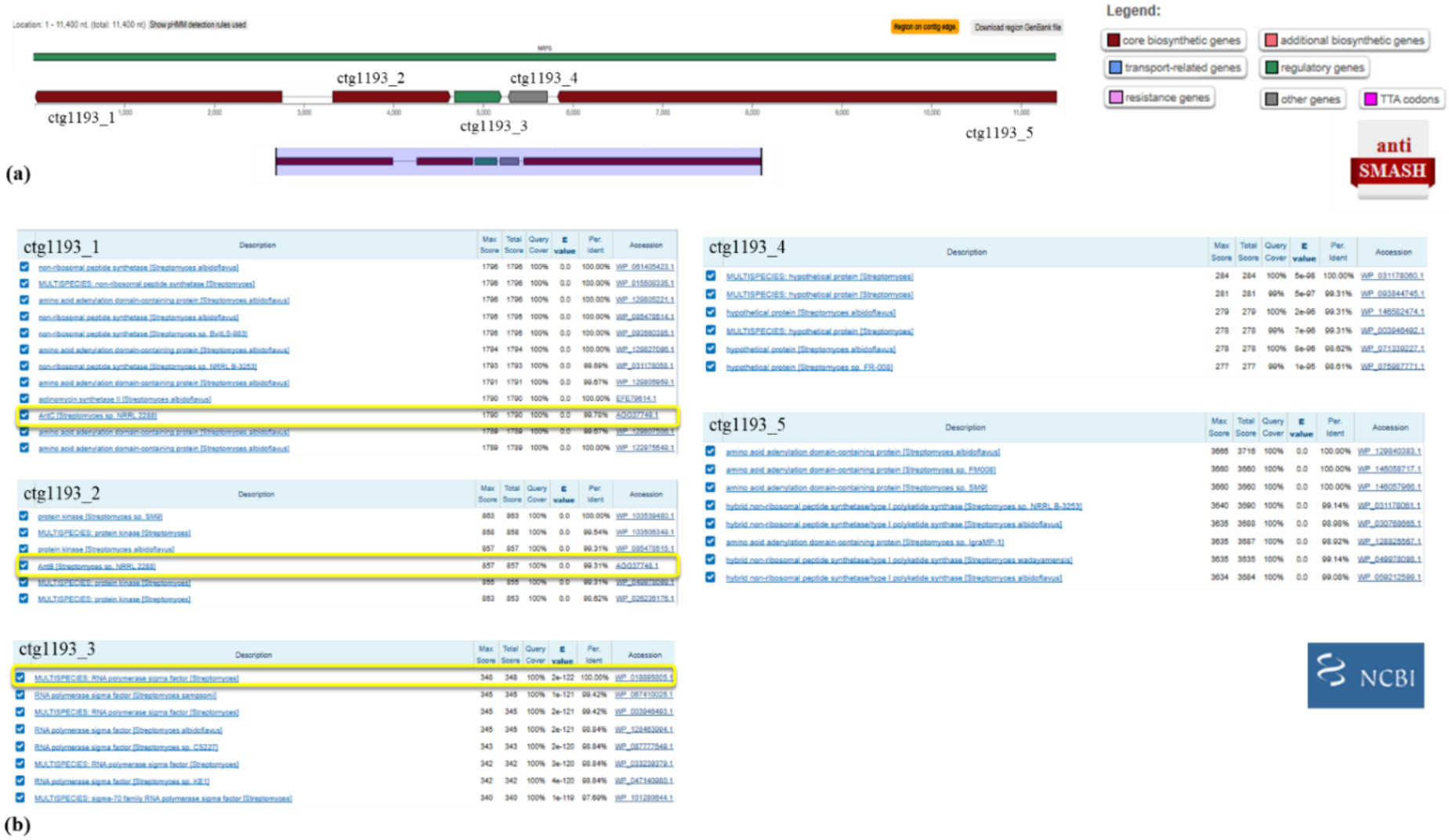


Figure 21: Gene legend (a) from the NRPS region 1193.1 identified by AntiSMASH version 5.0.0, with the Blastp identifications obtained from NCBI database (b).

Thus it is clear when considering the chemical, metabolomic and genomics data that the 1193 region gene cluster in SM9 is likely to encode for genes involved in the synthesis of Antimycin $A_{1.}$ In addition, it is also suggesting that the weak percentage of similarities (only 20%) with the referenced BGC0000958 Antimycin region cluster is due to the absence of Phenylacetate-CoA componments in our Antimycin A_1. This is more suggested when looking at the specific definition of similarity by AntiSMASH similarity is "a percentage of genes within the closest known compound that have a significant BLAST hit to genes within the current region" [available at this link: Understanding the output - antiSMASH Documentation (secondarymetabolites.org)]. Therefore the percentage of gene that have a hit with the referenced BGC0000958 antimycin, is lower, as the hit were only for 3 genes (antB, antC and σ^{AntA} sigma factor) out of the 15 total genes of the referenced BGC0000958 which were containing the missing Phenylacetate-CoA componments. Mathematically speaking: 3 of 15 = 20%.

3.2.2 Genomic results – Identification of the secondary metabolite regions responsible of the biosynthesis of the Desferrioxamine B and Desferrioxamine E

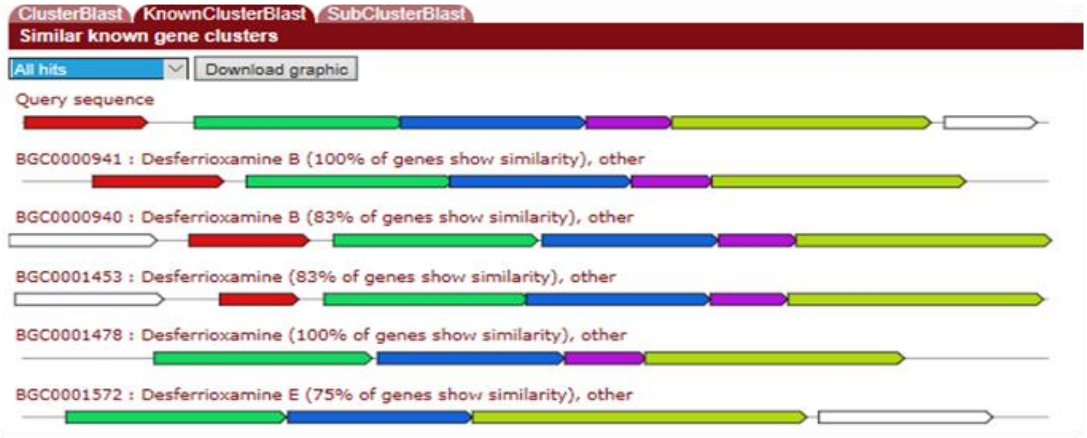
Following analysis using AntiSMASH it was possible to identify the region on the SM9 genome (Siderophore region 759.1); which encodes the potential production of both Desferrioxamine B and Desferrioxamine E.

When focusing on Desferrioxamine B, the siderophore region 759.1 displayed 100% similarity with the Desferrioxamine B region cluster from the referenced BGC0000941 cluster (from *Streptomyces griseus subsp. griseus* NBRC 13350) using the KnownClusterBalst analysis (Figure 22b) and possess both core biosynthetic genes and additional biosynthetic gene

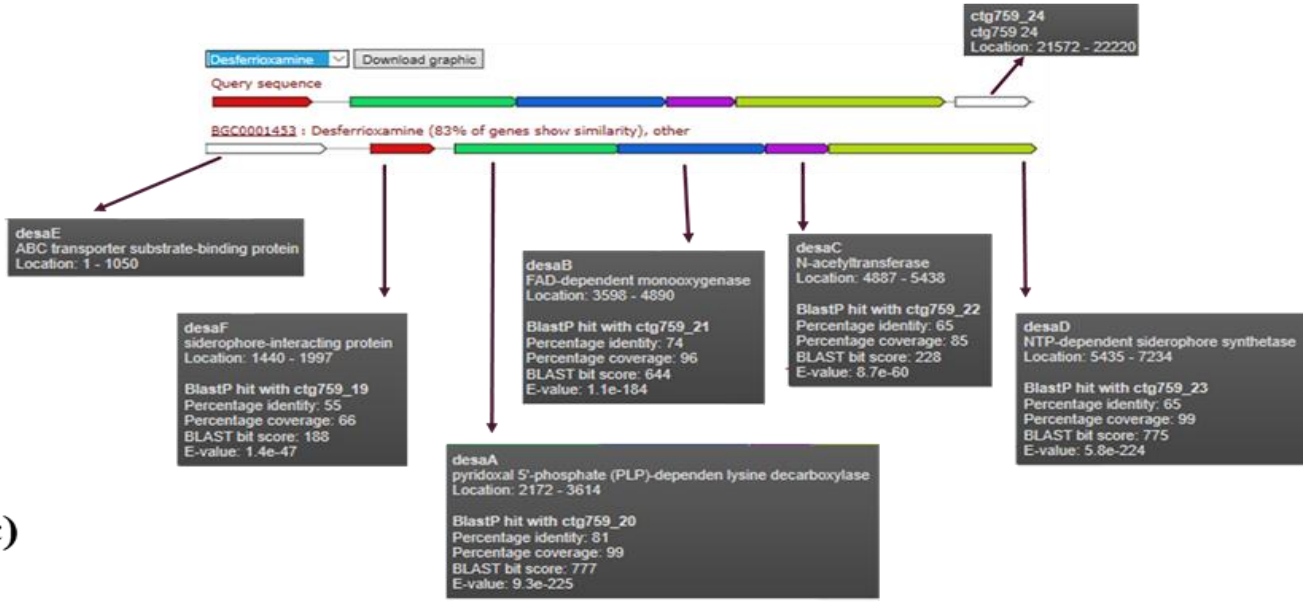
necessary for the synthesis synthetize of Desferrioxamine B (Figure 22a). The ctg759_21 gene in SM9 is most closely related to *desa*B gene, showing 74% identity with the referenced BGC0001453 (from *Streptomyces argillaceus*) (Figure 22c). In addition, the ctg759_19, ctg759_21, ctg759_22 and ctg759_23 genes displayed 55%, 81%, 74%, 65% and 65% identity with desaF, desaA, desaB, desaC and desaD with the reference strain BCG0001453 respectively (Figure 22c). Thus it is clear that the Desferrioxamine B metabolite present in the crude extract from SM9 is likely to have been synthetized from the ctg759_21 gene in the siderophore region 759.1 of SM9 genome.



(a)



(b)



(c)



(d)

Figure 22: Detail of the Siderophore region 759.1 identified by AntiSMASH version 5.0.0 (a). KnownClusterBlast of the region cluster showing similarities with referenced desferrioxamine B and desferrioxamine E gene clusters (b). Details of the similarities and differences with the following referenced gene clusters: BGC0001453 desferrioxamine B (c) and BGC0001572 desferrioxamine E (c).

Regarding Desferrioxamine E, the siderophore region 759.1 on the SM9 genome displayed 75% similarity with the Desferrioxamine E region cluster from the reference BGC0001572, using KnownClusterBlast analysis (Figure 22b). It also appears to possess the necessary core and additional biosynthetic genes to synthetize Desferrioxamine E (Figure 22a), as confirmed by detection of the compound in the LC-MS analysis (Table 8).

A notable difference between the reference clusters (both BGC001453 and BCG001572) and SM9 can be seen in the desaE gene, encoding an ABC transporter substrate-binding protein (Figure 22c) or encoding an MSF transporter DfoS (Figure 22d). In the reference BGC001453, desaE is observed at the 3' end of the cluster and is 1049 nt in length, whereas desaE appears to be absent in SM9, instead containing an open reading frame (ctg759_24; Figure 22c) of 649nt in length at the 5' end of the cluster. Similarly, In the reference BCG001572, which represent specifically the Desferrioxamine E gene cluster (Figure 22d), the gene AWD37884.1 which could be the desferrioxamine E gene have also a bigger size, size of 1226nt and like in SM9 query, it is located at the 5' end of the cluster. BLASTp analysis found similarities between ctg759_24 and DUF4429 domain-containing proteins in other *Streptomyces* species (see supplementary figures 11); however, the function of this domain is unknown. Both SM9 and the reference BCG001572 are capable of producing Desferrioxamine E; therefore suggesting either that ctg759_24 encodes a protein of similar function to DesaE or of a modified form of DesaE, such as for exemple a smaller DesaE.

Taken together, this confirms the chemical results from the SM9 extracts where Desferrioxamine B and Desferrioxamine E metabolites were both detected from SM9, and suggests the likelihood that the Siderophore region 759.1 in the SM9 genome, encodes for the genes responsible for the synthesis of the Desferrioxamine B and Desferrioxamine E.

3.2.3 Genomic results – Identification of the secondary metabolite regions responsible of the biosynthesis of the Surugamide A

Following AntiSMASH analysis it was also possible to link Surugamide A biosynthesis to the NRPS region 1026.1 on the SM9 genome, with the NRPS Region 1026.1 sharing 33% similarity with the surugamide A/D region BGC0001792 following KnownClusterBlast analysis (Figure 23b). While this is a somewhat low overall percentage similarity nonetheless it is clear that the region possess all the essential genes necessary to encode for the production of a Surugamide metabolite (core synthetic gene and transport related gene including core NRPS gene and MFS transporters) (Figure 23b). Indeed, it is a small percentage but as we see in the figure 23a and 23b, the Surugamide synthetized by SM9 genome will be mostly a Surugamide without ABC transporters and ATP binding cassettes, TetR/AcrR family transcriptional regulator, compared to the reference Surugamide A synthetized through the BGC0001792 surugamide gene cluster (Figure 23c).

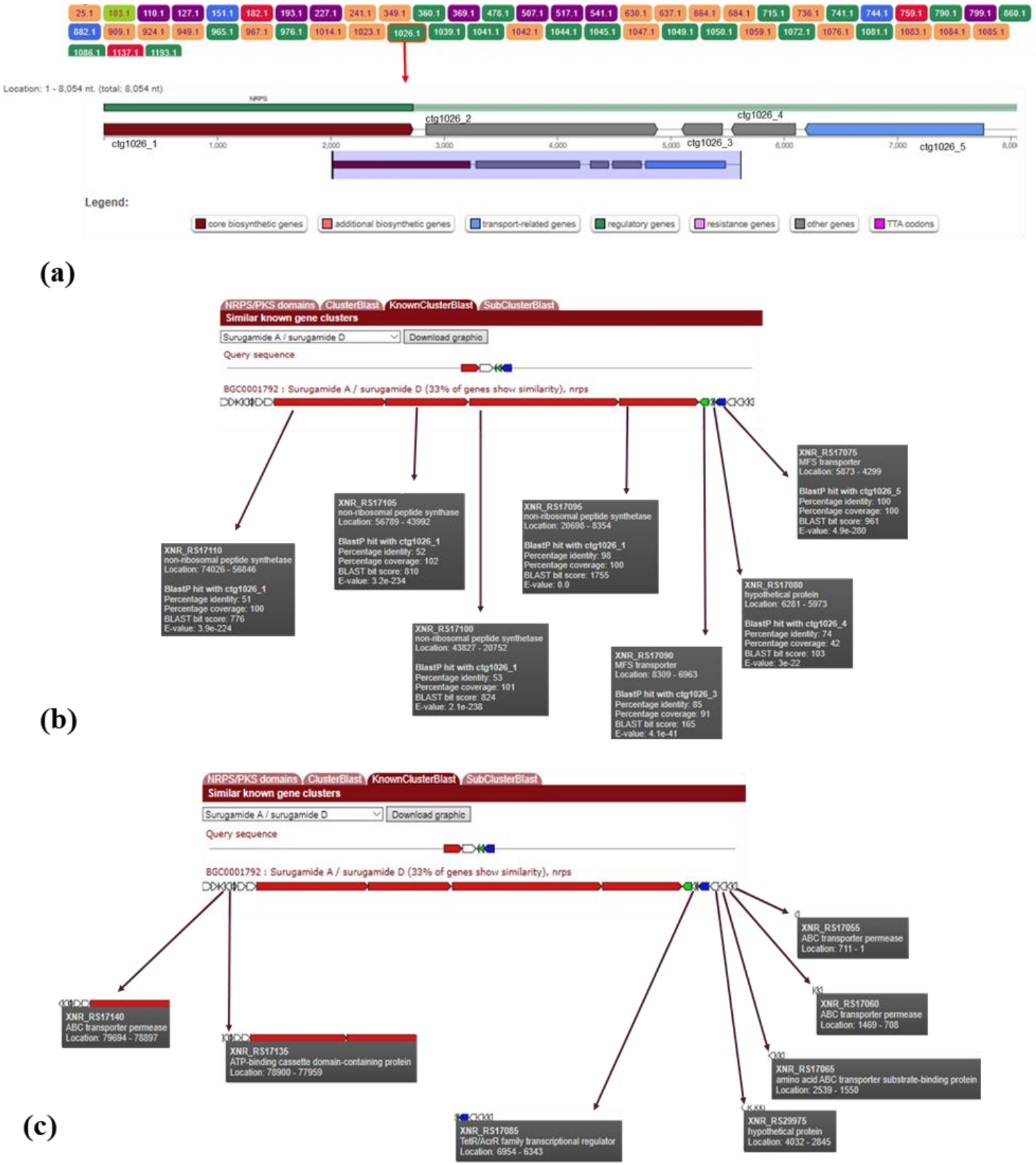


Figure 23: Detail of the NRPS region 1026.1 identified by AntiSMASH version 5.0.0 (a) KnownClusterBlast of the region cluster showing similarities (b) and differences (c) with the referenced BGC0001792 Surugamide A gene cluster.

In addition, from our chemical data, Surugamide A was only detected in SGG, OM and SYP extracts suggesting that this transcription of this NRPS region 1026.1 may be regulated by different nutrient regimes and may only be activated under specific conditions.

As previously mentioned surugamide A production in another marine Streptomyces strain SM17 has previously been reported to be differentially regulated by nutrient sources, with production being increased in growth media containing yeast extract, malt extract and dextrin sugars; and repressed in TSB media, suggesting some degree of control being mediated by carbon catabolite repression (CCR), which may be the case here as well with SM9 (Almeida et al., 2019).

Conclusions

Streptomyces strain SM9 which displayed anti-fungal activities against *C. glabrata* (CBS138) and *C. albicans* (Sc5314), was used to produce secondary metabolites following growth on 7 different nutrient sources using the OSMAC approach. The metabolites that were produced were extracted using LLE (Liquid liquid extraction) to obtain crude extracts. Additional chemical analysis (LC-MS and manual de-replication) was then performed to monitor metabolite production with both already known metabolites and potential new metabolites being detected using Chemspider, Reaxys, Antibase and the Dictionary of Natural Products libraries.

Overall, a total of 140 metabolites were detected by LC-MS analysis followed by manual de-replication using the aforementioned libraries from both organic and aqueous phases extracts from the 7 different growth media; including 60 potential new metabolites. Among the already known metabolites detected, the following were observed namely: we found Bisucaberin B, Maculosin, Bpagaome, glu-ala-pro-pro, Desferrioxamine B, pro-thr-ala-ile, Desferrioxamine

E, Concanamycin B, Lipoamicoumacin B, Probestin, Bhimamycin B, Surugamide A, Fiscalin B, Gibbestatin B and Antimycin A_1 in these extracts.

SM9 strain was selected as a promising strain for the anti-fungal activity, the production of secondary metabolites and the production of potential new metabolites.

SM9 crude extracts were also analyzed by metabolomics tools to monitor their chemical diversity and the SM9 metabolome through molecular networking using GNPS and Cytoscape. In total 36 metabolite clusters of different sizes (from a cluster of 2 metabolites to a cluster of 32 metabolites) were detected through molecular networking. Among those metabolites clusters termed as "sub-networks" for the following metabolites were identified. These included Desferioxamine B, Desferioxamine E, glu-ala-pro-pro and pro-thr-ala-ile which were present in a cluster of 32 metabolites; together with Antimycin A_1 which was present in a cluster of 4 metabolites and Surugamide A1 which was present in a cluster of 3 metabolites. In addition Concanamycin B and Maculosin which were not part of a sub-network were isolated in the molecular network. These metabolites were detected in only some of the specific media in which SM9 was grown. For example, Antimycin A_1 was only detected when SM9 was grown in SM and M400 media, while Surugamide A was detected in 5 of the growth media extracts (M19, M400, OM, SGG and SYP) which confirms the chemical data were some metabolites were only found in some extracts and not others. This demonstrates the importance of the nutrient source parameter in the culture conditions during growth of SM9 and that the production of secondary metabolites is regulated at least to some extent by the nutrient sources present in the growth media. This also highlights the complexity of this nutrient source regulation as those observed metabolic differences suggest activation of different BGCs encoding these metabolites under different culture regimes.

Following genomic analysis, 56 secondary metabolite encoding regions were identified with AntiSMASH, including 19 NRPS, 21 PKS, 9 terpene, 3 siderophore, 2 bacteriocin, 1 LAP and 1 ectoine encoding region within the SM9 genome. Among these, regions potentially encoding the following metabolites namely; Ectoine, Carotenoid, Hopene, Vazabitide A, Desferrioxamine B, Geosmin, Rhizomide A / B/ C, Surugamide A/D, Albaflavenone, Micromonolactam, Antimycin, and Frontalamides encoding metabolite were identified.

By linking the metabolomics and genomic data for SM9 it was then possible to identify regions/clusters of genes on the genome that were likely to be responsible for the production of a) Desferrioxamine B and Desferrioxamine E which were linked to the Siderophore region 759, b) Antimycin A_1 which was linked to the NRPS region 1193 together with c) Surugamide A which was linked to the NRPS region 1026.1 on the SM9 genome.

In addition, in the NRPS region 1193.1 responsible of the synthesis of the Antimycin A_1, *antC* could also be linked to the ctg1193_1 gene, *antB* could be linked to ctg1193_2 and a gene which is likely to encode the σ^{AntA} sigma factor was linked to ctg1193_3 using BlastP analysis (Figure 20 and 21). Given that the σ^{AntA} sigma factor has previously been shown to be involved in the biosynthesis of Antimycin A in *Streptomyces albus*(Seipke *et al.*, 2014), this suggest that the ctg1193_3 gene from the region 1193.1, is likely to be involved in the biosynthesis of the detected Antimycin A_1 in SM9 (Figure 20). In the siderophore region 759.1, potentially responsible for the synthesis of both Desferrioxamine B and Desferrioxamine E , the ctg 759_21 gene could be linked to *desaB* using AntiSMASH with 74% identity, while ctg759_19 was linked to *desaF* (55% identity) and the ctg759_20 gene was linked to *desaA* (81% identity). In addition the ctg 759_22 gene was linked to *desaC* (65% identity), the ctg759_23 gene was linked to *desaD* (65% identity) and finally the ctg759_24 gene could be linked to *desaE*. This suggest that the siderophore region 759.1 is responsible for the biosynthesis of both

Desferrioxamine B and Desferrioxamine E, with the ctg759_21 being involved in the synthesis of Desferrioxamine B, and ctg759_24 with the synthesis of Desferrioxamine E (Figure 22).

Thus the detection of Antimycin A_1, Desferrioxamine B, Desferrioxamine E and Surugamide A metabolites by chemical analysis using LC-MS and manual de-replication could be linked to metabolite clusters with bioinformatics and metabolomics analysis using molecular networking, GNPS and Cytoscape. Then together these could be linked to their respective gene cluster region by genomics analysis using AntiSMASH and genome mining. While the genomics data helped to corroborate the chemical data, it is clear that the metabolomics data helped both the genomics and chemical interpretations by giving a more in-depth perspective of the SM9 strain. In addition, both chemical analysis (LC-MS, manual de-replication) and genomic research analysis (genome mining, AntiSMASH,) can be very time-consuming, since chemical analysis requires analysis of a mixture of thousands of metabolites or more and genomics analysis require analysis of thousands of genes from several genomic regions, both fields can benefit from the use of metabolomics analysis. Indeed, metabolomics greatly increases the efficiency of this process by detecting most of the metabolites present in the mixture and helps in the identification of some of these metabolites by comparing their chemical profile simultaneously in several libraries - allowing us to determine which metabolite is interesting to further investigation in fewer experiments such as by LC-MS and NMR, instead of conducting experiments on thousands of metabolites and then attempting to identify metabolites one by one, and subsequently determining the more interesting metabolites. From a genomics perspective, metabolomics also increases the efficiency, as it allows more time to be spent in connecting metabolites to the genomic regions clusters from which they are likely to be produced. It also can provide a visual hint as to which metabolite could potentially be synthetized from a “silent” gene cluster, allowing us to focus our next analysis on the hypothesis of the presence of the potential “silent” gene cluster of interest to

directly confirm this hypothesis or not. In both chemical and genomics fields, the use of metabolomics analysis, allows for less time-consuming experiments, and more focused in-depth analysis of the potential of SM9 to produce secondary metabolites.

Material and Methods

- **Origin of strains**

The bacterial strains SM9 was obtained from the marine sponge *Haliclona simulans* (class Demospongiae, order Haplosclerida, family Chalinidae), at a depth of 15 m in Galway bay (Ireland) (Kennedy et al., 2009).

DNA extraction

Genomic DNA from SM9 was isolated using the phenol-chloroform-isoamyl alcohol extraction method (Wilson, 2001). Isolate was grown overnight (~16 h) in 5 mL liquid cultures. Cells were pelleted by centrifugation (6000 g), supernatants were decanted and discarded. Cell pellets were resuspended in 467 µL TE buffer. 30 µL of 10% SDS and 3 µL of 20 mg/mL Proteinase K (Fermentas, Sankt Leon-Rot, Germany) was added to each tube and incubated for 1 h at 37 °C. An equal volume of phenol/chloroform (phenol-chloroform-isoamyl alcohol mixture ratio 25:24:1, Sigma Aldrich, Arklow, Ireland) was added and mixed well. Tubes were centrifuged, at ~18,042 g for 10 min (Eppendorf Centrifuge, 5417r, Eppendorf UK Ltd., Stevenage, UK). The upper phase of each Eppendorf tube was aspirated to fresh 2 mL Eppendorf tubes, avoiding the interphase. Sodium acetate (NaOAc) 3M (100 µl) pH 5.2 was added to each tube and mixed well. 600 µL of isopropanol (Sigma Aldrich, Arklow, Ireland) was added, mixed well, and incubated at room temperature for 15 min. Tubes were then centrifuged at ~18,042 g for 20 min. Supernatants were removed and discarded. DNA pellets

were washed with cold (4 °C) 70 % EtOH. Tubes were centrifuged at ~18,042 g for 10 min. Ethanol was removed and discarded and DNA pellets were allowed to air-dry. DNA was re-suspended in 1 mL TE buffer. 1 µL of RNase A was added to the tubes, which were then incubated at 37 °C for 30 min. DNA was again purified by phenol extraction. DNA was analysed by gel electrophoresis and quantified using a spectrophotometer (NanoDrop ND-1000, Thermo Scientific, Gloucester, UK). The DNA solutions were stored at −20 °C.

- **Whole genome sequencing**

For whole genome sequencing, it was performed by Macrogen (Seoul, South 1334 Korea), mate pair libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. All libraries were sequenced in 250 bp paired read runs on the Illumina MiSeq platform. Reads were trimmed for quality with Sickle (https://github.com/najoshi/sickle) and Scythe (https://github.com/vsbuffalo/scythe).

- **Genomic and bioinformatics methods**

The SM9 genome was assembled by Dr. Eduardo Almeida by using SPAdes, version 3.1.1. The potential bioactive gene clusters were identified by searching the genomes with AntiSMASH version 3.0.5 and parsed with R methods, R is a free software environment for statistical computing and graphics (https://www.r-project.org/). Then, following advances in technologies in identify region clusters, the genome was later re-analysed with AntiSMASH new version 5.0.0, from the published assesion number: NEUA01000302.1, to be sure to have the most updated results in term of identifying region clusters.

The region clusters involved in the synthesis of the following detected metabolites (Desferrioxamine B, Desferrioxamine E, Surugamide A_1) were identified with genome minding methods using AntiSMASH version 5.0.0, Blastp and Blastn.

- **Production of secondary metabolites by fermentation**

The production of secondary metabolites was performed following the OSMAC approach used in chapter 1 (see chapter 1 material and methods) with fermentation of the strains in different nutrient-sources media and different temperature conditions. The composition of the fermentation media is provided in Table 2 of the chapter 1. The *Streptomyces* SM9 strain was first streaked on a plate and incubated at 28 °C for 3 days to allow for sporulation, then a single colony was selected to be cultivated in 250mL Erlenmeyer flasks, containing 50mL of rich media M410 with and without presence of synthetic sea salt (Instant Ocean). They were incubated for 7 days at 28°C, in a rotary shaker incubator at 200 rpm. 1mL of the culture media without salt, was inoculated to 250mL flasks containing 49mL of each respective media (OM/ MMM/ SGG/ M400/ M19/ SYP/ SM) and incubated for 14 days at 28°C, 200rpm. In addition, 1 mL of each the culture M410 media (with and without salt) was used for the analysis of anti-microbial activity. The fermented culture samples were centrifuged at 4°C, max speed for 10-15min. The pellet was freeze-dried at-80°C and the cell free supernatant (50mL) stored at -80°C or used immediately for extraction. Identical methods were employed for growth of the strains at 30°C. See table 2 for exact details about the media used for fermentation.

- **Antimicrobial activities screening - Well diffusion assays**

The isolated SM9 strain was screened for antimicrobial activity using the well diffusion assay method against different test strains such as *Staphylococcus aureus* (NCIMB 9518), *Pseudomonas aeruginosa* (PAO1), *Bacillus subtilis* (IE32), *Escherichia coli* (NCIMB 12210), *Candida glabrata* (CBS138), and *Candida albicans* (Sc5314). YPD (Yeast Extract-Peptone-

Dextrose) and LB (Luria-Bertani) agar plates were prepared and streaked with a single strain of each test strains at an OD=0.1, then wells were manually made in each plate by using sterile tips. A volume of 75 µL of the cell free supernatant from the M410-strain cultures broth, at day 7 of fermentation, were poured into each well. Plates were then incubated overnight at 37 °C for the bacterial plates and 28 °C for the fungal plates. Both anti-bacterial and anti-fungal activities, were deduced by observation of growth inhibition around the well. To confirm the validity of the results, in parallel, negative controls were performed by replacing the volume of 75 µL of culture by 75 µL of sterile water and the results were validated by the absence of growth inhibition surrounding the wells.

- **Liquid Liquid Extraction (LLE)**

Cell-free supernatants (50 mL) were chemically extracted using Liquid Liquid Extraction (LLE) with ethyl-acetate as solvent. The solvent was added to the filtrate in the ratio of 1:1(v/v), shaken vigorously and left to stand for 5 min to obtain the organic and aqueous phases. The organic phase was separated from the aqueous phase using a separating funnel. Both layers were concentrated by evaporating to dryness at 45 °C in a GeneVac to generate dry crude extracts.

- **Liquid Chromatography - Mass Spectrometry (LC-MS)**

LC-MS analysis was performed at the University of Aberdeen, in the Marine Biodiscovery Centre, during my secondment.

LC-MS was conducted on all the crude extracts on a Bruker MAXIS II QTOF. The Bruker MAXIS II QTOF was run in tandem with an Agilent 1260 HPLC using a Phenomenex Kinetex 2.6 mM XB-C18, 100 x 2.1 mm column. All samples were run at a concentration of 0.5 mg/ml using a gradient of 95 % Water / 5 % Acetonitrile - 100 % Acetonitrile.

- **Manual de-replication**

Manual de-replication was performed at the University of Aberdeen, in the Marine Biodiscovery Centre, during my secondment.

Manual de-replication of the metabolites from the LC-MS data was obtain by deduction of the possible molecular weights and molecular formulas predicted from the Bruker software using Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). For the interpretation of the MS data a number of different libraries were used, such as Chemspider, Antibase, Reaxys and The Dictionary of Natural Products. The potential new metabolites were deduced by the absence of any hits following searched on these databases. A second manual de-replication of the identified metabolites was obtained by matching the LC-MS raw data with the spectral data from the Molecular Network.

- **Venn diagrams and histogram**

The Venn diagram in Figure 17, was produced by using the website from (http://www.interactivenn.net/)- (Heberle *et al*., 2015). The histogram in Figure 17, was produced from the Jvenn website (http://jvenn.toulouse.inra.fr/app/example.html) - (Bardou *et al*., 2014).

- **Metabolomics analysis and Visualisation of Molecular Networking-**

The raw data generated from the LC-MS was converted into mzXLM file format by using the downloaded zip file GNPS_Vendor_Conversion obtained through the DataAnalysis software from the online portal of Global Natural Products Social Molecular Networking GNPS (https://gnps.ucsd.edu). All mzXLM files were larger than 20 MB so they were uploaded to ProteoSAFe (http://proteomics.ucsd.edu/ProteoSAFe/) using FileZilla as an ftp client. When all the files were uploaded to ProteoSAFe they became available within the GNPS system.

Separately, a metadata table composed of attributes (metadata) describing the samples properties was created by following the GNPS requirements workflow, in a text file format, and then used as an input to the molecular networking of GNPS. Molecular networking was performed using the online data analysis portal of Global Natural Products Social Molecular Networking. The data were filtered by removing all MS/MS peaks within ±17 Da of the precursor *m/z*. The MS/MS spectra were window filtered by choosing only the top 6 peaks in the ±50 Da window throughout the spectrum. The data was then clustered with MS-Cluster with a parent mass tolerance of 0.02 Da and an MS/MS fragment ion tolerance of 0.02 Da to create consensus spectra. Consensus spectra that contained less than 2 spectra were discarded. A network was then created, where edges were filtered to have a cosine score of above 0.7 and more than 6 matched peaks. Further edges between two nodes were retained in the network only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. All matches that were kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. The molecular network was visualized using Cytoscape version 3.7.1 and displayed using the settings “Organic layout” with “sample 2” style. To visualize each different media from all samples, regardless of the organics and water phases, color-codes was established manually for each media via the group columns from “pie” in “chart” in the Node window.

- **Colour-code**

To help visualize the results, a colour code was applied to all the different media used for production of secondary metabolites. This colour code was use in both Figure 18 and Figure 19. It is as follow: M19 red, M400 blue, MMM yellow, OM green, SGG purple, SM grey and SYP black.

References

Almeida, E. L., Kaur, N., Jennings, L. K., Carrillo Rincon, A. F., Jackson, S. A., Thomas, O. P., & Dobson, A. D. W. (2019). Genome Mining Coupled with OSMAC-Based Cultivation Reveal Differential Production of Surugamide A by the Marine Sponge Isolate *Streptomyces* sp. SM17 When Compared to Its Terrestrial Relative S. albidoflavus J1074. *Microorganisms, 7*(10). doi:10.3390/microorganisms7100394

Antoraz, S., Santamaria, R. I., Diaz, M., Sanz, D., & Rodriguez, H. (2015). Toward a new focus in antibiotic and drug discovery from the *Streptomyces* arsenal. *Front Microbiol, 6*, 461. doi:10.3389/fmicb.2015.00461

Bardou, P., Mariette, J., Escudie, F., Djemiel, C., & Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics, 15*, 293. doi:10.1186/1471-2105-15-293

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics, 27*(12), 1691-1692. doi:10.1093/bioinformatics/btr174

Barona-Gomez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J., & Challis, G. L. (2004). Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J Am Chem Soc, 126*(50), 16282-16283. doi:10.1021/ja045774k

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018). GenBank. *Nucleic Acids Res, 46*(D1), D41-D47.

doi:10.1093/nar/gkx1094

Bezabih, M., Abegaz, B. M., Dufall, K., Croft, K., Skinner-Adams, T., & Davis, T. M. (2001). Antiplasmodial and antioxidant isofuranonaphthoquinones from the roots of *Bulbine capitata*. *Planta Med, 67*(4), 340-344. doi:10.1055/s-2001-14329

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res, 47*(W1), W81-W87. doi:10.1093/nar/gkz310

Chen, J., Guo, Y., Lu, Y., Wang, B., Sun, J., Zhang, H., & Wang, H. (2019). Chemistry and Biology of Siderophores from Marine Microbes. *Marine Drugs, 17*(10). doi:10.3390/md17100562

Ejje, N., Soe, C. Z., Gu, J., & Codd, R. (2013). The variable hydroxamic acid siderophore metabolome of the marine actinomycete *Salinispora tropica* CNB-440. *Metallomics, 5*(11), 1519-1528. doi:10.1039/c3mt00230f

Fan, B., Parrot, D., Blumel, M., Labes, A., & Tasdemir, D. (2019). Influence of OSMAC-Based Cultivation in Metabolome and Anticancer Activity of Fungi Associated with the Brown Alga *Fucus vesiculosus*. *Marine Drugs, 17*(1). doi:10.3390/md17010067

Fotso, S., Maskey, R. P., Grun-Wollny, I., Schulz, K. P., Munk, M., & Laatsch, H. (2003). Bhimamycin A to approximately E and bhimanone: isolation, structure elucidation and biological activity of novel quinone antibiotics from a terrestrial *Streptomycete*. *J*

Antibiot (Tokyo), 56(11), 931-941. doi:10.7164/antibiotics.56.931

Fujita, M. J., Nakano, K., & Sakai, R. (2013). Bisucaberin B, a linear hydroxamate class siderophore from the marine bacterium *Tenacibaculum mesophilum*. *Molecules, 18*(4), 3917-3926. doi:10.3390/molecules18043917

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086

Hayashi, K.I., Inoguchi, M., Kondo, H., Nozaki, H. Gibbestatin B inhibits the GA-induced expression of α-amylase expression in cereal seeds. *Phytochemistry*. Volume 55, Issue 1, September 2000, Pages 1-9. ISSN 0031-9422. https://doi.org/10.1016/S0031-9422(00)00204-1.

Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., & Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics, 16*, 169. doi:10.1186/s12859-015-0611-3

Ishida, S., Arai, M., Niikawa, H., & Kobayashi, M. (2011). Inhibitory effect of cyclic trihydroxamate siderophore, desferrioxamine E, on the biofilm formation of *Mycobacterium* species. *Biol Pharm Bull, 34*(6), 917-920. doi:10.1248/bpb.34.917

Ito, K., Kobayashi, T., Moriyama, Y., Toshima, K., Tatsuta, K., Kakiuchi, T., Futai, M., Ploegh, H. L., Miwa, K. (1995). Concanamycin B inhibits the expression of newly-

synthesized MHC class II molecules on the cell surface. *J Antibiot (Tokyo), 48*(6), 488-494. doi:10.7164/antibiotics.48.488

Jackson, S. A., Crossman, L., Almeida, E. L., Margassery, L. M., Kennedy, J., & Dobson, A. D. (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene Clusters in the Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates. *Marine Drugs, 16*(2). doi:ARTN 67
10.3390/md16020067

Kalinovskaya, N. I., Romanenko, L. A., Irisawa, T., Ermakova, S. P., & Kalinovsky, A. I. (2011). Marine isolate *Citricoccus* sp. KMM 3890 as a source of a cyclic siderophore nocardamine with antitumor activity. *Microbiol Res, 166*(8), 654-661. doi:10.1016/j.micres.2011.01.004

Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., Bourke, M. B., Rea, M. C., O'Connor, P. M., Ross, R. P., Hill, C., O'Gara, F., Marchesi, J. R., Dobson, A. D. (2009). Isolation and analysis of bacteria with antimicrobial activities from the marine sponge *Haliclona simulans* collected from Irish waters. *Mar Biotechnol (NY), 11*(3), 384-396. doi:10.1007/s10126-008-9154-1

Kinashi, H., Someno, K., & Sakaguchi, K. (1984). Isolation and characterization of concanamycins A, B and C. *J Antibiot (Tokyo), 37*(11), 1333-1343. doi:10.7164/antibiotics.37.1333

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M.

(2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res, 27*(5), 722-736. doi:10.1101/gr.215087.116

Lee, M. H., Kataoka, T., Honjo, N., Magae, J., & Nagai, K. (2000). *In vivo* rapid reduction of alloantigen-activated CD8+ mature cytotoxic T cells by inhibitors of acidification of intracellular organelles, prodigiosin 25-C and concanamycin B. *Immunology, 99*(2), 243-248. doi:10.1046/j.1365-2567.2000.00961.x

Li, Y., Li, Z., Yamanaka, K., Xu, Y., Zhang, W., Vlamakis, H., Kolter, R., Moore, B. S., Qian, P. Y. (2015). Directed natural product biosynthesis gene cluster capture and expression in the model bacterium *Bacillus subtilis*. *Sci Rep, 5*, 9383. doi:10.1038/srep09383

Li, Y., Zhang, C., Liu, C., Ju, J., & Ma, J. (2018). Genome Sequencing of *Streptomyces atratus* SCSIOZH16 and Activation Production of Nocardamine via Metabolic Engineering. *Front Microbiol, 9*, 1269. doi:10.3389/fmicb.2018.01269

Liu, J., Zhu, X., Kim, S. J., & Zhang, W. (2016). Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivities. *Nat Prod Rep, 33*(10), 1146-1165. doi:10.1039/c6np00004e

Long, S., Resende, D., Kijjoa, A., Silva, A. M. S., Pina, A., Fernandez-Marcelo, T., Vasconcelos, M. H., Sousa, E., Pinto, M. M. M. (2018). Antitumor Activity of Quinazolinone Alkaloids Inspired by Marine Natural Products. *Marine Drugs*, *16*(8). doi:10.3390/md16080261

Martin, J. F., Sola-Landa, A., Santos-Beneit, F., Fernandez-Martinez, L. T., Prieto, C., & Rodriguez-Garcia, A. (2011). Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in *Streptomyces*. *Microb Biotechnol*, *4*(2), 165-174. doi:10.1111/j.1751-7915.2010.00235.x

Meena, M., & Samal, S. (2019). Alternaria host-specific (HSTs) toxins: An overview of chemical characterization, target sites, regulation and their toxic effects. *Toxicol Rep*, *6*, 745-758. doi:10.1016/j.toxrep.2019.06.021

Mobarra, N., Shanaki, M., Ehteram, H., Nasiri, H., Sahmani, M., Saeidi, M., Goudarzi, M., Pourkarim, H., Azad, M. (2016). A Review on Iron Chelators in Treatment of Iron Overload Syndromes. *Int J Hematol Oncol Stem Cell Res, 10*(4), 239-247.

Moeschlin, S., & Schnider, U. (1964). Treatment of Primary and Secondary Hemochromatosis and Acute Iron Poisoning with a New Potent Iron Eliminating Agent (Desferrioxamine B-Dfom). *Ned Tijdschr Geneeskd, 108*, 1648-1653.

Neubauer, U., Nowack, B., Furrer, G., Schulin, R. Heavy Metal Sorption on Clay Minerals Affected by the Siderophore Desferrioxamine B. Environ. Sci. Technol. 2000, 34, 13, 2749–2755. https://doi.org/10.1021/es990495w

Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., Wakimoto, T., Ohnishi, Y., Matsunaga, S., Takada, K. (2016). Biosynthetic Gene Cluster for Surugamide A Encompasses an Unrelated Decapeptide, Surugamide F.

Chembiochem, 17(18), 1709-1712. doi:10.1002/cbic.201600350

Niu, S., Li, S., Chen, Y., Tian, X., Zhang, H., Zhang, G., Ju, J., Zhang, C. (2011). Lobophorins E and F, new spirotetronate antibiotics from a South China Sea-derived *Streptomyces* sp. SCSIO 01127. *J Antibiot (Tokyo), 64*(11), 711-716. doi:10.1038/ja.2011.78

Okazaki, T., Kitahara, T., & Okami, Y. (1975). Studies on marine microorganisms. IV. A new antibiotic SS-228 Y produced by Chainia isolated from shallow sea mud. *J Antibiot (Tokyo), 28*(3), 176-184. doi:10.7164/antibiotics.28.176

Oppong-Danquah, E., Parrot, D., Blumel, M., Labes, A., & Tasdemir, D. (2018). Molecular Networking-Based Metabolome and Bioactivity Analyses of Marine-Adapted Fungi Co-cultivated With Phytopathogens. *Front Microbiol, 9*, 2072. doi:10.3389/fmicb.2018.02072

Pan, R., Bai, X., Chen, J., Zhang, H., & Wang, H. (2019). Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review. *Front Microbiol, 10*, 294. doi:10.3389/fmicb.2019.00294

Pathuri, G., Thorpe, J. E., Disch, B. C., Bailey-Downs, L. C., Ihnat, M. A., & Gali, H. (2013). Solid phase synthesis and biological evaluation of probestin as an angiogenesis inhibitor. *Bioorg Med Chem Lett, 23*(12), 3561-3564. doi:10.1016/j.bmcl.2013.04.031

Petersen, L.E., Kellermann, M..Y., Schupp, P.J. (2020) Secondary Metabolites of Marine

Microbes: From Natural Products Chemistry to Chemical Ecology. In: Jungblut, S., Liebich, V., Bode-Dalby, M. (eds) YOUMARES 9 - The Oceans: Our Research, Our Future. Springer, Cham. https://doi.org/10.1007/978-3-030-20389-4_8.

Reen, F. J., Gutierrez-Barranquero, J. A., Dobson, A. D., Adams, C., & O'Gara, F. (2015). Emerging concepts promising new horizons for marine biodiscovery and synthetic biology. *Marine Drugs, 13*(5), 2924-2954. doi:10.3390/md13052924

Romano, S., Jackson, S. A., Patry, S., & Dobson, A. D. W. (2018). Extending the "One Strain Many Compounds" (OSMAC) Principle to Marine Microorganisms. *Marine Drugs, 16*(7). doi:10.3390/md16070244

Sargent, M. (Editor). (2013). Guide to achieving reliable quantitative LC-MS measurements, RSC Analytical Methods Committee, 2013. ISBN 978-0-948926-27-3. Available at https://www.rsc.org/images/AMC%20LCMS%20Guide_tcm18-240030.pdf

Schupp, T., Toupet, C., & Divers, M. (1988). Cloning and expression of two genes of *Streptomyces pilosus* involved in the biosynthesis of the siderophore desferrioxamine B. *Gene, 64*(2), 179-188. doi:10.1016/0378-1119(88)90333-2

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069. doi:10.1093/bioinformatics/btu153

Seipke, R. F., Barke, J., Brearley, C., Hill, L., Yu, D. W., Goss, R. J., & Hutchings, M. I. (2011). A single *Streptomyces* symbiont makes multiple antifungals to support the

fungus farming ant *Acromyrmex octospinosus*. *PLoS One, 6*(8), e22028. doi:10.1371/journal.pone.0022028

Seipke, R. F., Patrick, E., & Hutchings, M. I. (2014). Regulation of antimycin biosynthesis by the orphan ECF RNA polymerase sigma factor sigma (AntA.). *PeerJ, 2*, e253. doi:10.7717/peerj.253

Sharom, F. J., Yu, X., Chu, J. W., & Doige, C. A. (1995). Characterization of the ATPase activity of P-glycoprotein from multidrug-resistant Chinese hamster ovary cells. *Biochem J, 308 (Pt 2)*, 381-390. doi:10.1042/bj3080381

Song, Y., Li, Q., Liu, X., Chen, Y., Zhang, Y., Sun, A., Zhang, W., Zhang, J., Ju, J. (2014). Cyclic Hexapeptides from the Deep South China Sea-Derived *Streptomyces scopuliridis* SCSIO ZJ46 Active Against Pathogenic Gram-Positive Bacteria. *J Nat Prod, 77*(8), 1937-1941. doi:10.1021/np500399v

Sproule, A., Correa, H., Decken, A., Haltli, B., Berrue, F., Overy, D. P., & Kerr, R. G. (2019). Terrosamycins A and B, Bioactive Polyether Ionophores from *Streptomyces sp.* RKND004 from Prince Edward Island Sediment. *Marine Drugs, 17*(6). doi:10.3390/md17060347

Stierle, A. C., Cardellina, J. H., & Strobel, G. A. (1988). Maculosin, a host-specific phytotoxin for spotted knapweed from *Alternaria alternata*. *Proc Natl Acad Sci U S A, 85*(21), 8008-8011. doi:10.1073/pnas.85.21.8008

Sujatha, P., Bapi Raju, K. V., & Ramana, T. (2005). Studies on a new marine streptomycete BT-408 producing polyketide antibiotic SBR-22 effective against methicillin resistant *Staphylococcus aureus*. *Microbiol Res, 160*(2), 119-126. doi:10.1016/j.micres.2004.10.006

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res, 44*(14), 6614-6624. doi:10.1093/nar/gkw569

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crusemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderon, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., P, C. A. B., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V.,

Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodriguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. F., Alexandrov, T., Litaudon, M., Wolfender, J. L. Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Muller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B. O., Pogliano, K., Linington, R. G., Gutierrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C., Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol, 34*(8), 828-837. doi:10.1038/nbt.3597

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. (2012). Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A, 109*(26), E1743-1752. doi:10.1073/pnas.1203689109

Wilson K. (2001). Preparation of genomic DNA from bacteria. *Current protocols in molecular biology*, *Chapter 2*, . https://doi.org/10.1002/0471142727.mb0204s56

Wong, S. M., Musza, L. L., Kydd, G. C., Kullnig, R., Gillum, A. M., & Cooper, R. (1993). Fiscalins: new substance P inhibitors produced by the fungus Neosartorya fischeri. Taxonomy, fermentation, structures, and biological properties. *J Antibiot (Tokyo), 46*(4), 545-553. doi:10.7164/antibiotics.46.545

Woo, J. T., Ohba, Y., Tagami, K., Sumitani, K., Yamaguchi, K., & Tsuji, T. (1996).

Concanamycin B, a vacuolar H(+)-ATPase specific inhibitor suppresses bone resorption in vitro. *Biol Pharm Bull, 19*(2), 297-299. doi:10.1248/bpb.19.297

Yilla, M., Tan, A., Ito, K., Miwa, K., & Ploegh, H. L. (1993). Involvement of the vacuolar H(+)-ATPases in the secretory pathway of HepG2 cells. *J Biol Chem, 268*(25), 19092-19100.

Ying, Y. M., Huang, L., Tian, T., Li, C. Y., Wang, S. L., Ma, L. F., . . . Zhan, Z. J. (2018). Studies on the Chemical Diversities of Secondary Metabolites Produced by Neosartorya fischeri via the OSMAC Method. *Molecules, 23*(11). doi:10.3390/molecules23112772

Zotchev, S. B. (2012). Marine actinomycetes as an emerging resource for the drug development pipelines. *J Biotechnol, 158*(4), 168-175. doi:10.1016/j.jbiotec.2011.06.002

4 Chapter 4: Detection of bioactive secondary metabolites, from the marine *Streptomyces* SM9 crude extracts.

Abstract

Microbes associated with marine sponges are exposed to highly competitive environments both physiologically and nutritionally which is likely to promote the production of novel secondary metabolites. This study focuses on the detection of potential bioactive secondary metabolites from a *Streptomyces* isolate SM9, using the two complementary methods of bioassays-guided dereplication and bioactive molecular networking, with the goal of discovering novel bioactive metabolites with potential biopharmaceutical utility. SM9 was subjected to different growth conditions using the OSMAC (One Strain MAny Compounds) approach for the production of secondary metabolites. Cell free supernatants were chemically extracted using Liquid Liquid Extraction (LLE) with ethyl-acetate. Crude extracts containing the potential secondary metabolites from both aqueous and organic phases were then screened for anti-microbial activities. The antimicrobial screening of the crude extracts demonstrated that they possessed secondary metabolites with anti-fungal activity against *C. glabrata* CBS138; anti-inflammatory activity against LPS induced TNF-α production in Thp-1 cells, with an inhibition of 44-45%; and an anti-cancer activity against the human melanoma cell line A2058 allowing a percentage of 10-41% of cell survival. These activities varied depending on the culturing conditions; with , anti-cancer activity being observed following growth of SM9 in SGG and SM media; a strong antifungal activity when cultured in M19, M400, SGG, OM and SM media; and partial anti-inflammatory activity when cultured in M19 and SM media.

Bioinformatics and metabolomics analysis allowed the mapping of these bioactivities through the establishment of a bioactive molecular network and revealed the presence of approximately 19 promising molecules from a total of 177, with high bioactive scores and low p_value. Bioactivity scores from 60% to 89% for potential anti-cancer activity against the human melanoma cell line A2058; 62% to 88% for potential anti-inflammatory activity against the LPS induced TNF-α production in Thp-1 cells; and 60% to 64% for potential anti-fungal activity against *C. glabrata* CBS138 were observed. Some promising molecules were matched to our previous LC-MS dataset, through their $[M+H]^+$ masses, such as, Desferrioxamine E, pro-thr-ala-ile, val-glu-pro-pro and val-ser-pro-pro metabolites. Other molecules were also matched to potential new metabolites from our previous LC-MS dataset, through their $[M+H]^+$ masses.

Additional antioxidant activities (including radical scavenging activity) and antimicrobial activities against *C. glabratra* CBS138, *S.aureus* 6538p, *A.baumannii* AB13, *B.metallica* LMG 24068 and *P.aeruginosa* PA01 were also observed.

Introduction

We have already described in chapter 1 and 2 how changes in fermentation parameters such as nutrient source media, following the OSMAC approach, results in the production of bioactive secondary metabolites and potentially new metabolites (Reen *et al*., 2015; Pan *et al*., 2019).

In this chapter, we focused on the use of both bioassay guided and molecular networking approaches to further investigate the bioactivity profile of the SM9 isolate. This involved on the one hand employing bioactivity assays, together with refocusing our efforts on re analysing the chemical and molecular network data obtained from the crude extracts previously described in Chapter 2; with the overall aim of discovering both potential new and bioactive secondary metabolites being produced in this strain.

In the context of bioassay guided dereplication, we investigated the biological activities of our SM9 crude extracts by subjecting each sample to a diverse range of bioassays such as a) anti-fungal assays against *C. glabrata* CBS138 (a strain which had been previously used in several similar studies (Xu *et al*., 2016 ; Roetzer *et al*., 2011), b) anti-cancer assays against the human melanoma cell line A2058, which had previously been used in similar screens on extracts from *Streptomyces* sp. R(K44 Fang *et al*., 2020) and c) anti-inflammatory assays measured in Thp-1 cells which had previously been used in the isolation of Barettin from extracts of the marine sponge *Geodia barrette* (Lind *et al*., 2013). The overall aim was to not only identify the most promising extracts containing high levels of biological activity from the *Streptomyces* SM9 crude extracts, but also by employing dereplication strategies, together with the molecular networking to identify and select the most promising potential bioactive molecules of interest (Figure 24).

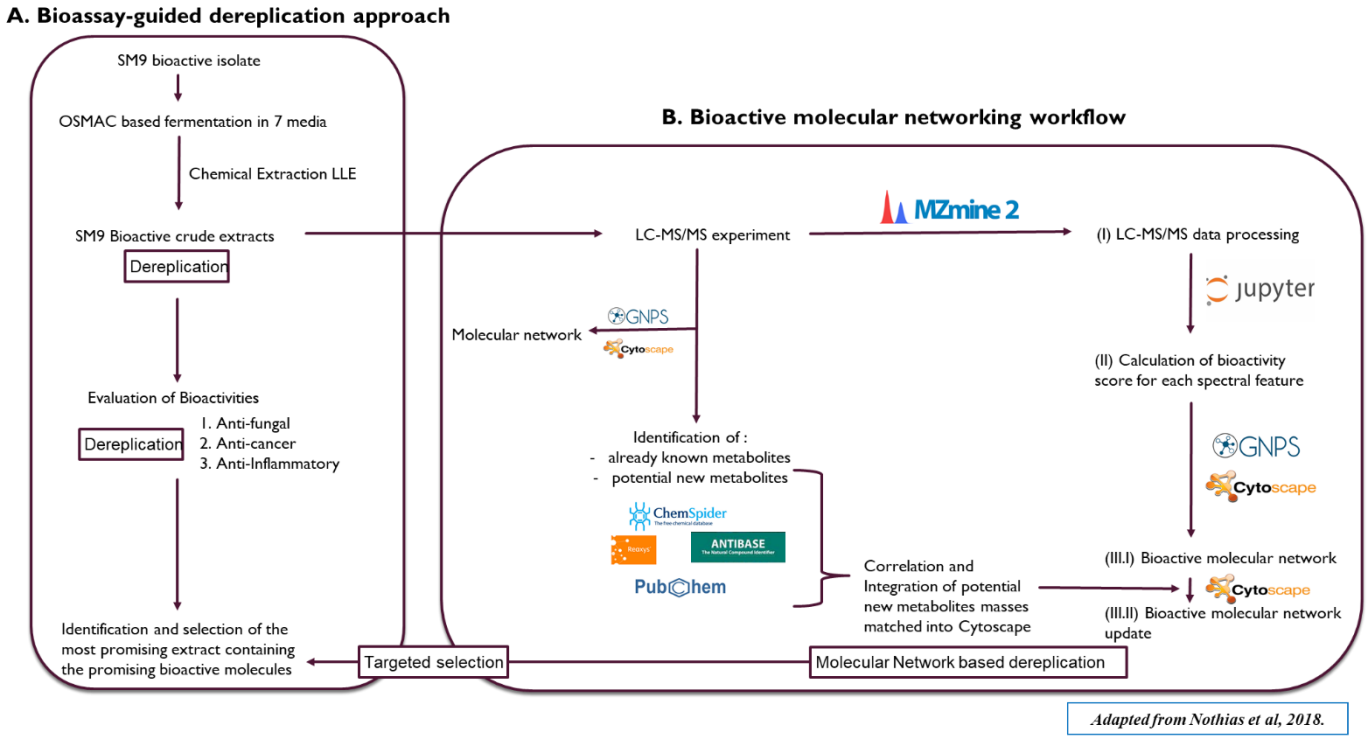


Figure 24: Bioactive natural discovery workflow, involving a bioassay-guided dereplication (A) and bioactive molecular networking (B).

Additional information: Chemical extraction LLE=Liquid liquid extraction; Evaluation of bioactivities 1. Antifungal against *C. glabrata* CBS138; Anti-cancer against the human melanoma cell line A2058 and anti-inflammatory assay measured in Thp-1 cells. Bioactive molecular networked updated consist of the correlation of the previous potential new metabolites masses detected in our previous study (chapter 2) with the masses of the bioactive molecular network. When the masses were matched, the value "no hit" was integrated into Cytoscape to allow direct visualization in the network of molecules already known/suspected (from chapter 2 study) to be potential new metabolites (using metabolome mining and Cytoscape).

In the context of bioactive molecular networking, we used both bioinformatics and metabolomics tools, such as such as MZmine2, the Jupyter notebook, the GNPS web platform, together with Cytoscape and metabolome "mining" to create the bioactive molecular networks. Since we did not know whether or not each activity was related to the other activities, a bioactive molecular network was generated for each of the individual bioactivities screened for in the SM9 crude extracts. A total of 3 bioactivity maps involving anti-fungal activity, anti-cancer activity and anti-inflammatory activities were created following the process previously outlined by the Dorrestein group (*Nothias et al*, 2018). Following this the maps were

subsequently employed together with the targeted selection of the most bioactive molecules of interest, to identify the most promising extract (Figure 24).

Current approaches that are often employed in natural product discovery programs can be problematic given despite good bioassay results in the initial extract, the active compound(s) may not be isolated during the subsequent bioassay-guided purification (*Nothias et al*., 2018). To help overcome this the Dorrestein group introduced the concept of bioactive molecular networking to identify candidate active molecules directly from fractionated bioactive extracts. This is the approach that we chose to use, which typically involved a three step process (Figure 24). The first step (I) involves processing LC-MS/MS spectral data with the MZmine2 toolbox. This step allows the detection and relative quantification of LC-MS/MS spectral features (ions) across the chromatographic extracts (*Nothias et al*, 2018). The second step (II) involves calculation of a bioactive score using the Pearson correlation between feature intensity across sample and the bioactivity level associated with each sample (*Nothias et al*, 2018). The score is calculated following the Jupyter notebook script using R language, which is available on the open source web application website http://jupyter.org/. The final step (III) involves two parts, with the first part (III.I), entailing an analysis of the MS/MS data on the GNPS web-platform using Cytoscape to visualize the corresponding molecular network. The objective of this step is to annotate molecules that are detected using both spectral library annotations and the network topology, and in particular those molecules with significant predicted bioactivity scores (*Nothias et al*, 2018).

Finally, the last part (III.II) involves correlating and annotating the bioactive molecular networks with our previous data regarding potential new metabolites masses that match with our network, using metabolome mining and Cytoscape by manually incorporating the value "no hit" in the related exact masses, which were previously identify as potentially new

metabolite masses, in our Cytoscape network. This allows the direct visualization of some of the potential new metabolites identified from Chapter 2, into this bioactive molecular network.

For this workflow, adapted from *Nothias et al*, 2018., the data curation relies on two complementary approaches; firstly a targeted approach, where prior knowledge can be used, thus we targeted a set of bioactive extracts (*Nothias et al.*, 2018). Second an untargeted approach, by looking for consistent patterns of potential bioactive candidates within the network, such as individual molecule with a high bioactivity score and or cluster with high frequency of bioactive candidate (*Nothias et al.*, 2018). This new concept of bioactive molecular networking (BMN) using molecular networking and dereplication tools (LC-MS/MS) to detect candidate active molecules is gaining increased interest in the field of natural products chemistry as evidenced by another recent study, by the Tasdemir group (Buedenbender *et al*., 2020). In this study dereplication tools coupled with bioactive molecular networking (BMN) was used to identify cluster and individual compounds responsible for antimicrobial activity, involving the identification of a galactolipids bioactive cluster, and the subsequent isolation and identification of several bioactive compounds including 6 monogalactosyldiacylglycerol derivatives that were active against *Staphylococcus aureus* from the alga *Fucus vesiculosus* (Buedenbender et al., 2020).

Having established the optimum media for SM9 to produce secondary metabolites, potential new metabolites and also bioactive metabolites, the strain was grown up under these conditions on a larger scale to facilitate the generation of extracts for subsequent further anti-microbial screening

Results and Discussion

4.1 Bioactive molecular network approach

4.1.1 Bioactivities of crude extracts

All the SM9 crude extract samples generated from the different culture conditions were screened for: anti-fungal activity (Table 9), anti-cancer activity (Table 10) and anti-inflammatory activity (Table 11).

Strain	Media of fermentation	Extract phases	Result
SM9	M19	aqueous phase	A*
		organic phase	A
	M400	aqueous phase	A *
		organic phase	A
	MMM	aqueous phase	I
		organic phase	I
	OM	aqueous phase	A
		organic phase	I
	SGG	aqueous phase	I
		organic phase	A
	SM	aqueous phase	I
		organic phase	A
	SYP	aqueous phase	I
		organic phase	I

Table 9: Table of the NCCLS assay (National Committee for Clinical Laboratory Standards, broth microdilution method) against *C. glabrata* CBS138.

A= active, A*= very active, I= inactive.

Media	Plate n°	Well ID	Extract phases	Activity	Activity average	Units	Result
SM9_M19	1	A06	aqueous phase	82.3888916051435	82	PCT_SURVIVAL	I
	1	A07	organic phase	71.3978389478316	71	PCT_SURVIVAL	I
SM9_M400	1	B06	aqueous phase	70.3613161883342	70	PCT_SURVIVAL	I
	1	B07	organic phase	74.0918202708649	74	PCT_SURVIVAL	I
SM9_MMM	1	C06	aqueous phase	86.202513078766	86	PCT_SURVIVAL	I
	1	C07	organic phase	72.9526230870777	73	PCT_SURVIVAL	I
SM9_OM	1	D06	aqueous phase	73.6713440571065	74	PCT_SURVIVAL	I
	1	D07	organic phase	69.0901090304601	69	PCT_SURVIVAL	I
SM9_SGG	1	E06	aqueous phase	41.3289004058084	41	PCT_SURVIVAL	A
	1	E07	organic phase	59.2089180071383	59	PCT_SURVIVAL	U
SM9_SM	1	F06	aqueous phase	59.8249645528773	60	PCT_SURVIVAL	U
	1	F07	organic phase	10.4190094362685	10	PCT_SURVIVAL	A
SM9_SYP	1	G06	aqueous phase	81.9977509411822	82	PCT_SURVIVAL	I
	1	G07	organic phase	74.0967095291644	74	PCT_SURVIVAL	I
Blank_M19	2	A09	aqueous phase	101.571235526351	102	PCT_SURVIVAL	I
	2	A08	organic phase	102.901353965184	103	PCT_SURVIVAL	I
Blank_M400	2	B09	aqueous phase	94.7669643093718	95	PCT_SURVIVAL	I
	2	B08	organic phase	96.3885429638854	96	PCT_SURVIVAL	I
Blank_MMM	2	C09	aqueous phase	98.699027582735	99	PCT_SURVIVAL	I
	2	C08	organic phase	97.2470257809809	97	PCT_SURVIVAL	I
Blank_OM	2	D09	aqueous phase	94.4013142206089	94	PCT_SURVIVAL	I
	2	D08	organic phase	97.0668503749238	97	PCT_SURVIVAL	I
Blank_SGG	2	E09	aqueous phase	95.1220158448372	95	PCT_SURVIVAL	I
	2	E08	organic phase	103.553164992978	104	PCT_SURVIVAL	I
Blank_SM	2	F09	aqueous phase	97.4536975702817	97	PCT_SURVIVAL	I
	2	F08	organic phase	102.143557404414	102	PCT_SURVIVAL	I
Blank_SYP	2	G09	aqueous phase	94.5125990302324	95	PCT_SURVIVAL	I
	2	G08	organic phase	92.4511804138735	92	PCT_SURVIVAL	I

Table 10: Assay category: anti-cancers, assay name: A2058_MTT.

Extracts were tested against the human melanoma cell line A2058 (ATCC-CRL-11147) at a concentration of 100 µg /mL. Extracts are considered to be active if the cell survival is less than 50 %. A= active; U= unclear but can be considered partly active; I= inactive.

Media	Plate n°	Well ID	Extract phases	Activity	Activity average	Units	Result
SM9_M19	1	A06	aqueous phase	-3.93736109764762	-4	PCT_INHIBITION	I
	1	A07	organic phase	45.1785325947576	45	PCT_INHIBITION	U
SM9_M400	1	B06	aqueous phase	17.5064048803915	18	PCT_INHIBITION	I
	1	B07	organic phase	4.13486115270618	4	PCT_INHIBITION	I
SM9_MMM	1	C06	aqueous phase	13.0792436461986	13	PCT_INHIBITION	I
	1	C07	organic phase	-8.45458235694718	-8	PCT_INHIBITION	I
SM9_OM	1	D06	aqueous phase	-4.28812119543163	-4	PCT_INHIBITION	I
	1	D07	organic phase	-11.654083248896	-12	PCT_INHIBITION	I
SM9_SGG	1	E06	aqueous phase	-28.8413280403128	-29	PCT_INHIBITION	I
	1	E07	organic phase	-7.95214221687819	-8	PCT_INHIBITION	I
SM9_SM	1	F06	aqueous phase	-9.35044260669287	-9	PCT_INHIBITION	I
	1	F07	organic phase	44.1641723119768	44	PCT_INHIBITION	U
SM9_SYP	1	G06	aqueous phase	-6.49696181120667	-6	PCT_INHIBITION	I
	1	G07	organic phase	-15.393944291485	-15	PCT_INHIBITION	I
Blank_M19	2	A09	aqueous phase	-15.6567923974916	-16	PCT_INHIBITION	I
	2	A08	organic phase	-2.10002212103608	-2	PCT_INHIBITION	I
Blank_M400	2	B09	aqueous phase	-10.9010123355822	-11	PCT_INHIBITION	I
	2	B08	organic phase	-13.8862908525537	-14	PCT_INHIBITION	I
Blank_MMM	2	C09	aqueous phase	0.119051147794806	0	PCT_INHIBITION	I
	2	C08	organic phase	-5.46933472085097	-5	PCT_INHIBITION	I
Blank_OM	2	D09	aqueous phase	5.05340793139645	5	PCT_INHIBITION	I
	2	D08	organic phase	2.20544491827684	2	PCT_INHIBITION	I
Blank_SGG	2	E09	aqueous phase	-4.52394361620365	-5	PCT_INHIBITION	I
	2	E08	organic phase	6.7976619660846	7	PCT_INHIBITION	I
Blank_SM	2	F09	aqueous phase	-3.43368573639833	-3	PCT_INHIBITION	I
	2	F08	organic phase	-4.33920233442994	-4	PCT_INHIBITION	I
Blank_SYP	2	G09	aqueous phase	4.0383402501984	4	PCT_INHIBITION	I
	2	G08	organic phase	-2.37412139386205	-2	PCT_INHIBITION	I

Table 11: Assay category: anti-inflammatory, Assay name: TNF_THPAIF.

Concentration 100 µg /ml. Extracts are considered to be active if the inhibition is at least 50 %. A= active; U=unclear but can be considered partly active; I= inactive.

Table 9 shows the anti-fungal activity results, with SM9 showing anti-fungal activity against *C. glabrata* CBS138 in crude extracts from SGG, SM, M19, M400 media and in the organic phase extracts from aqueous extracts from the M19, M400 and OM media . This suggests that SM9 produces bioactive anti-fungal metabolites which are active against *C. glabrata* CBS138 when grown in OM, M19, M400, SGG and SM culture media (Table 9). The M19 and M400 extracts displayed the highest bioactivity, thus it is clear that the optimum nutrient source for the SM9 isolate to produce potential antifungal metabolites that are active against C. *glabrata* CBS138, are both the M19 and M400 media, as extracts from these growth media displayed the best anti-fungal activity (Table 9).

Table 10 shows the result obtained from the anti-cancer activity screen, with both SGG and SM extracts containing anti-cancer metabolites with anti-cancer activity using the human melanoma cell line A2058 assay (Table 10). In this assay system, extracts are considered to be active if the cell survival rate is <50%. Aqueous extracts from SM9 grown on SGG media resulted in survival rates of 41%, while organic phase extracts from SM9 grown on SM medium resulted in survival rates of 10% (Table 10). Overall, an average of 50% and 35% of cell survival rates were observed from SGG and SM crude extracts respectively, calculated from both organic phase and aqueous phase extracts together. This indicates that both the SGG and SM growth media are optimal for the production of anti-cancer metabolites by SM9.

Crude extracts from SM9 grown in M19 and SM media displayed partial anti-inflammatory activity by inhibiting LPS induced TNF-α production in Thp-1 cells, a monocytic cell line (Table 11). This is reflected in a total of 45% and 44% inhibition of LPS induced TNF-α production in Thp-1 cells in the crude extracts from both M19 and SM organic phase crude extracts respectively (Table 11). This indicates that the SM9 isolate produces bioactive metabolites with partial anti-inflammatory activity when cultured in both M19 and SM culture

media. Thus, both M19 and SM media appear to be optimal to induce production of metabolites with potential anti-inflammatory activity in SM9.

Taking together, these bioactivity results show that in some instances there is a lack of uniformity with respect to the observed activities from the extracts. For example, SM and SGG extracts were the most active in terms of anti-cancer activity, while the M19 and M400 extracts were the most active in terms of anti-fungal activity (Table 12).

Media of fermentation	Crude extract sample	Number of potential new metabolites detected	Relevant tested bioactivities		
			Anti-fungal against *C. glabrata* CBS138	Anti-cancer against human melanoma cell line A2058	Anti-inflammatory against LPS induced TNF-α production in Thp-1 cells
M19	organic phase	7			
	aqueous phase	3			
M400	organic phase	10			
	aqueous phase	3			
MMM	organic phase	5			
	aqueous phase	3			
OM	organic phase	7			
	aqueous phase	7			
SGG	organic phase	3			
	aqueous phase	3			
SM	organic phase	2			
	aqueous phase	9			
SYP	organic phase	11			
	aqueous phase	8			

Legend	Visualization	Activity level
Bioactivities		Inactive
		Partly active
		Active
		Very active
Chemistry	Visualization	Definition
		>5 potential new metabolites

Table 12: Composite table of SM9 crude extracts bioactivities results, together with the number of potential new metabolites detected by previous LC-MS/MS analysis from chapter 2.

In addition, when focusing on the extracts containing the highest number of potential new metabolites, together with the observed bioactivities, some of the extracts that contained the highest number of potential new metabolites were those which displayed less activity. For example, SYP and OM organic phase extracts contained 11 and 7 potential new metabolites respectively; but no activity was observed in any of the bioassays tested (Table 12). In contrast, the extracts which displayed the best activity appeared to contain the lowest number of potential new metabolites. For example, the SGG and SM extracts which displayed the highest levels of bioactivity had the lowest number of potential new metabolites (Table 12). However despite these trends some of the other extract such as those from the M400 and M19 growth media, showed high number of potential new metabolites and good levels of bioactivity. (Table 12).

Based on this the media extracts which were the most interesting in term of both production of potential new metabolites and metabolites with potential bioactivity from the SM9 isolate were identified as : M19 (organic phase), M400 (organic phase) and OM (aqueous phase). Thus, the media extract which was selected for large scale fermentation for the production of potential new and bioactive metabolite, for further analysis was the M19-organic phase extract.

4.1.2 Bioactivity mapping of the anti-cancer, anti-inflammatory and anti-fungal activities

4.1.2.1 Anti-cancer activity mapping

Regarding anti-cancer activity mapping, from Figure 25, it is clear that a total of 13 of the 177 molecules (=7.34%) were predicted as having a statistically significant anti-cancer activity score (r> 0.6) against the human melanoma cell line A2058. Twelve of these 177 molecules (6.79%) were matched to potential new metabolites already predicted by the manual de-

replication dataset; and 2 of the 177 molecules (1.13%) were predicted as having both a statistically significant anti-cancer bioactivity score and were matched to the potential new metabolites (Figure 25a). Although, the possibility does exist however that some of the bioactive molecules could be new metabolites and not be matched to our LC-MS dataset.

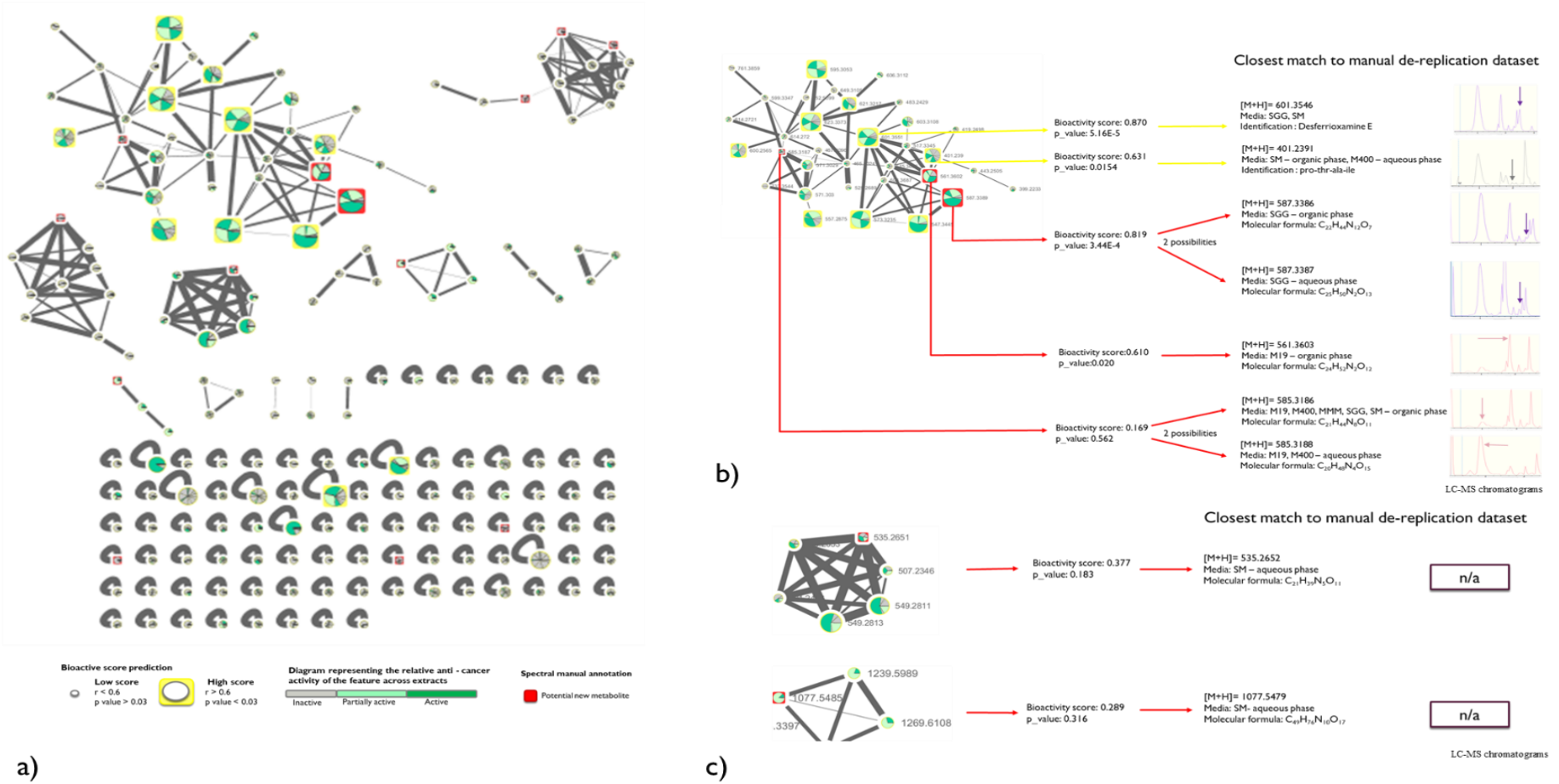


Figure 25: Bioactivity mapping of the anti-cancer activity against the human melanoma cell line A2058, from SM9 crude extracts, with legend of the bioactive molecular network (a). View of the bioactive clusters containing the bioactive metabolites with significant bioactivity score and/or potential new metabolites (b and c).

The bioactivity scoring is visualized directly in the molecular network with the node size proportionally representing the molecule predicted bioactivity score. The bioactivity score was defined herein as the Pearson correlation coefficient (r) between the molecule relative abundance derived from the LC-MS peak (area

under the curve) and the selectivity index (SI) value obtained from the evaluation of the extract anti-cancer activity. The largest nodes represent molecules with statistically significant bioactivity score [strong Pearson correlation value ($r > 0.6$) and a significance (FDR corrected, p value <0.03) after Bonferroni correction]. In addition, the relative abundance of detected molecule in extract can be visualized in each node with a pie chart diagram [colours correspond to the fraction bioactivity level]. Some of the metabolite were manually matched to the LC-MS and manual de-replication result dataset by metabolome mining through there $[M+H]^+$ masses. Only closest masses up to 0.0006 of differences were matched. n/a= not available. Additional information: the following molecules 547.3445, 557.2875, 573.3235, 595.3053, 600.2565, 621.3217 and 623.3373, have shown high anti-cancer bioactivity scores of 0.737, 0.689, 0.885, 0.868, 0.628, 0.601 and 0.806 respectively.

When considering the bioactive cluster containing the metabolites with a significant bioactivity score in Figure 25b, it is clear that among this cluster, 11 metabolites displayed a low p_value and a high bioactivity score in the human melanoma cell line A2058, with bioactivity score from 0.610 to 0.870, which represent a bioactivity level of approximately 61% to 87% (Figure 25b). In addition, 3 metabolites in this cluster were linked as being potential new metabolites (from a correlation with the LC-MS dataset) including 2 metabolites which were predicted as both potential new metabolites and having statistically significant anti-cancer bioactivity scores.

During this bioinformatics and metabolomics analysis, the interesting potentially new metabolites were represented in red cube in the bioactive molecular networks by metabolome mining after correlation of their $[M+H]^+$ masses to previous LC-MS/MS analysis (Figure 25 b, c). However, while the not-red metabolites in the bioactive molecular network, which were not linked to our potential new metabolites list may still be potentially new metabolite as well as already known metabolites. Only chemical characterization by NMR of these metabolites will confirm if this is the case or not.

Furthermore, some of the metabolites were matched to the previous LC-MS and manual de-replication dataset through their $[M+H]^+$ masses, with only closest masses up to 0.0006 of differences being matched. Two interesting molecules with high significant bioactivity scores of 0.870 and 0.631 were matched to the previously identified Desferrioxamine E and pro-thr-ala-ile metabolites from the LC-MS and manual de-replication dataset (Figure 25b; also in Table 7 from Chapter 2). Another interesting molecule, with a 0.610 bioactivity score of 0.610 and a 0.020 p_value, was matched to a potential new metabolite which was only detected in the M19 organic phase extract, had the following molecular formula $C_{24}H_{52}N_2O_{12}$ (Figure 25b). The two remaining molecules with bioactivities scores of 0.819 and 0.169 were matched to several possible potential new metabolites detected with LC-MS and the manual-de-

replication dataset. Indeed, their $[M+H]^+$ masses could be linked to two other $[M+H]^+$ masses from our LC-MS and manual dereplication dataset. For example, the bioactive molecule with $[M+H]^+$ masse of 587.3389 in the bioactive molecular network (with the high bioactivity score of 0.819) could not only be linked to the potential new metabolite with the $[M+H]^+$ mass of 587.3386 and the molecular formula $C_{22}H_{44}N_{12}O_7$, but could also be linked to the another potential new metabolite with the $[M+H]^+$ mass of 587.3387 and the molecular formula $C_{25}H_{50}N_2O_{13}$ (Figure 25b). In the same way, the bioactive molecule with the $[M+H]^+$ mass of 585.3187 in the bioactive molecular network (with low bioactivity score of 0.169) could be linked to both $[M+H]^+$ masses of 585.3186 and 585.3187 from both metabolites with the molecular formula $C_{21}H_{44}N_8O_{11}$ and $C_{20}H_{48}N_4O_{15}$ respectively (Figure 25b). Only subsequent full chemical characterisation using NMR analysis will allow the definitive identification of the metabolite, thus in the meantime both of the aforementioned options are possible.

4.1.2.2 Anti-inflammatory mapping

The anti-inflammatory activity mapping is presented Figure 26. This resulted in the identification of a total of 14 of the 177 molecules being predicted as having a high bioactivity score ($r > 0.6$) with an acceptable p_value <0.03 (7.91%). This included 6 molecule predicted as having a very high and statistically significant anti-inflammatory activity score ($r > 0.6$) and a very low p_value <0.003 (3.39%). In addition 13 of the 177 molecules (7.34%) were matched to the potential new metabolites already predicted by the manual de-replication dataset but which did not have a significant bioactivity score (low r score) (Figure 26a). When looking closely at the two bioactive clusters containing the metabolites with a significant bioactivity score (Figure 26b) it was clear that among these clusters, there were 4 metabolites which displayed significant high bioactivity scores and low p_values against the LPS induced TNF-

α production in Thp-1 cells, with bioactivity scores ranging from 0.732 to 0.881 (which represents a bioactivity level of approximately 73% to 88%) making them very interesting metabolites to further analysis (Figure 26b). In addition, 2 metabolites predicted as having high bioactivity scores of 0.620 and 0.655 but which also had high p-values (p_value >0.003) could be linked to our LC-MS and manual de-replication dataset, through their $[M+H]^+$ masses, were matched and identified as pro-glu-val-val and val-ser-pro-pro metabolites respectively (Figure 26b).

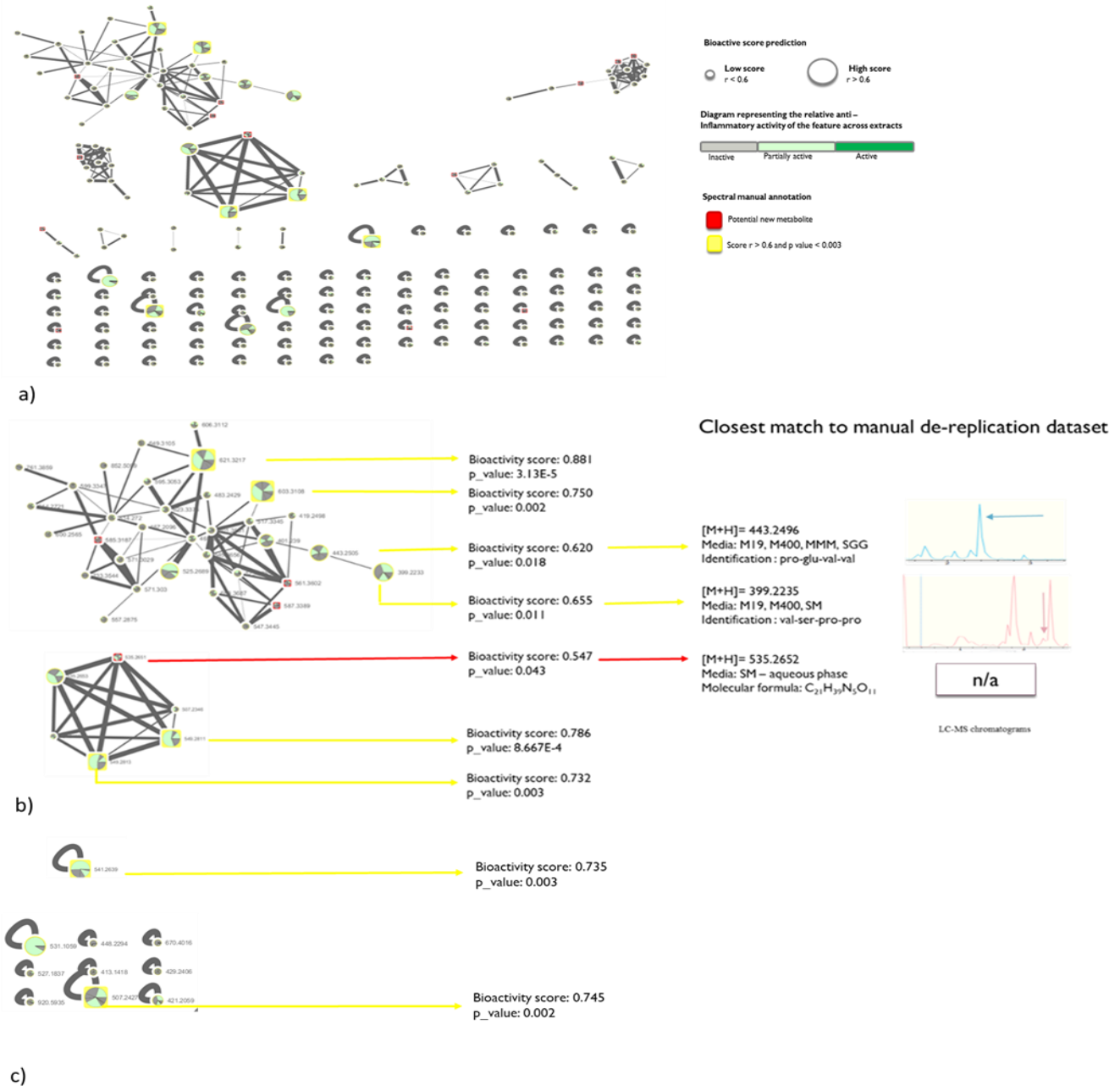


Figure 26: Bioactivity mapping of the anti-inflammatory activity of SM9 extracts against the LPS induced TNF-α production in Thp-1 cells, with legend of bioactive molecular network (a). View of the bioactive clusters containing the bioactive molecules with significant bioactive score (b) as well as isolated bioactive molecules in the network (c).

The bioactivity scoring is visualized directly in the molecular network with the node size proportionally representing the molecule predicted bioactivity score. The bioactivity score was defined herein as the Pearson correlation coefficient (r) between the molecule relative abundance derived from the LC-MS peak (area under the curve) and the selectivity index (SI) value obtained from the evaluation of the extract anti-cancer activity. The largest nodes represent molecules with statistically significant bioactivity score [strong Pearson correlation value ($r > 0.6$) and a significance (FDR corrected, p value <0.03) after Bonferroni correction]. In addition, the relative abundance of detected molecule in extract can be visualized in each node with a pie chart diagram [colours correspond to the fraction bioactivity level]. Some of the metabolite were manually matched to the LC-MS and manual de-replication result dataset through there $[M+H]^+$ masses. Only closest masses up to 0.0006 of differences were matched. n/a= not available.

Among the bioactive molecules isolated in the network, with significant anti-inflammatory activity, 2 molecules with $[M+H]^+$ masses of 541.2639 and 507.2427, with predicted high bioactivity score of 0.735 and 0.745 and with very low p_value were detected but could not be linked to our previous LC-MS dataset. Therefore, they may still be potentially new metabolite as well as already known metabolites and they remain interesting molecules for further investigations (Figure 26c).

4.1.2.3 Anti-fungal activity mapping

Regarding the anti-fungal activity mapping, which is represented in Figure 27, 2 of the 177 molecules were predicted as having a high and significant bioactivity score against *Candida glabrata* CBS138, with an r> 0.6 and a p_value <0.03, representing (1.13% of the total predicted molecules). In addition, 13 of the 177 molecules (7.34%) were matched to the potential new metabolites already predicted by the manual de-replication dataset but which did not have a significant bioactivity score and high p_value (Figure 27a). Although, as previously mentioned for the anti-cancer activities some of these bioactive molecules could be new metabolites which have no matches in our LC-MS dataset.

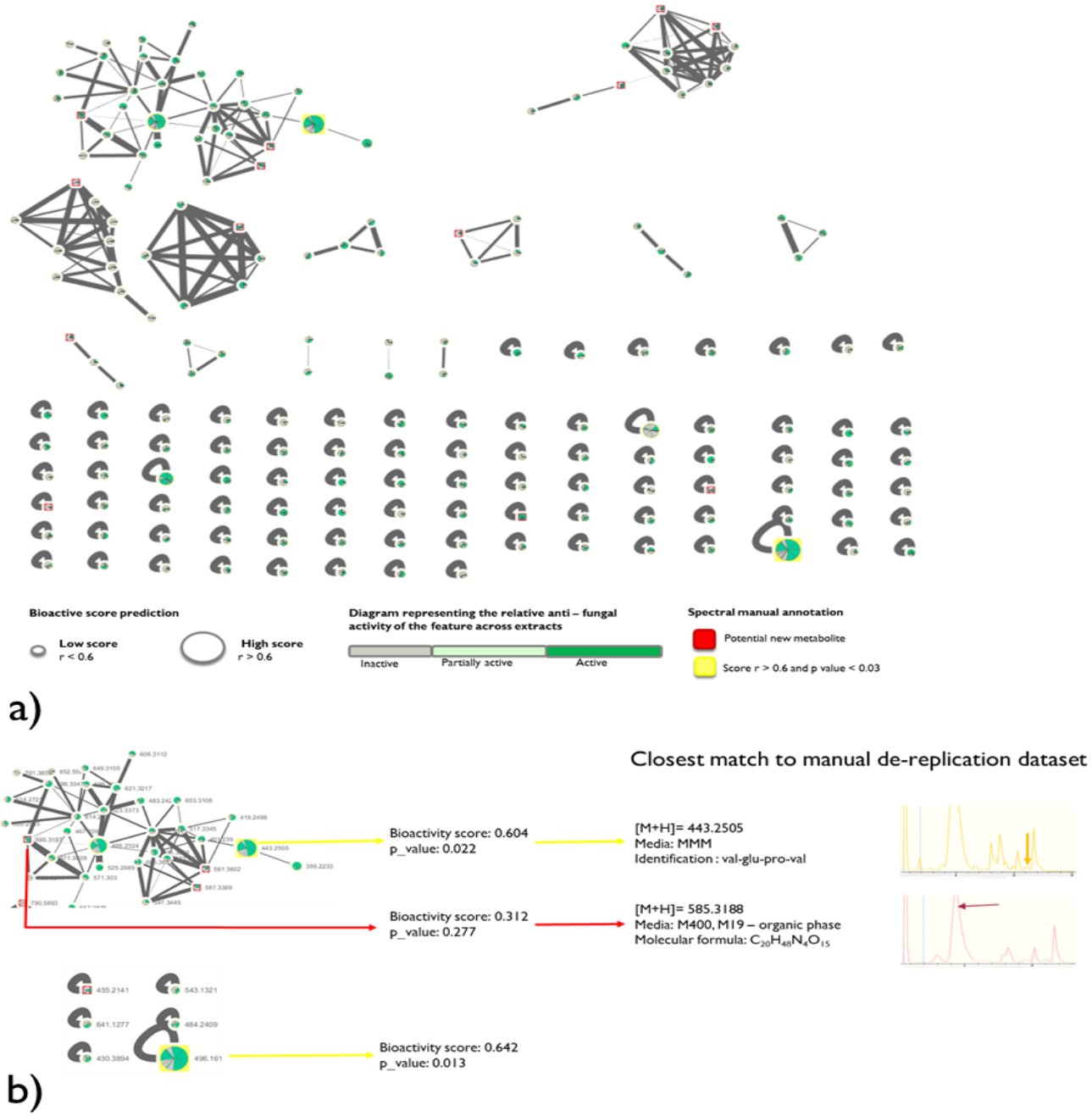


Figure 27: Bioactivity mapping of the anti-fungal activity of SM9 extracts against *C. glabrata* CBS138, with legend of bioactive molecular network (a). View of the bioactive cluster containing the bioactive molecule with significant bioactive score as well as isolated bioactive molecule in the network (b).

The bioactivity scoring is visualized directly in the molecular network with the node size proportionally representing the molecule predicted bioactivity score. The bioactivity score was defined herein as the Pearson correlation coefficient (r) between the molecule relative abundance derived from the LC-MS peak (area under the curve) and the selectivity index (SI) value obtained from the evaluation of the extract anti-cancer activity. The largest nodes represent molecules with statistically significant bioactivity score [strong Pearson correlation value ($r > 0.6$) and a significance (FDR corrected, p value <0.03) after Bonferroni correction]. In addition, the relative abundance of detected molecule in extract can be visualized in each node with a pie chart diagram [colours correspond to the fraction bioactivity level]. Some of the metabolite were manually matched to the LC-MS and manual de-replication result dataset through there $[M+H]^+$ masses. Only closest masses up to 0.0006 of differences were matched. n/a= not available.

When focusing our attention on these bioactive molecules 443.2305 and 496.1610, we can see that their bioactivity score is 0.604 and 0.642 representing approximate 60% and 64% of bioactivity against *Candida glabrata* CBS138, respectively. Furthermore, the 443.2305 molecule could be matched to the val-glu-pro-val metabolite from our LC-MS dataset, through its $[M+H]^+$ mass.

4.1.3 Bioactive molecules of interest

When combining all three bioactivities, and focusing on each individual molecule of interest, it is clear that the bioactivity mappings of SM9 extracts has highlighted the potential of 19 promising molecules of the 177 molecules analysed, in term of anti-cancer, anti-inflammatory and/or anti-fungal activities, with the use of bioactivity scores and p_value (using bioinformatics and metabolomics analysis).

Of these 19 promising predicted bioactive molecules, 4 bioactive molecules predicted to have significant bioactivity scores and low p_value have been identified by matching the bioactivity mapping data with the LC-MS and manual de-replication data through their $[M+H]^+$ masses. The two molecules named by their masses in the bioactivity mapping 601.3551 and 401.239, both have high bioactivity score of 87% and 63% with anti-cancer activities (0.870 and 0.631) were matched to the Desferrioxamine E and pro-thr-ala-ile metabolites. The molecule 443.2505, which had a high bioactivity score in both the anti-fungal (0.604) and anti-inflammatory (0.620) assays can be matched to the val-glu-pro-pro metabolite. Finally, the molecule 399.2223, which had a high bioactivity score of 0.655 with a high p_value of 0.011 in the anti-inflammatory activity assay can be matched to the val-ser-pro-pro metabolite. Those matched identifications confirm the bioactivity mapping predictions as some of these known

metabolites, such as Desferrioxamine E, has already been reported to possess similar types of bioactivity. Indeed, iron chelators such as Desferrioxamine (DFO) have been reported to inhibit the proliferation of tumor cells and induce cell apoptosis; thus having potential clinical importance in the treatment of leukemia. A large number of both in vitro and clinical studies have shown that DFO can inhibit the growth of tumor cells by chelating intracellular iron and DFO has also been shown to have a clear anti-tumor cell proliferation effect [Wang et al., 2017; Benadiba et al., 2017]. Indeed Wang and co-workers have shown that DFO had both growth-arresting and apoptosis-inducing effects on tumor-associated mesenchymal stromal cells (TAMSCs) and on bone marrow associated mesenchymal stromal cells (BMMSCs) from EG-7 tumors. Benadiba et al., 2017 have reported that transformed cells from a murine PTEN-deficient T-cell lymphoma model and from T-cell acute lymphoblastic leukemia/lymphoma (T-ALL/T-LL) cell lines were overexpressing CD71, leading to the development of an iron "addiction" whose chelation by deferoxamine (DFO) dramatically affected their survival to induce apoptosis. Another recent study has reported that DFO inhibits leukemia cell viability, induces cell apoptosis, and regulates expression of apoptosis-related genes (Yang et al., 2018). Yang and co-workers have also demonstrated that DFO treatment significantly decreased iron content in K562 cells, as well as, inhibiting the proliferation ability of K562 cells, both in a dose-dependent manner. They also demonstrated that iron chelation was in fact responsible for suppression of proliferation of the K562 cell lines by diminishing its antiproliferative effect due to the quenching of the iron chelator and that DFO can induce leukemia cell apoptosis and may involve a variety of apoptosis-related signal transduction pathways (Wang et al., 2018). Finally, Desferrioxamine E, also known as Nocardamin, has previously been reported to exhibit inhibitory effects on colony formation of T-47D, SK-Mel-5, SK-Mel-28 and PRMI-7951 tumor cell lines (Kalinovskaya et al., 2011).

Furthermore, 2 of the 19 promising molecules, which have high bioactivity scores and low p_value scores in the anti-cancer assay have been matched to potential new metabolites already detected in our LC-MS and manual de-replication dataset, making them even more promising molecules of interest. Indeed, the molecule 561.3602 which had a bioactivity score of 0.610 and can be matched to a potential new metabolite detected in the SM9-M19 extract, with the following molecular formula $C_{24}H_{52}N_2O_{12}$ (Figure 25b). The molecule 587.3389 which has a bioactivity score of 0.819, could also be matched to two possibilities of potential new metabolites detected in the SM9-SGG extracts. These two possibilities were as follows: the potential new metabolite with the $[M+H]^+$ mass of 587.3386 and the molecular formula $C_{22}H_{44}N_{12}O_{7;}$ and the potential new metabolite with the $[M+H]^+$ mass of 587.3387 and the molecular formula $C_{25}H_{50}N_2O_{13}$ (Figure 26b).

The remaining 13 of the 19 promising molecules predicted as having high bioactivity scores and low p_values could not be matched to our LC-MS and manual de-replication dataset, and thus, it is not possible to determine whether these molecule are known metabolites or potentially new metabolites. As previously mentioned, only NMR analysis will allow the definitive identification through full chemical characterisation of these metabolites. Among these are: 6 molecules namely 547.3445, 557.2875, 573.3235, 595.3053, 600.2565 and 623.3373 which have shown high anti-cancer bioactivity scores of 0.737, 0.689, 0.885, 0.868, 0.628 and 0.806 respectively (Figure 25) ; 5 molecules, namely 507.2427, 541.2639, 549.2811, 549.2813 and 603.3108, which have anti-inflammatory scores of 0.745, 0.735, 0.786, 0.732 and 0.750 respectively (Figure 26); 1 molecule, namely 496.1610, which has an anti-fungal score of 0.642 (Figure 27) and finally 1 molecule, namely 621.3217, which has both an anti-cancer score of 0.601 and an anti-inflammatory score of 0.881 (Figure 25 and 26).

In addition, some metabolites were matched to potential new metabolites from our LC-MS & manual de-replication dataset, but their low bioactivity score and/or high p_value make them less interesting to pursue for further analysis. For example, in the bioactivity mapping of the anti-cancer activity, the molecules 535.2651 and 1077.5485 with $[M+H]^+$ masses of 535.2651 and 1077.5485, low bioactivity scores of 0.377 and 0.316, and high p_values of 0.183 and 0.316 were both matched to potential new metabolite with the $[M+H]^+$ masses of 535.2652 and 1077.5479 with the molecular formula of $C_{21}H_{39}N_5O_{11}$ and $C_{49}H_{76}N_{10}O_{17}$ (Figure 25c). In the bioactivity mapping of the anti-inflammatory activity, in the second bioactive cluster, the same molecule 535.2651 with $[M+H]^+$ with a mass of 535.2651, which had a bioactivity score of 0.547 and a 0.043 p_value was matched to the same potential new metabolite with the $[M+H]^+$ masses of 535.2652 and the molecular formula $C_{21}H_{39}N_5O_{11}$ (Figure 26b). In the bioactivity mapping of the anti-fungal activity against *C. glabrata* CBS138, the molecule 585.3187 with $[M+H]^+$ mass of 585.3187, with the low bioactivity score of 0.312 and the high p_value of 0.277 was matched to the potential new metabolite with the $[M+H]^+$ masses of 585.3188 and the molecular formula $C_{20}H_{48}N_4O_{15}$ (Figure 27b). As previously mentioned only NMR analysis will ultimately allow the definitive identification through full chemical characterisation of this metabolite, therefore currently both possibilities exist.

4.2 Additional anti-oxidant and anti-microbial screening of SM9-M19 crude extract against human bacterial pathogens.

4.2.1 Antimicrobial screening

Following reculturing of SM9 in M19 media, the organic phase crude extract was subjected to an additional anti-microbial screen targeting the following human bacterial pathogens: *Staphylococcus aureus* 6538p, *Burkholderia metallica* LMG 24068, *Acinetobacter baumannii* AB13 and *Pseudomonas aeruginosa* PA01 (Figure 28). The Minimal Inhibitory Concentration (MIC) as well as the Minimal Bactericidal Concentration (MBLC) and the Minimal Bacteriostatic Concentration (MBSC) were determined using the extracts in a liquid inhibition assay.

mg/mL	B1	E4	DMSO	*Sa 6538P*	LB
1	0.547	-0.018	0.313	0.320	0.001
0.5	0.461	-0.024	0.273	0.306	0.004
0.25	0.394	-0.010	0.305	0.293	0.004
0.125	0.467	0.008	0.340	0.302	0.005
0.0625	0.463	-0.007	0.300	0.284	0.000
0.03125	0.425	-0.003	0.332	0.333	0.007
0.015625	0.482	0.327	0.318	0.304	0.007
0.0078125	0.483	0.451	0.338	0.302	0.003

(a)

mg/mL	B1	E4	DMSO	Bm	LB
1	0.905	0.463	0.804	0.882	0.002
0.5	0.871	0.528	0.838	1.032	0.001
0.25	0.932	0.537	1.006	0.983	0.001
0.125	0.967	0.655	1.083	1.026	0.002
0.0625	1.003	0.772	1.095	0.973	0.004
0.03125	1.022	0.729	1.019	1.005	0.005
0.015625	0.935	0.866	1.093	1.101	0.004
0.0078125	0.851	0.943	1.177	1.148	0.005

(b)

mg/mL	B1	E4	DMSO	Ab13	LB
1	0.913	0.587	0.920	0.986	0.002
0.5	0.926	0.666	0.929	0.997	0.003
0.25	0.954	0.781	0.973	0.993	0.005
0.125	0.957	0.846	0.962	0.992	0.005
0.0625	0.960	0.878	0.933	1.011	0.003
0.03125	0.949	0.914	0.921	0.975	0.004
0.015625	0.964	0.956	0.916	0.984	0.006
0.0078125	0.963	0.969	0.825	0.950	0.006

(c)

mg/mL	B1	E4	DMSO	Pa01	LB
1	0.782	0.566	0.784	0.788	0.002
0.5	0.844	0.627	0.784	0.841	0.001
0.25	0.866	0.671	0.839	0.855	0.002
0.125	0.880	0.753	0.888	0.833	0.002
0.0625	0.935	0.798	0.942	0.896	0.003
0.03125	0.882	0.792	0.933	0.885	0.002
0.015625	0.919	0.794	0.888	0.835	0.002
0.0078125	0.913	0.858	0.924	0.890	0.003

(d)

Legend

inhibition-bactericidal
inhibition-bacteriostatic
no to medium inhibition
no inhibition

Figure 28: **Anti-bacterial assays against: (a) *S.aureus* 6538p ; (b) *B.metallica* LMG 24068, (c) *A.baumannii AB13* and (d) *P.aeruginosa* PA01**.

E4= SM9-M19 organic phase crude extract (culturing media M19 + SM9, fermentation, then chemically extracted); B1= biological control (same culturing condition than SM9-M19 without SM9); control negative = LB; control positive =*Sa* 6538p, *B.m*, *A.b13*, *P.a* . Absorbance in each well was measured at 630nm, at T0 and then at T24h.

Absolute value of antibacterial activity= (T18 A630nm sample well – T0 A630 nm sample well).

The SM9-M19 organic phase crude extract was found to be active against *S.aureus* 6538p at a MIC and MBLC of 0.03125 mg/ml (Figure 28a). The SM9-M19 organic phase crude extract also displayed bacteriostatic activity against *B.metallica* LMG 24068, *A.baumannii* AB13 and *P.aeruginosa* PA01; at a MIC and Minimal MBSC of 0.125 mg/ml, 0.25 mg/ml and 0.25

mg/ml respectively (Figure 28b, 28c and 28d). The activity was found to be dose-dependent. The bacteriostatic activity of the extract is a good indicator of potential anti-biofilm activity.

4.2.2 Antioxidant activity

In parallel, the anti-oxidant activity of SM9-M19 crude extracts was assessed, using the DPPH antioxidant assay (2,2-diphenyl-1-picrylhydrazyl) with Vitamin C (Ascorbic acid) being used as a positive control. The SM9-M19 organic phase crude extract displayed a total antioxidant activity at 0.03125 mg/ml (concentration dependent visible in Figure 29) and displayed a radical scavenging activity of 49 % at a concentration of 1mg/ml (Figure 29). Given that the positive control Vitamin C, is known to have high antioxidant effect, and displayed 50 % activity in this assay, the 49 % activity observed with the crude extract (SM9-M19) is high. Our results are similar to those previously reported from other marine bacteria, with Saurav and co-workers who reported the identification of a 5-(2,4-dimethylbenzyl)pyrrolidin-2-one (DMBPO) compound extracted from a marine *Streptomyces* VITSVK5 spp, which exhibited significant DPPH radical scavenging activity and total antioxidant activity of 50.10 % at 5 µg/ml DMBPO (Saurav et al., 2011). Kumaqai et al. (1993), also reported a PC-766 B compound, which exhibited a dose dependent antioxidant activity, isolated from the marine actinomycete *Nocardia brasiliensis*. While Dholakiya and co-workers also reported antioxidant activity in extracts from a marine *Streptomyces variabilis* RD-5 strain, which showed 82.86 and 89 % of 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical scavenging activity and metal chelating activity, respectively, at 5.0 mg/mL (Dholakiya et al., 2017).

mg/mL	E4	B1	C+	C-
1	49%	-33%	50%	-33%
0.5	24%	-39%	48%	-40%
0.25	-1%	-43%	47%	-31%
0.125	-11%	-46%	47%	-45%
0.0625	-14%	-42%	46%	-41%
0.03125	-8%	-44%	48%	-45%

Radical scavenger activity
Anti-oxidant activity
No activity

***Figure 29*: Anti-oxidant activity of SM9-M19 crude extract, at T30min, 490nm, visible in percentage of activity *(%)*.**

Additional information: E4= SM9-M19 organic phase crude extract (culturing media M19 + SM9, fermentation, then chemically extracted); B1= biological control (same culturing condition than SM9-M19 without SM9); Positive Control = Vitamin C (Ascorbic acid); Negative Control = Methanol (MeOH). Absorbance value was measure at T0, then at T30 min at 490 nm.

% anti-oxidant activity = [(T30 min A490 nm control negative average value - T30 min A490 nm sample value) / T30 min A490 nm control negative average value] x100.

It is important to note that the control-M19 extract displayed no activity in any the anti-bacterial assays or in the anti-oxidant assay (Figure 28 and 29). This demonstrate that the activities of the SM9-M19 crude extract doesn't derive from the media itself but from the secondary metabolites that are produced following the fermentation of the SM9 isolate in the M19 media.

Thus, it is clear that the SM9-M19 organic phase crude extract contains secondary metabolites which displayed a radical scavenging activity and total anti-oxidant activity; which are

bactericidal against *S.aureus* 6538p; bacteriostatic against *B.metallica* LMG 24068, *A.baumannii* AB13 and *P.aeruginosa* PA01. This antioxidant property of the extract could be further studied by *in vivo* animal model studies. It would also be interesting to investigate on the anti-biofilm activity of these metabolites against *B.metallica* LMG 24068, *A.baumannii* AB13 and *P.aeruginosa* PA01 because of their bacteriostatic activity.

Conclusions

The SM9 isolate was selected from amongst the others strain isolates because of the high bioactivity potential that it exhibited. The OSMAC approach, revealed different bioactivity following cultivation in the various culture conditions, such as an anti-cancer activity when SM9 was fermented in SGG and SM media; a strong antifungal activity when SM9 was cultured in M19, M400, SGG, OM and SM media; and partial anti-inflammatory activity when SM9 was fermented in M19 and SM media. Indeed, the antimicrobial activity screening of the crude extracts tested was shown to possess secondary metabolites containing anti-fungal activity against *C. glabrata* CBS138; anti-inflammatory activity against the LPS induced TNF-α production in Thp-1 cells, with an inhibition of 44-45%; and an anti-cancer activity against the human melanoma cell line A2058 allowing a percentage of 10-41% cell survival. This demonstrates that SM9 crude extracts contain secondary metabolites that have an anti-fungal activity against *C. glabrata* CBS138; anti-inflammatory activity against LPS induced TNF-α production in Thp-1 cells and or an anti-cancer activity against the human melanoma cell line A2058, when SM9 cultured in specific growth media.

In addition, bioinformatics and metabolomics investigation of these separate bioactivities have allowed the mapping of each bioactivity within the chemical diversity of the SM9 isolate, though variation of cultivation parameters and variation of bioactivity, by the establishment of

a bioactive molecular network for the following activities namely, anti-fungal activity against *C. glabrata* CBS138; anti-inflammatory activity against the LPS induced TNF-α production in Thp-1 cells and or an anti-cancer activity against the human melanoma cell line A2058.

The bioactivity mapping of the anti-cancer activity against the human melanoma cell line A2058, has highlighted one bioactive cluster containing 11 promising molecules of interest and an additional 2 molecules of interest in the network. Their bioactivity scores ranged from 0.601 to 0.885, representing 60% to 89% of bioactivity and included 2 potential new metabolites. The bioactivity mapping of the anti-inflammatory activity against the LPS induced TNF-α production in Thp-1 cells, highlighted 14 promising molecules and 2 potential bioactive clusters. The promising molecules have shown bioactivity scores ranging from 0.620 to 0.881, representing 62% to 88% of bioactivity. The bioactivity mapping of the anti-fungal activity against *C. glabrata* CBS138, has highlighted 2 promising molecules and 1 potential bioactive cluster. These two promising molecules displayed bioactivity scores of 0.604 and 0.642, representing 60% and 64% of bioactivity.

Furthermore, the bioactivity mapping has highlighted a total of 19 promising bioactive molecules from a total of 177, with a high bioactive score and low p_value. Four of these 177 promising molecules were matched to our previous LC-MS dataset, through their $[M+H]^+$ mass, such as, Desferrioxamine E, pro-thr-ala-ile, val-glu-pro-pro, val-ser-pro-pro metabolites. Two others out of 177 promising molecules were matched to potential new metabolites from our previous LC-MS dataset, through their $[M+H]^+$ masses. While the remaining 13 promising molecules out of 177 could not be matched, they do remain promising as it is not clear as yet whether these molecule are known metabolites or potentially new metabolites. Only NMR analysis will allow the definitive identification through the full chemical characterisation of these metabolites. Among them, 6 molecules have shown high anti-cancer bioactivity scores ranging from 0.628 to 0.885, representing 63%-89% of bioactivity. 5 molecules have shown

anti-inflammatory scores ranging from 0.732 to 0.786, representing 73%-79% of bioactivity; 1 molecule has shown an anti-fungal score of 0.642 (=64%) and lastly, 1 molecule has shown both anti-cancer score anti-inflammatory activity with scores of 0.601 and 0.881, respectively; representing 60% of the bioactivity against the human melanoma cell line A2058cancer and 88% of the bioactivity against the LPS induced TNF-α production in Thp-1 cells.

From all the SM9 extracts tested, the SM9-M19 extract was specifically selected for additional antimicrobial and antioxidant screening analysis because of its strong antifungal activity and partial anti-inflammatory activity, as well as, its high potential for production of secondary metabolites including potential new metabolites. The antimicrobial activity screening of the extract was shown to possess both bactericidal and bacteriostatic antibacterial activity against *S.aureus* 6538p and *A.baumannii* AB13, *B.metallica* LMG 24068, *P.aeruginosa* PA01, strains, respectively, when cultured in the M19 media at a temperature of 28°C. In addition, the extract also showed a radical scavenging anti-oxidant activity at a concentration of 1 mg/ml which could be further investigated by *in vivo* animal model studies.

This demonstrated that the crude extract of interest (SM9-M19, also called E4), contain secondary metabolites which have an antioxidant effect (including radical scavenging activity) and which are active against *C. glabratra* CBS , *S.aureus* 6538p, *A.baumannii* AB13, *B.metallica* LMG 24068, *P.aeruginosa* PA01. Fractionation and further characterizations such as chemical characterisation NMR followed by toxicity testing on these metabolites would be required to test their potential efficacy again these pathogens.

Material and Methods

- **Origin of strains**

The bacterial strains SM9 was obtained from the marine sponge *Haliclona simulans* (class Demospongiae, order Haplosclerida, family Chalinidae), at a depth of 15 m in Galway bay (Ireland). (See chapter 2 Material and Methods, p188 for more detail).

- **Production of secondary metabolites by fermentation**

The production of secondary metabolites was originally performed following the OSMAC approach used in chapter 1 and 2 (see chapter 1 material and methods) with fermentation of the strains in different nutrient-sources media and different temperature conditions. In this study, we repeated the fermentation process on the *Streptomyces* SM9 strain in one media, namely M19, at a temperature of 28 °C. The composition of the fermentation media M19 is provided in Table 2 of the chapter 1.

Identical methods for production of secondary metabolites by SM9 in M19 media were employed (see chapter 2 material and methods for more details), with two exceptions: for the large scale production: 8L of fermentation media was produced instead of 50 mL as in chapter 1; and at the end of the experiment the pellet was discarded instead of being freeze dried. The cell free supernatant (8 L approximate) was stored at -80 °C and/or used immediately for extraction.

- **Liquid Liquid Extraction (LLE)**

Cell-free supernatants (8 L) were chemically extracted using Liquid Liquid Extraction (LLE) with ethyl-acetate as solvent. The solvent was added to the filtrate in the ratio of 1:1(v/v), shaken vigorously and left to stand for 5 min to obtain the organic and aqueous phases. The organic phase was separated from the aqueous phase using a separating funnel. The organic

phase extract were then concentrated by evaporating to dryness at 45 °C in a GeneVac to generate dry crude extracts. The aqueous phase extract was discarded.

Bioactivity bioassays- guided dereplication

- **Anti-fungal NCCLS assay**

All the crude extract samples were checked for their ability to inhibit the growth of *C.glabrata* CBS138, by performing an NCCLS assay (National Committee for Clinical Laboratory Standards) following the protocol of Liu *et al*,.2007. The NCCLS assay is a broth colorimetric broth microdilution method, which uses the colorimetric indicator Resazurin, which is blue in sterile condition and turns pink in presence of yeast, due to a pH change. For this experiment, we used a 96-well microtiter plate, the dried crude extracts were mixed with 10 µL of ethyl acetate for the organic phase extract samples, or 10 µL of sterile water for the aqueous phases extracts. Then the 10µL of the liquid extracts were placed into the well containing 290 of RPMI solution, and were serially diluted following a 3-fold dilution to observe the level of activity, with a final volume of 200 in each well. 100 µL of Rcs solution, containing both *C. glabrata* CBS138 inoculum and Resazurin, was added to the wells, except the sterility control wells. The sterility control wells were then supplemented of 100 µL of R solution containing distilled sterile water with Resazurin. The plate was then incubated at 28 °C and results were taken after 24 h and after 48 h following incubation. The antifungal activity of the extracts was then visually observable by the pink to blue colour change in the wells. For confirmation of the validity of the method and of the bioactivity of the extracts, the following controls were performed: media control, ethyl acetate control, growth control and sterility control. The RPMI solution and the R solution was prepared following the protocol of Liu et al,.2007. The Rcs solution was prepared as follow: *C.glabrata* CBS138 was plated on YPD agar and grown

overnight at 28 °C. A single colony of the yeast strain was used to inoculate 5mL of sterile distilled water in sterile bacteriological tubes, then suspensions were mixed for 15 s to ensure homogeneity and subsequently diluted to a final OD = 0.12–0.15 . Then the working suspension of 20mL was supplemented with 0.1ml sterilized solution of Resazurin (20 mg/ ml in water).

- **Anti-cancer assay: CellTiter96 Aqueous One Solution Cell Proliferation assay**

The anti-cancer assays were performed as part of a collaboration with the University of Tromso, in the Marbio facilities and by the group of Prof. Jeanette Hammer Anderson, namely Kirsti Helland and Marte Albrigtsen .

The CellTiter 96 ® AQueous One Solution Cell Proliferation assay is a colorimetric method for determining the number of viable cells in proliferation or cytotoxicity assays. Metabolically active cells will reduce the tetrazolium compound MTS (yellow) into a colored formazan product (blue) that is soluble in tissue culture medium. The assays are performed by adding AQueous One Solution to cells in a tissue culture plate and incubating the cells for 1-4 hours and reading Absorbance at 490 nm (450-540 nm). The quantity of formazan product as measured by the absorbance at 490 nm is directly proportional to the number of living cells in culture.

The cell line A2058 were seeded in 96 -wellplates at a density of 2000 cells/well and then incubated overnight at 37°C and 5 % CO_2. The cells were then observed under the microscope to check that they were healthy and evenly distributed in the plate. The cell media were removed then, RPMI w/10 % FBS were added. Then samples were added (in triplicates) to a total volume of 100 µL/well. Samples were placed in a 96-well plate and incubated for 72 h at 37 °C and 5 % CO_2. After 72 h, 10 µl AQOS were added to each well and the plates were then incubated for 1 h at 37°C and 5 % CO_2. Finally the absorbance was read at ~490 nm in a 96

well plate reader. Results were represented as follow: **A= active 50% > survival; Q= questionable 50 - 60 % survival; I= Inactive > 60 % survival.**

- **Anti-inflammatory assay- ELISA assay: TNF_THPAIF assay**

The anti-inflammatory assays were also performed as part of a collaboration with the University of Tromso, in the Marbio facilities with the group of Prof. Jeanette Hammer Anderson, namely Kirsti Helland and Marte Albrigtsen.

The anti-inflammatory effect of the extracts was measured in Thp-1 cells, a monocytic cell line. Thp-1 cells will secrete the cytokine TNF-α when treated with lipopolysaccharides (LPS). LPS are components of the Gram-negative bacterial cell wall and will induce an immune response in the Thp-1 cells. Extracts with anti-inflammatory effect will inhibit the LPS induced TNF-α production in the cells. Thp-1 cells differentiated from monocytes to macrophages are added to the crude extract samples (Concentration 100 µg /ml of crude extracts). After one hour of incubation LPS is added and the cells are incubated for 6 hours. TNF-α secreted into the cell culture media is measured by a TNF-α ELISA assay. Duplicate of each sample were generated.

Cells were seeded in microtiter plates at a density of 1 x 10^6 cells/mL. 50 ng/mL PMA was added to the cell suspension (0,5 µl PMA stock solution pr 10 ml) to differentiate cells from monocytes to macrophages. A 10 mL cell suspension was generated for each microtiter plate and 5 mL extra were added. 100 µl of cell suspension was added to each well in a microtiter plate, then incubated at 37 °C with 5 % CO2, for 48 hours. The cells were monitored after 24h by microscopy to ensure that they had started to differentiate. After 48h, the cells were washed with endotoxin tested PBS and new growth media was added before incubation at 37 °C with 5 % CO_2 for a further 24 hours. The cell media was removed and 80 µl RPMI growth media was added to each well. 90 µl RPMI was added to LPS controls and 100 µl RPMI was also

added to cell controls.10 µl of samples were added to each respective well (and 10 µl of controls solution was added to the control wells). Cells were then incubated at 37 °C with 5 % CO_2 for 1 hour. Then 10 µl of 10 ng/mL LPS solution was added to each well except the cell control. Cells were incubated for 6 h at 37°C with 5 % CO_2. The reactions were stopped by immediately freezing the plates at -80°C.

One day prior to the ELISA testing, Affinity purified anti-human TNF-α to 2 µg/mL was diluted into in 10 mM TBS then 100 µL of the mixture was added to each well in a 96-well Maxisorp plate before being incubated overnight at 4°C. Plates were washed with the Aquamax plate washer, then 200 µl 10 mM TBS with 2 % BSA were added to each well before incubation for 1hour with on shaker. The washing step was repeated 3 times, then samples and standards were diluted in the assay diluent, directly in ELISA plate for a total volume of 100 µL in each plate before incubation for 2 hours on shaker. The washing step was repeated three times and then the Biotin anti-human TNF-α was diluted to 3 µg/mL in 10 mM TBS. 100µl of the mixture was then added to each well and incubated for 1 hour on shaker. The plates were then washed again. Dilution of 1:20000 of ExtrAvidin-Alkaline phosphatase in 10 mM TBS was made, then 100µl of the mixture was then added to each well and incubated for 30 min on shaker. Plates were washed again. 100 µL of 1 mg/mL Solved pNPP in 1 M diethanolamin buffer were added to each well for 30-45 minutes then the absorbance was readed at 405 nm.

Extracts which display an anti-inflammatory effect will inhibit the LPS induced TNF-α by at least 50 %. A= active; U = unclear but can be considered partly active; I= inactive. U=unclear but can be considered partly active

- **Liquid inhibition assays: Anti-microbial screening**

The liquid inhibition assays were performed in the Consiligo Nazionale Delle Ricerche, Institute of Protein Biochemistry (IBP-CNR), at Napoli (Italia) during my secondment.

The crude extracts (sample and biological controls) were checked for their ability to inhibit the growth of a selected panel of human pathogenic bacteria using a 96-well microtiter plate and in vitro Minimal Inhibitory Concentration (MIC) assay. The selected panel of human bacterial pathogens used for the antimicrobial screening were as follows: *Acinetobacter baumannii* Ab13*; Burkholderia metallica* LMG 24068; *Pseudomonas aeruginosa* PA01 and *Staphylococcus aureus* 6538P. These bacteria were plated on LB agar and grown overnight at 37 °C. The extracts were placed into the well of the 96-well microtiter plate at an initial concentration of 1 mg/mL and were serially diluted 2-fold in LB. A positive control column did not contain any extract but contain the human bacteria pathogens used as a growth control. Two percent (v/v) DMSO was also used as a control, to assess the toxicity of the solvent on the bacterial growth. B1 extract was used as a biological control and LB was used as a sterility control. A single colony of each pathogen strain was used to inoculate 3 mL of liquid medium in sterile bacteriological tubes. After 5-8 hr of incubation, growth was measured by monitoring the Abs600 nm and approximately $4x10^3$ CFU/mL was dispensed into each well (with the exception of the sterility control). The plates were then incubated at 37 °C for 18 hours. Growth was measured at T0 and T18, by using a 96- well plate reader (VICTOR X Multilabel Plate Reader, PerkinElmer, USA) by monitoring the Abs630 nm. The activity level was determined by calculation of the antibacterial activity:

Absolute value of antibacterial activity= (T18 A630nm sample well – T0 A630 nm sample well).

- **2,2-diphenyl-1-picrylhydrazyl (DPPH) assay : Antioxidant activity**

The liquid inhibition assays were also performed in the Consiligo Nazionale Delle Ricerche, Institute of Protein Biochemistry (IBP-CNR), at Napoli (Italia) during my secondment.

The 2,2-diphenyl-1-picrylhydrazyl (DPPH) assay was conducted as previously described (Raes et al., 2004). 100 µL of Methanol (MeOH) was added to each well, followed by an additional 96 µL MeOH to the first row + 8 µL of extract or control was added to the same row, and a 2x-fold serial dilution was then performed to dilute the extracts in the plate. Finally, 100 µL of 0.3 mM DDPH (prepared in MeOH) was then added to all the wells, including the controls. Ascorbic acid (Vitamin C) was prepared at 50 mg/mL in Milli-Q water (MQ H_2O) and used as a positive control and methanol was used as a negative control. The plates were measured at an Abs 490 nm and incubated at 15 °C in the dark for thirty minutes. Following this he plates were again measured at an Abs 490 nm, and the % antioxidant activity was calculated as follows:

% anti-oxidant activity = [(T30 min A490 nm control negative average value - T30 min A490 nm sample value) / T30 min A490 nm control negative average value] x100.

Metabolomics analysis

- **Spectral Feature Detection.**

This step was performed following as previously described by the Dorrestein laboratory, with the corresponding online documentation on GitHub : https://github.com/DorresteinLaboratory/Bioactive_Molecular_Networks . The popular MZmine2 toolbox, version MZmine2.51, was used for LC-MS feature detection, integration, filtering, and visualization. Raw data from the previous LC-MS/MS experiment (Chapter 2), were re-invested and imported into MZmine2, in mzXLM format. MZmine 2 data processing was analysed following the manual.pdf file present inside the download MZmine2.51-windows

zip file and the README file from the Dorrestein laboratory. In the Peak list row filter module, two options were selected: "Keep only features with MS/MS scan (GNPS)", 'Reset the peak number ID". The MGF file and the CSV table were both exported using the module “Export for GNPS” which includes 'Export row ID', 'Export row m/z', 'Export row retention time', and 'Peak area'. The MGF file generated, containing MS/MS spectral data for each feature was uploaded on GNPS, and a Data Analysis was performed as described below. The CSV table, named (feature_quantification_matrix.csv), containing relative quantitative information such as the aligned list of features and their intensity across the crude extracts analyzed by LC-MS/MS, was used for both the calculation of the bioactivity score and the visualization of the relative abundance in the molecular networks.

The CSV table file was used to predict the: anti-cancer, anti-inflammatory and anti-fungal activity scores. Then the value of the bioassay experiments for each sample analyzed by LC-MS/MS were manually added into each relevant tables (features_quantification_matrix.csv to feature_quantification_matrix_edited_anticanceractivity.csv;
feature_quantification_matrix_edited_antifungalactivity.csv;
feature_quantification_matrix_edited_antiinflammatoryactivity.csv).

- **Prediction of Bioactivity Score Significance : Jupyter notebook**

An R-based Jupyter notebook, called Bioactive Molecular Networks, was used from GitHub, https://github.com/DorresteinLaboratory/Bioactive_ Molecular_Networks along with a step-by-step tutorial and the representative input/output files needed (Nothias et al, 2018). We obtained new CSV format tables (.csv) of spectral features’ intensity edited with bioactivity data. In our study, three bioactivity were analysed via Jupyter notebook, therefore we obtained 3 csv tables.

This step involved three steps: scaling samples by normalizing the intensity to the total ion current TIC (van den Berg et al, 2006), calculating the Pearson correlation score (r) and its significance (r_pval) between each feature and the bioactivity value, as well as the multiple hypothesis testing correction (the Bonferroni correction), and output the node attribute table (Nothias et al, 2018). The three tables were outputted, namely features_quantification_matrix_anticanceractivity_transposed_with_significant_correlation_pvalue_corrected.csv ; features_quantification_matrix_antifungalactivity_transposed_with_significant_correlation_pvalue_corrected.csv and features_quantification_matrix_antiinflammatoryactivity_transposed_with_significant_correlation_pvalue_corrected.csv) can all be imported (separately) into the Cytoscape software to map back the bioactivity score onto the corresponding MS/MS molecular network.

- **Bioactivity mapping into the Molecular Network.**

The molecular network was created using the GNPS platform (http://gnps.ucsd. edu). The spectral information (MGF file) previously generated by MZmine2 was uploaded on GNPS. The precursor ion mass tolerance was set to 0.01 Da and to a product ion tolerance of 0.0075 Da (allowing a maximum error of 25 ppm at m/z 300 and 12 ppm at m/z 600) as previously described (Nothias et al., 2018). The fragment ions below 10 counts were removed from the MS/ MS spectra. MNs networks were generated using 12 minimum matched peaks and a cosine score of 0.7. The molecular networking data was imported using the Cytoscape 3.4.0 software for visualization of the molecular network. Then, one of our previous output tables of the Jupyter notebook was imported into Cytoscape into the molecular network (features_quantification_matrix_anticanceractivity_transposed_with_significant_correlation_pvalue_corrected.csv) creating the bioactivity mapping of the anti-cancer activity. This

workflow was subsequently repeated to create molecular networks for the anti-fungal and the anti-inflammatory activity mapping.

Finally, metabolome "mining" was used on each bioactive molecular networks. A cosine score factor was used and nodes were filtered for their significant bioactivity based on their r score and their false discovery rate (FDR) corrected p value, in the Seclect function of Cytoscape. The style of these nodes was then bypassed with a larger node size to allow the direct visualization of the bioactive molecules and a piechart diagrams was generated to display the distribution of the ions across the different samples. Only molecules having a bioactivity score above $r > 0.6$ and p value <0.03 (for anti-fungal and anti-cancer activity mapping) or p value <0.003 (for anti-inflammatory activity mapping) were represented as having a larger node on the bioactive molecular networks.

- **Potential new metabolite into the molecular network- bioactivity network updated**

For this part, we re- analysed our previous data from LC-MS/MS experiment from chapter2 (prior MZmine2 data processing). This data included the identification of potential new metabolites which were not in the ChemSpider, Antibase, PubChem, Reaxys and the Dictionary of Natural Products libraries. The $[M+H]^+$ masses of these potential new metabolites were compared and matched to the $[M+H]^+$ masses of the Cytoscape molecules. The matched masses were then highlighted using metabolome mining by integrating the value "no hit" to Cytoscape then bypassing the style of the nodes on all "no hit" nodes with a red colour square to allow the direct visualization of the potential new metabolites already analyzed by LC-MS/MS.

References

Benadiba, J., Rosilio, C., Nebout, M., Heimeroth, V., Neffati, Z., Popa, A., Mary, D., Griessinger, E., Imbert, V., Sirvent, N., Peyron, J. F. (2017). Iron chelation: an adjuvant therapy to target metabolism, growth and survival of murine PTEN-deficient T lymphoma and human T lymphoblastic leukemia/lymphoma. *Leuk Lymphoma, 58*(6), 1433-1445. doi:10.1080/10428194.2016.1239257

Buedenbender, L., Astone, F. A., & Tasdemir, D. (2020). Bioactive Molecular Networking for Mapping the Antimicrobial Constituents of the Baltic Brown Alga Fucus vesiculosus. *Marine Drugs, 18*(6). doi:10.3390/md18060311

Dholakiya, R. N., Kumar, R., Mishra, A., Mody, K. H., & Jha, B. (2017). Antibacterial and Antioxidant Activities of Novel Actinobacteria Strain Isolated from Gulf of Khambhat, Gujarat. *Front Microbiol, 8*, 2420. doi:10.3389/fmicb.2017.02420

Fang, Q., Maglangit, F., Wu, L., Ebel, R., Kyeremeh, K., Andersen, J. H., Annang, F., Perez-Moreno, G., Reyes, F., Deng, H. (2020). Signalling and Bioactive Metabolites from *Streptomyces* sp. RK44. *Molecules, 25*(3). doi:10.3390/molecules25030460

Kalinovskaya, N. I., Romanenko, L. A., Irisawa, T., Ermakova, S. P., & Kalinovsky, A. I. (2011). Marine isolate *Citricoccus* sp. KMM 3890 as a source of a cyclic siderophore nocardamine with antitumor activity. *Microbiol Res, 166*(8), 654-661. doi:10.1016/j.micres.2011.01.004

Kumagai, K., Fukui, A., Tanaka, S., Ikemoto, M., Moriguchi, K., & Nabeshima, S. (1993).

PC-766B, a new macrolide antibiotic produced by *Nocardia brasiliensis*. II. Isolation, physico-chemical properties and structure elucidation. *J Antibiot (Tokyo), 46*(7), 1139-1144. doi:10.7164/antibiotics.46.1139

Kumagai, K., Taya, K., Fukui, A., Fukasawa, M., Fukui, M., & Nabeshima, S. (1993). PC-766B, a new macrolide antibiotic produced by *Nocardia brasiliensis*. I. Taxonomy, fermentation and biological activity. *J Antibiot (Tokyo), 46*(6), 972-978. doi:10.7164/antibiotics.46.972

Lind, K. F., Hansen, E., Osterud, B., Eilertsen, K. E., Bayer, A., Engqvist, M., Leszczak, K., Jorgensen, T. O., Andersen, J. H. (2013). Antioxidant and anti-inflammatory activities of barettin. *Marine Drugs, 11*(7), 2655-2666. doi:10.3390/md11072655

Liu, M., Seidel, V., Katerere, D. R., & Gray, A. I. (2007). Colorimetric broth microdilution method for the antifungal screening of plant extracts against yeasts. *Methods, 42*(4), 325-329. doi:10.1016/j.ymeth.2007.02.013

Nothias, L. F., Nothias-Esposito, M., da Silva, R., Wang, M., Protsyuk, I., Zhang, Z., Sarvepalli, A., Leyssen, P., Touboul, D., Costa, J., Paolini, J., Alexandrov, T., Litaudon, M., Dorrestein, P. C. (2018). Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *J Nat Prod, 81*(4), 758-767. doi:10.1021/acs.jnatprod.7b00737

Pan, R., Bai, X., Chen, J., Zhang, H., & Wang, H. (2019). Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review. *Front*

Microbiol, 10, 294. doi:10.3389/fmicb.2019.00294

Raes, K., De Smet, S. and Demeyer, D. (2004). Effect of dietary fatty acids on incorporation of long chain polyunsaturated fatty acids and conjugated linoleic acid in lamb, beef and pork meat: a review. *Animal Feed Science and Technology*, 113(1-4), pp.199-221. doi:10.1016/j.anifeedsci.2003.09.001

Reen, F. J., Gutierrez-Barranquero, J. A., Dobson, A. D., Adams, C., & O'Gara, F. (2015). Emerging concepts promising new horizons for marine biodiscovery and synthetic biology. *Marine Drugs, 13*(5), 2924-2954. doi:10.3390/md13052924

Roetzer, A., Gabaldon, T., & Schuller, C. (2011). From *Saccharomyces cerevisiae* to Candida glabrata in a few easy steps: important adaptations for an opportunistic pathogen. *FEMS Microbiol Lett, 314*(1), 1-9. doi:10.1111/j.1574-6968.2010.02102.x

Saurav, K., & Kannabiran, K. (2012). Cytotoxicity and antioxidant activity of 5-(2,4-dimethylbenzyl)pyrrolidin-2-one extracted from marine *Streptomyces* VITSVK5 spp. *Saudi J Biol Sci, 19*(1), 81-86. doi:10.1016/j.sjbs.2011.07.003

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142. doi:10.1186/1471-2164-7-142

Wang, G., Shen, G., & Yin, T. (2017). In vitro assessment of deferoxamine on mesenchymal

stromal cells from tumor and bone marrow. *Environ Toxicol Pharmacol, 49*, 58-64. doi:10.1016/j.etap.2016.11.014

Weller, M. G. (2012). A unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques. *Sensors (Basel), 12*(7), 9181-9209. doi:10.3390/s120709181

Xu, N., Ye, C., Chen, X., Liu, J., Liu, L., & Chen, J. (2016). Genome Sequencing of the Pyruvate-producing Strain *Candida glabrata* CCTCC M202019 and Genomic Comparison with Strain CBS138. *Sci Rep, 6*, 34893. doi:10.1038/srep34893

Yang, Y., Xu, Y., Su, A., Yang, D., & Zhang, X. (2018). Effects of Deferoxamine on Leukemia In Vitro and Its Related Mechanism. *Med Sci Monit, 24*, 6735-6741. doi:10.12659/MSM.910325

5 General Discussion

Discussion

Contributing to the fight against Global AMR threat

AMR has been recognized as a serious global threat to public health worldwide; which has resulted in a large number of high-level policy initiatives, being led by institutions such as the World Health Organisation (WHO), the United Nations (UN), and many others. These initiatives have been undertaken due to the potential negative effects of AMR which are likely to adverse morbidity, mortality and financial consequences such as forcing up to 24 million people into extreme poverty by 2030; the prediction of 10 Million deaths globally due to AMR by 2050; and the loss of a total Gross Domestic Product (GDP) of $100.2 trillion by 2050 if nothing is done to prevent AMR (Dadgostar.2019; O'Neill.2014).

In an effort to address this growing and challenging threat, this study aimed to contribute to the global fight against AMR by employing an OSMAC based approach, together with new marine "omics" based methods such as metabolomics and genomics, to study marine sponge derived *Streptomyces* isolates. By characterizing these new marine bacterial strains and targeting specific bioactivity towards clinical pathogens that have been prioritised by the WHO; we hoped to identify secondary metabolites produced by these isolates that may have potentially new bioactivities. The other aim was to investigate and describe potential novel biosynthetic pathways that may encode for any of the new bioactivities that were observed in these strains.

Following the screening regime isolates were identified that produced metabolites which were active against categorized "Urgent" and "Serious threat" microorganisms from the WHO/CDC list of priorities (Table 1). In this respect, we found biological activity against *Acinetobacter*

baumannii and *Pseudomonas aeruginosa,* which are registered as being an "urgent threat" and classified as "critical pathogens", together with activity against *Candida spp* (*Candida glabrata* and *Candida albicans* in our study) and *Staphylococcus aureus* which is registered as being a "serious threat". In addition, activities were identified which are not in the list of priorities but which are also important and of biomedical interest, namely anti-cancer, anti-inflammatory and antioxidant activities were also observed which are also of biomedical interest.

Specifically the deep sea isolate *B226SN104* displayed bioactivity against *C. glabrata* (CBS138), the shallow water isolate SM3 displayed bioactivity against *P. aeruginosa (PAO1) and B. subtilis (IE32)* while the shallow water isolate SM9 displayed bioactivity against *C. glabrata* (CBS138) and *Candida albicans* (Table 3 from chapter 1 and Table 6 from chapter 2). Furthermore, crude extracts from SM9 displayed bactericidal activities against *Candida glabrata, Candida albicans* and *Staphylococcus aureus (*6538p)*;* bacteriostatic activities against *Acinetobacter baumannii* (AB13) and *P. aeruginosa (PAO1)* ; as well as biostatic activity against *B. metallica (*LMG 24068). In addition, SM9 crude extract were also active against a human melanoma cell line A2058 (indicating potential anti-cancer activity) and also displayed activity against LPS induced TNF-α production in Thp-1 cells (indicating potential anti-inflammatory activity). Therefore, in total, we found extracts containing metabolites which displayed activity against five of the bacteria on the "emergent microbial threat" pathogen list of the CDC and WHO organisations (Table 1), and considered as serious and urgent threats.

One of the First Large-Scale marine OSMAC – multi-fields perspectives study

As previously mentioned in the introduction, marine ecosystems offer a large variety of different environmental conditions, such as variations in temperature, oxygen levels, nutrient

availability, salinity and pressure, which taken together has resulted in the evolution of marine microorganisms which has help them to adapt and survive in these extreme environments (Romano et al., 2018; Gribble et al., 2015; Antoraz S et al., 2015). The repertoire of NP which are isolated from marine ecosystems keeps expanding day by day as researchers around the globe/world start to used MGR more and more; together with the increasing visibility and benefits of using culture based approaches such as the OSMAC approach on marine strains, to facilitate the discovery of new natural product (NPs) (Romano et al., 2018; Bentley et al., 2002; Williamson et al., 2006). In this context the work presented here represents to our knowledge, the first studies to investigate a large scale based OSMAC approach involving 3 different bioactive marine bacterial strains, cultured in 7 different growth media, at 2 different temperatures; from which both organic and aqueous phases extracts were analysed from a genomics, metabolomics and chemical perspective.

Having employed different OSMAC conditions including different temperatures and different nutrient sources, when culturing the bioactive B226SN140 and SM3 isolates, we could clearly see that each strain had its own particular metabolome, cluster, and network. It was evident that subtle changes in nutrient source or in temperature, even if only involving a difference of 2 degrees (28-30°C) resulted in quite marked changes to their metabolome. Thus it is clear that production of these metabolites is likely to result from the differential regulation of transcription of different BGCs encoding these metabolites in both B226SN140 and SM3 under different physiological conditions. This is also evident in changes in their molecular networks.

In addition to revealing the likely complexity of the regulation of these metabolites in both strains, it is also indicates that OSMAC based approaches may prove to be useful in re-culturing already cultured strains from existing culture collections, to induce production of different sets of metabolites and potentially new ones (Romano et al., 2018; Schwarz et al., 2021). It also highlights the fact that such an approach can help elucidate the full metabolic

potential of strains that have already been isolated in many different laboratories, but that are likely to possess the potential to produce more and different types of metabolites, by subjecting these strains to growth under different sets of physiological parameters. Such an approach will of course be both strain dependant, and parameter dependant, as certain change could result in either a higher or reduced number of metabolites being produced. However not withstanding this, these approaches involving existing strain collections could reduce the need to re-collect new samples, thereby saving time and expenses such as expedition costs, and also reduce barriers that may exist in the future in sample collection in certain oceanic regions. However, it should also be considered that finding new physiological parameters that induce the production of metabolites in specific strains, may also be time consuming.

Streptomyces are good producer of secondary metabolites and bioactive metabolites

From all the nutrient sources and temperatures modifications tested, we detected and identified a total of 472 metabolites produced by SM3 and B226SN104, including 192 potential new metabolites; together with 140 metabolites produced by SM9 including 60 potential new metabolites. These high number of metabolites confirms that these three marine *Streptomyces* strains like other marine *Streptomyces* strains (Yang et al., 2020) are potent producers of secondary metabolites and that they are likely to be very promising strains for production of potential new metabolites. Further investigations on those potential new metabolites would be interesting to pursue, to confirm the potentiality novelty of the metabolites themselves, by performing chemical characterization of the metabolites by NMR but based on our evidence to date, they do appear to be very promising from a novelty perspective.

Another interesting finding from the study were the results of using M19 media and 28°C as fermentation parameters, for both SM3 and B226SN104. For both marine bacterial strains, the

combination of mannitol and peptone present in the culture media when fermented at 28°C resulted in an increase in the production of secondary metabolites and of potential new metabolites when compared to the other conditions tested, indicating that growth on mannitol and peptone at 28°C appear to be the optimal parameters for the production of secondary metabolites from the various parameters tested. In contrast for the SM9 isolate, the optimum culture conditions to produce the highest number of secondary metabolite including potential new metabolites was when the strain was fermented in SYP media at 28°C, while in contrast extracts from the strain grown on SYP did not result in any interesting activities (Table 4, Chapter 3). The extract which showed the best activity results was the organic phase from the SM –grown cultures, however this extract did not appear to contain any potential new metabolites. Therefore, for strains such as SM9, the optimum media which promotes the production of potentially novel metabolites may be not necessarily be the same media that results in promising biological activity.

Regarding biological activity , this study also demonstrated that the SM9 strain is a good producer of bioactive secondary metabolites with both bactericidal and bacteriostatic antibacterial activity against *S.aureus* 6538p and *A.baumannii* AB13, *B.metallica* LMG 24068, *P.aeruginosa* PA01, an anti-fungal activity against both *C. glabrata* CBS138 and *C. albicans* Sc5314. In addition SM9 produced metabolites with radical scavenging anti-oxidant activity; anti-inflammatory activity against the LPS induced TNF-α production in Thp-1 cells and anti-cancer activity against the human melanoma cell line A2058, when cultured in specific fermentation parameters, as highlighted in chapter 3.

Overall this study also illustrates that data from different analytical approaches can be combined together with other types of data such as the origin and morphology of the strain,

biological activity, extract phase, chemical weigh, chemical identification, chemical prediction; together with genomic characteristics such as the identification of BGCs. This data coupled with metabolomics characteristics such as metabolites identification of components of the metabolome of the strain can then be used to allow us to visualize all the relationships between the data combined together through a molecular network and in this way help to facilitate our understanding of the metabolomic and genomic potential of the strain. This allows one to 'zone in' on a specific research topic (e.g. analysis of specific metabolite and or specific metabolite clusters) while at the same time broadening the scope of the research to allow exploration of unexpected avenues of interest.

Opening potential for metabolomics based prediction of "silent" gene clusters

The aforementioned multi-field perspective, allows us to link microbiological culture regime conditions to metabolite production, and to couple this to the genomics and chemical diversity of the strain; thereby allowing us to link metabolites and chemical diversity involving potential BGCs under different culture regimes. Indeed, the annotation of metabolites into the molecular network that mapped exclusively to a certain medium, is likely to be indicative of the activation of silent BGCs due to growth of the strain in the varying culture parameters. For example, a specific metabolite represented in a single colored node in the metabolomic network represent a metabolite only produced following growth under one specific set of conditions, such as Cyclo (L-leucyl-L-propyl) metabolite, which was only detected in M19 extract from the SM3 isolate (Figure 4a and 4c, Chapter 1); while metabolite represented in a multi-colored node into the network represent a metabolite that can be produced following several

fermentation process, such as Sarmentoside B which was detected in 5 different media extracts from SM3 (Figure 4a and 4b, Chapter1).

Therefore, the more "shared metabolites" there are among the different media extracts, the easier it is to produce these metabolites. Easy in this context refers to the fact that these metabolites can be produced in different culture regimes or following not specific cultures parameters and are thus likely to be commonly produced metabolites and therefore potentially less interesting. In contrast, the more metabolites that are "not shared" or which are specifically produced under only one cultural parameter regime, the more unique they are potentially likely to be. Thus metabolomic based analysis helps to provide a visual hint as to which metabolites are likely to be synthetized from a cryptic or "silent" BGCs, allowing us to focus and subsequently analyze the hypothetical production of this metabolite; using additional relevant genomics analysis as a confirmatory step. It is important to note that metabolomics predictions only provide visual hints that need to be confirmed by supplementary analysis.

In addition, if the metabolite which is present in a metabolite sub network cluster with a high cosine score (observable with the thickness of the edge), it also suggests that the metabolites in this cluster (including the one you are interested in) may be synthetized by the same biosynthetic route (NRPS, PKS, or others). This is the case in our chapter 1 with the SM3 metabolome investigation revealing the presence of a specific metabolite cluster of 7 unique metabolites with a high cosine score, all of which were only present when the strain is cultured in one specific set of growth condition, namely in M19 media.

Reducing time of "time-consuming" analysis

Both chemical analysis (LC-MS, manual de-replication) and genomic analysis (genome mining, AntiSMASH,) can be very time-consuming, since chemical analysis requires the analysis of a mixture of thousands of metabolites or more, while genomics based analysis requires the investigation of thousands of genes from several genomic regions. Both fields can therefore benefit from an overall time perspective from the use of metabolomics based analysis.

Indeed, employing metabolomics based approaches greatly increases the efficiency of the discovery process by allowing the detection of the vast majority of the metabolites present in a mixture prior to the manual dereplication and helps in the identification of some of these metabolites by comparing their chemical profile simultaneously using several libraries. Instead of conducting experiments on thousands of metabolites by LC-MS, to attempt to identify all the detected metabolites one by one using manual dereplication and then identify the interesting ones; metabolomics analysis allows us to predict metabolites as interesting compounds, in a much shorter time frame; because they can be identified prior to manual dereplication as a result of the libraries comparisons. Manual dereplication can then be conducted by focusing primarily on the interesting metabolites, as opposed to having to consider all the metabolites in the sample.

Manual dereplication is by far the most time consuming component of the chemical analysis of extracts from bacterial cultures, because it is not yet clear what is precisely being looked for in the sample, which means that everything requires analysis. However by employing metabolomic analysis, it becomes be clearer what specific metabolite was being targeted or conversely what not to target if the metabolite is already a known one; and then redirect the manual dereplication in the direction of interest.

Of course metabolomics based analysis as with bioinformatic analysis is also time consuming, but comparing the time of metabolomics analysis with the manual dereplication targeting specific metabolites of interest, to the time involved in the manual dereplication time of potentially thousands of metabolites present in a which has not been subjected to metabolomic analysis, is vastly different.

From a genomics perspective, metabolomics analysis allows for a more focused in-depth analysis and also increases the efficiency in metabolite identification, as it is allows more time to be spent in connecting metabolites to the genomic regions clusters from which they are likely to be produced. This was the case in Chapter 2, with Antimycin A_1 being linked to a NRPS region; Desferrioxamine B and Desferrioxamine E being linked to a siderophore region and Surugamide A linked to a NRPS region, in the genome of SM9.

Employing both genomics and metabolomics together should be standard practice in drug discovery

While data from both metabolomics and genomics are only "predictions", the use of both sets of predictions together can offer a more accurate overall assessment of the likelihood of identifying novel metabolites, thereby reducing the instances of identification of previously identified and well characterised natural products, resulting in a reduction in resources and manpower as previously eluded to.

The data from this study clearly shows the benefits of using both genomics and metabolomics together in the discovery of natural products; indicating that this should perhaps become standard practice. Many laboratories favour using either genomics or metabolomics based approaches for natural product discovery, but the time involved could be drastically reduced if both were employed simultaneously.

The advantages of bioactivity mapping

As previously mentioned, the optimal conditions for bacteria to produce potential new metabolites may not be the same as the optimal conditions required for them to produce metabolites which are bioactive. In general, in drug discovery research, there are usually two approaches that researchers employ : one is to focus on biological activities to ensure that metabolites that are discovered exhibit the desired bioactivity, but in many cases, this results in uncovering a new activity from an already known compound (Firsova et al., 2017; Jakubiec-Krzesniak et al., 2018; Romano et al., 2018). The second and opposite approach, is to focus on the discovery of novel metabolites, which unfortunately in many instances results in the identification of new metabolites but which do not display the desired bioactivity.

However, in this study, we were able to overcome these limitations and generated a far greater list of potential new bioactive metabolites than is typically reported (Romano et al., 2018;

Pullen et al., 2002; Nothias et al, 2018) by elucidating both biological activity and novelty in parallel. This allowed us to not only, determine which conditions promoted the discovery of potential new metabolites and separately, which conditions promoted the production of biological activity in our isolates, (with of course taking into account the proviso that NMR analysis will be required to confirm the novelty of these metabolite); but also, to visualize more potential new metabolites and more biological activities which were all clearly linked

together in one picture through bioactivity mapping. This visual mapping allowed us to highlight which of the metabolites had high bioactivity scores and which one has not been identified by comparison within the GNPS libraries, which, were therefore more interesting for us on further analyse. This was evident in Chapter 3 with the bioactivity mapping of the anti-fungal activity against *C. glabrata* CBS138, which highlighted two promising molecules which showed bioactivity scores of 60% and 64% of bioactivity respectively, together with one bioactive cluster.

In addition, we expanded upon the approach routinely used in previous studies on bioactivity mapping by analysing three different bioactivities simultaneously (antifungal, anti-inflammatory and anti-cancer), as opposed to focusing on one activity alone which is typical of similar types of studies (*Nothias et al*, 2018). To our knowledge, this is one of the first studies to maps several different bioactivities to one metabolome using molecular networks, offering a template by which to greatly expand upon and exploit future large-scale molecular networks for novel bioactive compounds.

Furthermore, by linking LC-MS data to bioactivity mapping, we were able to predict potential novel bioactive metabolites within molecular networks with greater certainty and refine the pool of potential candidates for analysis. For example, combining anti-cancer activity of the human melanoma cell line, A2058, to $[M+H]^+$ mass data (generated by LC-MS) highlighted one bioactive cluster of interest containing 13 bioactive molecules (out of a total of 177 metabolites) with bioactivity scores ranged from 60% to 89% of bioactivity, including two that are potentially new metabolites. With that, the total pool of candidates identified through three bioactivity mappings reduced the pool of potential candidates for analysis from 117 to 19; however, cross-referencing with LC-MS data reduced this pool even further. Some candidates,

for example, contained $[M+H]^+$ data that mapped to previously identified compounds such as Desferrioxamine E. Some other candidate also showed double activity, such as for example, 1 of the 19 interesting molecules of interest showed 60% bioactivity against the human melanoma cancer cell line A2058 and 88% of anti-inflammatory activity against the LPS induced TNF-α production in Thp-1 cells (Chapter 3).

Therefore, the approach applied in this study can offer greater clarity when identifying potential novel candidates for analysis and when identifying bioactive candidates, thereby greatly increasing the efficiency of future novel bioactive discovery studies.

References

Antoraz, S., Santamaria, R. I., Diaz, M., Sanz, D., & Rodriguez, H. (2015). Toward a new focus in antibiotic and drug discovery from the *Streptomyces* arsenal. *Front Microbiol, 6*, 461. doi:10.3389/fmicb.2015.00461

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. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417*, 141–147.

Dadgostar, P. (2019). Antimicrobial Resistance: Implications and Costs. Infection and Drug Resistance 2019:12 3903–3910. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6929930/pdf/idr-12-3903.pdf.

Firsova, D., Mahajan, N., Solanki, H., Morrow, C. and Thomas, O.P. 2017. Current Status and Perspectives in Marine Biodiscovery. In Bioprospecting (pp. 29-50). Springer, Cham.

Gribble, G. W. (2015). Biological Activity of Recently Discovered Halogenated Marine Natural Products. *Marine Drugs, 13*(7), 4044-4136. doi:10.3390/md13074044.

Jakubiec-Krzesniak, K., Rajnisz-Mateusiak, A., Guspiel, A., Ziemska, J., Solecka, J. Secondary Metabolites of Actinomycetes and their Antibacterial, Antifungal and Antiviral Properties. Pol J Microbiol. 2018 Sep; 67(3): 259–272. doi: 10.21307/pjm-2018-048.

Nothias, L. F., Nothias-Esposito, M., da Silva, R., Wang, M., Protsyuk, I., Zhang, Z., Sarvepalli, A., Leyssen, P., Touboul, D., Costa, J., Paolini, J., Alexandrov, T., Litaudon, M., & Dorrestein, P. C. (2018). Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. Journal of natural products, 81(4), 758–767. https://doi.org/10.1021/acs.jnatprod.7b00737.

O'Neill, J. 2014. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. 2014. Available at: https://amr-review.org/Publications.html .

Pullen, C., Schmitz, P., Meurer, K., Bamberg, D. D., Lohmann, S., De Castro França, S., Groth, I., Schlegel, B., Möllmann, U., Gollmick, F., Gräfe, U., & Leistner, E. (2002). New and bioactive compounds from *Streptomyces* strains residing in the wood of

Celastraceae. *Planta*, *216*(1), 162–167. https://doi.org/10.1007/s00425-002-0874-6.

Romano, S., Jackson, S. A., Patry, S., Dobson, A. (2018). Extending the "One Strain Many Compounds" (OSMAC) Principle to Marine Microorganisms. *Mar. Drugs* **2018**, *16*(7), 244; **https://doi.org/10.3390/md16070244.**

Schwarz, J., Hubmann, G., Rosenthal, K., Lütz, S. (2021). Triaging of Culture Conditions for Enhanced Secondary Metabolite Diversity from Different Bacteria. Biomolecules 11, 193. https://doi.org/10.3390/biom11020193.

Williamson, N.R., Fineran, P.C., Leeper, F.J. and Salmond, G.P. 2006. The biosynthesis and regulation of bacterial prodiginines. Nature Reviews Microbiology, 4(12), p.887.

Yang, Z., He, J., Wei, X., Ju, J., Ma, J. (2020). Exploration and genome mining of natural products from marine *Streptomyces*. Appl Micro Biotech 104, 67-76.

Appendix

Publication by the author

Romano, Stefano; Jackson, Stephen A.; Patry, Sloane; Dobson, Alan D. W. Extending the "one strain many compounds" (OSMAC) principle to marine microorganisms. *Mar. Drugs* **2018**, *16*(7), 244.

Other activities

Poster presentation communication

Sloane Patry, Lekha Margassery, Alan D.W. Dobson. Genomic mining to identify novel bioactive compounds in marine bacterial strains. *UCC Annual poster session*, UCC, Cork, Ireland, December 05, 2017.

Sloane Patry, Lekha Margassery, Alan D.W. Dobson. Genomic mining to identify novel bioactive compounds in marine bacterial strains. *Microbiology society, Annual Conference*, Birmingham, UK, 10-13 April 2018.

Sloane Patry, Antigoni-Angeliki Kyritsi, Scott A Jarmusch, Rainer Ebel, Marcel Jaspars and Alan D.W. Dobson. Utilization of Genomic mining and LC-MS to identify novel biosynthetic pathways encoding novel bioactive compounds in marine bacterial strains. *BIOPROSP_19, 9th International Conference on Marine Biotechnology*, Tromsø, Norway, 25 – 27 February 2019.

Sloane Patry, Antigoni-Angeliki Kyritsi, Scott A Jarmusch, Rainer Ebel, Stephen Jackson, Marcel Jaspars and Alan D.W. Dobson. Utilization of OSMAC, Liquid Chromatography-Mass Spectrometry and Metabolomic analysis to identify secondary metabolites and potential novel

bioactive compounds in marine bacterial strains. *UCC Annual poster session*, UCC, Cork, Ireland, December 02, 2019.

Marketing communications

May 2018: Creation of Marpipe's flyer. *Eight countries & eleven partners: improving the flow in the pipeline of the next generation of marine bio-discovery scientists*.

2017-2020: Working alongside the others ESRs for updating the Marpipe Facebook and Marpipe Twitter accounts.

2017-2020: International networking activities and talks at international conferences, workshops and meetings.

PhD progress meeting communications

University College Cork:

April 2017-2020: Several Lab meetings with 10 min oral presentations, UCC, Cork, Ireland.

December 2017: *UCC Annual poster session*, UCC, Cork, Ireland, December 05, 2017.

December 2019: *UCC Annual poster session*, UCC, Cork, Ireland, December 02, 2019.

Marine Biodiscovery centre, University of Aberdeen:

October 2018: Lab meeting – Updated Project's presentation *(05.10.2018).*

Marpipe:

October 2017: PhD progress meeting 1 – KU Leuven, Leuven (Belgium).

May 2018: PhD progress meeting 2 – CNR, Naples (Italia).

October 2018: Marpipe Midterm review – Brussel (Belgium).

February 2019: PhD progress meeting 3 – UiT, Tromsø (Norway).

December 2019: PhD progress meeting 4 – SZN, Naples (Italia).

Short technical or scientific courses

Short scientific course: Zebrafish as an *in vivo* model for safety assessment of drugs. KU Leuven, Leuven, Belgium. October 16-19, 2017.

Short technical course: Scientific communication and grant proposal writing. CNR, Naples, Italy. May 8-10, 2018.

Short scientific course: Organic structure elucidation, Fundacion MEDINA, Granada, Spain. December 10-11, 2018.

BIOPROSP_19 pre-workshop: Marine biotech entrepreneurship: Capturing the value of scientific results, University of Tromso, Tromsø, Norway. February 25, 2019.

Short scientific and technical course: Bioinformatics workshop, MG-RAST Practical. University College Cork, Cork, Ireland. August 20, 2019.

Demonstration activities

As part of UCC, I did some demonstrations to Bachelor's students, during my second year of PhD in 2018. I demonstrated MB2006 and ML2001 modules with specific sessions involving the following topics: Minimal Inhibitory Concentration, Nitrate and Nitrification; Bacteriophage and Controls of microbes using chemical agents.

International collaborations:

2019: University of Tromso, Marbio facilities – anticancer and anti-inflammatory assays have been performed on all the OSMAC crude extracts - performed by the technicians Kirsti Helland and Marte Albrigtsen.

2020: University of Leuven – for toxicity assays on the fractions of the SM9-M19 crude extract – performed by ESR6 Arianna Guisti.

2020: University of Tromso - LC-MS/MS analysis on the fractions – performed by ESR3 Yannik Schneider.

International secondments:

As part of the EU MSCA Horizon 2020 MarPipe ESR Programme, I was required to spend several secondments at academic institutions partners and/or industrial partners. As this is an International Training Network (ITN), the goal for spending several secondments abroad is to acquire as much transferable knowledge and skills as possible during the PhD study, but also to learn about new cultures, new languages, and also to live and work in different environments

with different nationalities. My ESR project, namely ESR7, required me to do 3 secondment abroad for a total length of 6 to 7 months flexible abroad out of 36 months of my PhD.

I went to my first MarPipe ESR secondment to the Marine Biodiscovery Centre, at the University of Aberdeen, in Scotland (UK) to learn about LC-MS/MS experiments and analysis. The goal of this secondment was for me to perform LC-MS/MS analysis on all my crude extracts samples, including both organic phase and aqueous phases from the following isolate: B226SN104, SM9, SM3, SM18 and B188M101; to analyses and process the data of all the crude extracts using manual de-replication. During this secondment, I was under the supervision of Prof. Marcel Jaspars, Dr. Rainer Ebel and ESR4 Antigoni Kyritsi. This secondment was for a duration of 5 months (one month in July 2018 for most of the LC-MS/MS experiments, then from September to December 2018 for the remaining lab experiments and the learning of the processing data using manual de-replication). In addition, this secondment allowed me to develop contacts with scientists in chemistry and metabolomics who provided me with the tools and support I needed to conduct metabolomics analysis when I returned to UCC.

For my second MarPipe ESR secondment, I went to the ABS-int in Bruges (Belgium) with all the other ESRs to learn about Marine biotechnology entrepreneurship, Biosafety regulations and Intellectual property. The goal of the secondment was to familiarize all the ESRs with the non-scientific aspects of the marine biodiscovery value chain related to entrepreneurship, law and policy. The following key topics were given: commercialisation of scientific results; creation of business models and business plans; pitching, communication and presentations; negotiation tactics; basics of intellectual property and biosafety regulations. Through this

secondment, we all gained valuable knowledge and experience in terms of how to valorise our scientific research, as well as, how to create our own business. During this secondment, I was under the supervision of Dr. Thomas Vanagt and ESR8 Jane Eva Collins. The duration of this secondment was 7 days in Bruges, Belgium.

For my third and last MarPipe ESR secondment, I went to the Consiligo Nazionale Delle Ricerche, Institute of Protein Biochemistry (IBP-CNR), at Napoli (Italia) to learn about liquid inhibition assays, anti-biofilm assays and SPE fractionation. The goal of this secondment was for me to perform liquid inhibition assays, anti-biofilms assays and SPE fractionation. During this secondment I was under the supervision of Donatella De Pascale and ESR1 Grant January, for a duration of one month (December 2019).

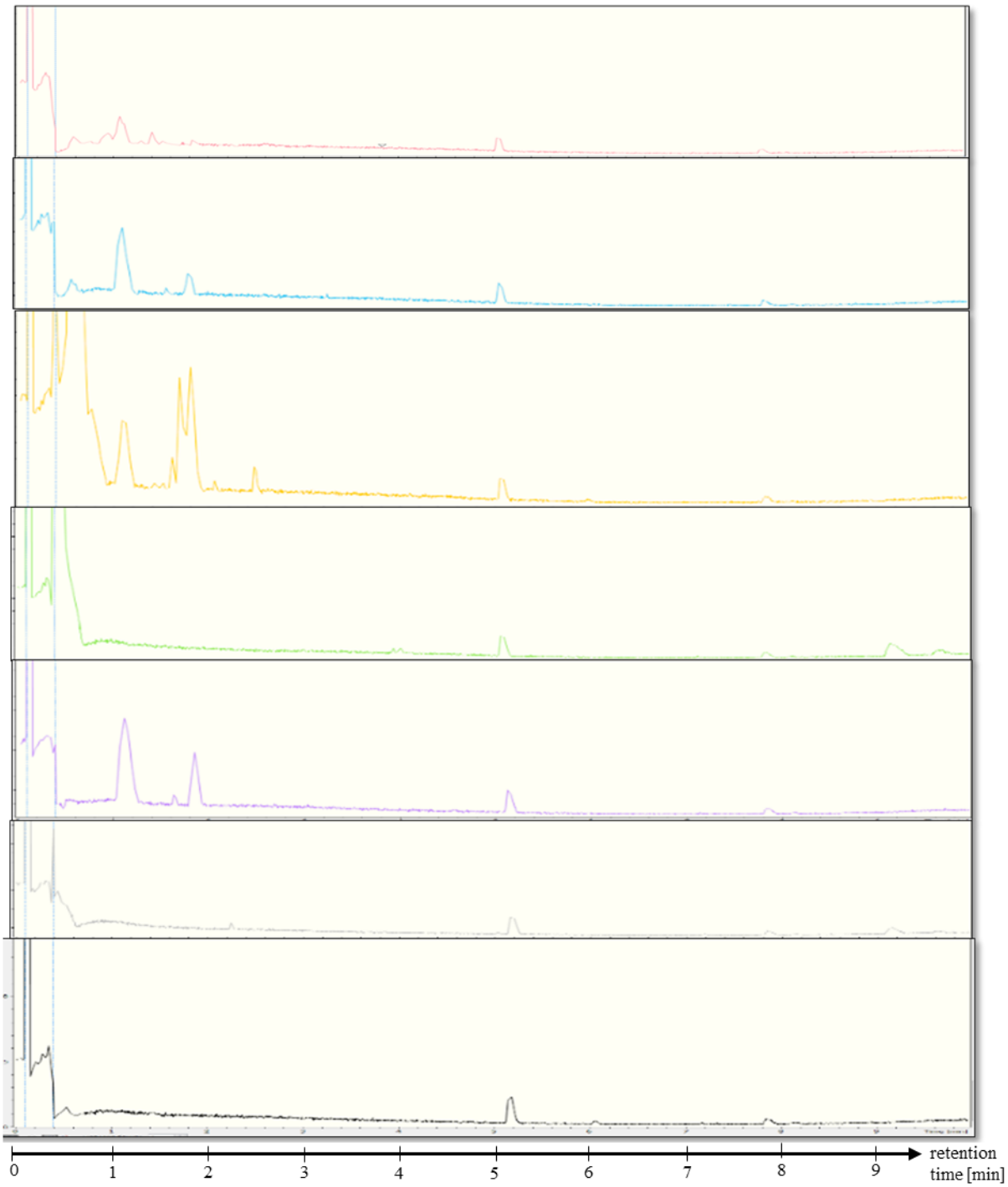
Overall, my secondment experiences were very beneficial to me, both in terms of their relevance to my scientific studies, but also to my own personal growth and development. In terms of the scientific project itself, by giving me new viewpoints and perspectives outside of those I had as a microbiologist I was able to adopt new strategies to overcome difficulties or obtain new understanding. Furthermore, I obtained scientific transferable skills such as chemical analysis, metabolomics, entrepreneurship, biosafety regulation and intellectual property skills, as further development of my microbiological skills e.g. via the liquid inhibition, antioxidant, and anti-biofilm assays. In addition, my PhD-ESR in UCC, the secondments and the Marpipe progress meeting allowed to develop my skills of working on international multi-culture environments.

There is a huge bridge between the starting position as the ‘PhD student’, who often is only able to understand and follow the instruction and expertise of their supervisors while

discovering and learning about their projects, and the destination of the ‘PhD expert’ who is capable of investigating their own research independently, following their own knowledge, research, background and expertise. I feel this is similar to the bridge between a teenager and a full adult. Instead of being simply a PhD student but also an ESR professional employee, I was given a greater sense of independence, thus making it easier to follow my scientific intuition. Furthermore, the secondments really helped me in the evolution of my scientific knowledge and expertise. Thanks to these, I feel like I have been able to cross the bridge from student to expert – I am really grateful to have been able to benefit from it. This carries into my development as a human being, since I was allowed to develop my adaptation skills with new cultures, new languages, new environments, and new people (many of whom are now new friends).

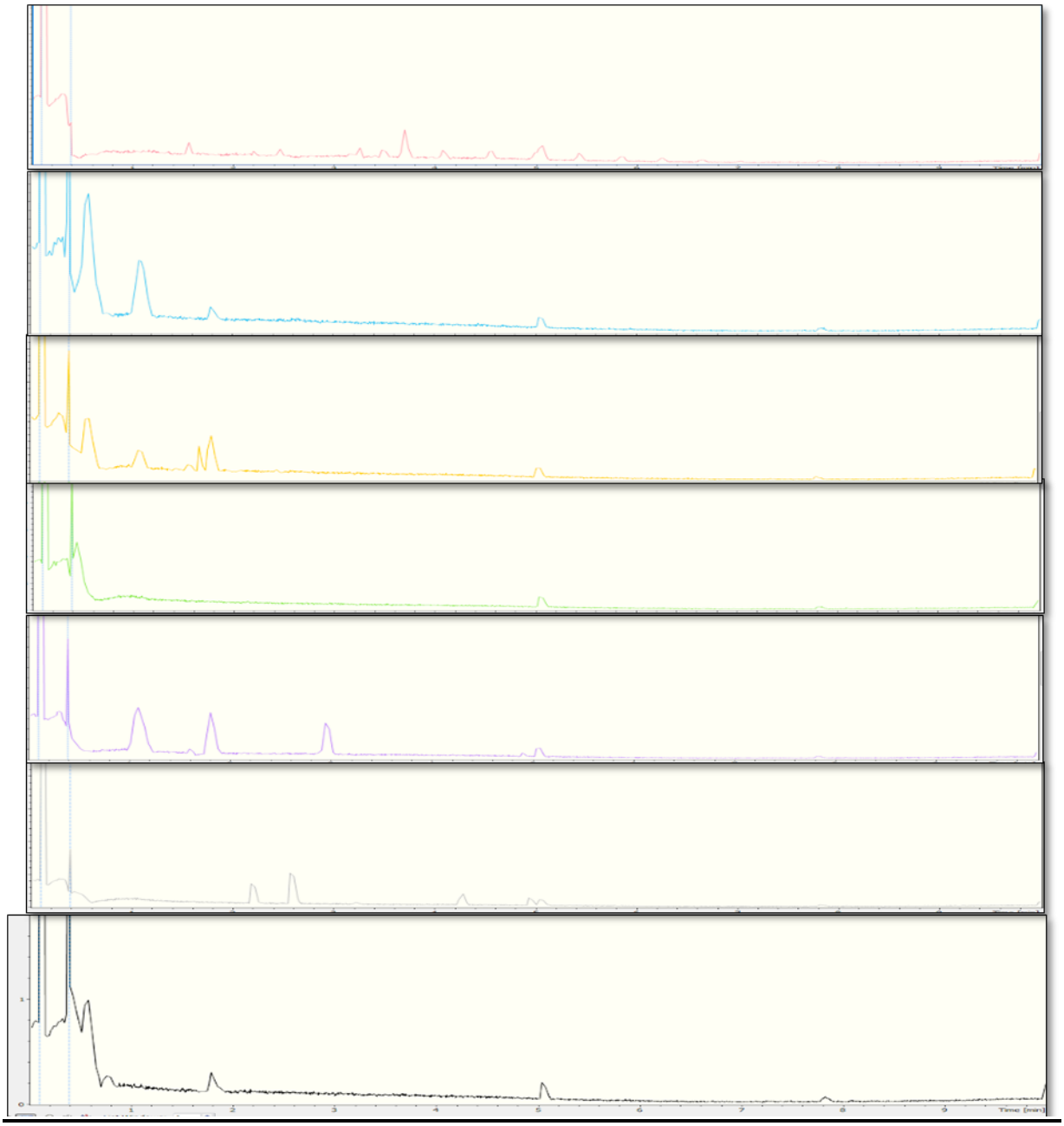
Supplementary figures

Supplementary figure 1 need to be read together with supplementary table 1, same for supplementary figure 2 to 10, with supplementary table 2 to 10.



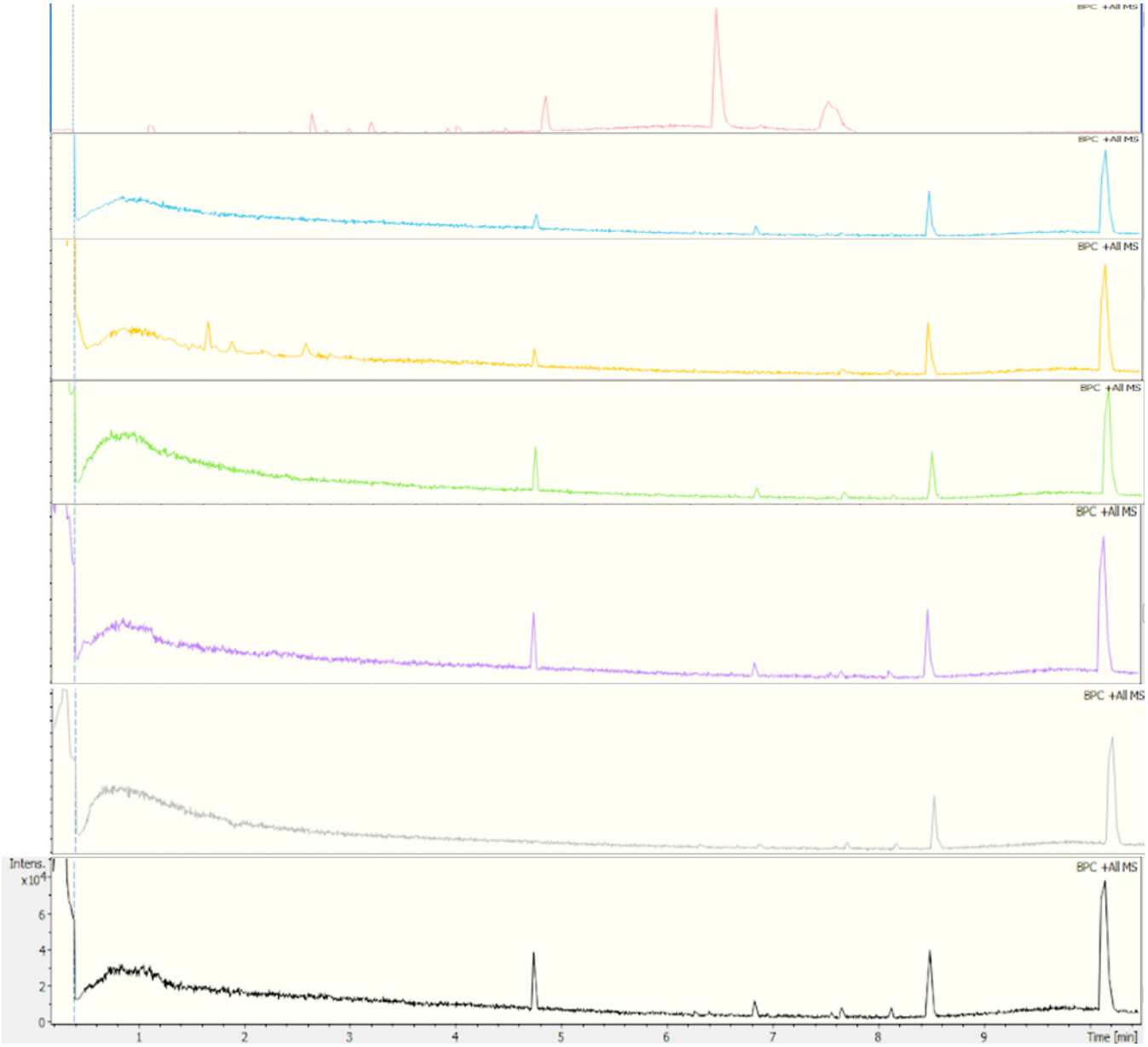
Supplementary Figure 1: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from SM3 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Colour code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



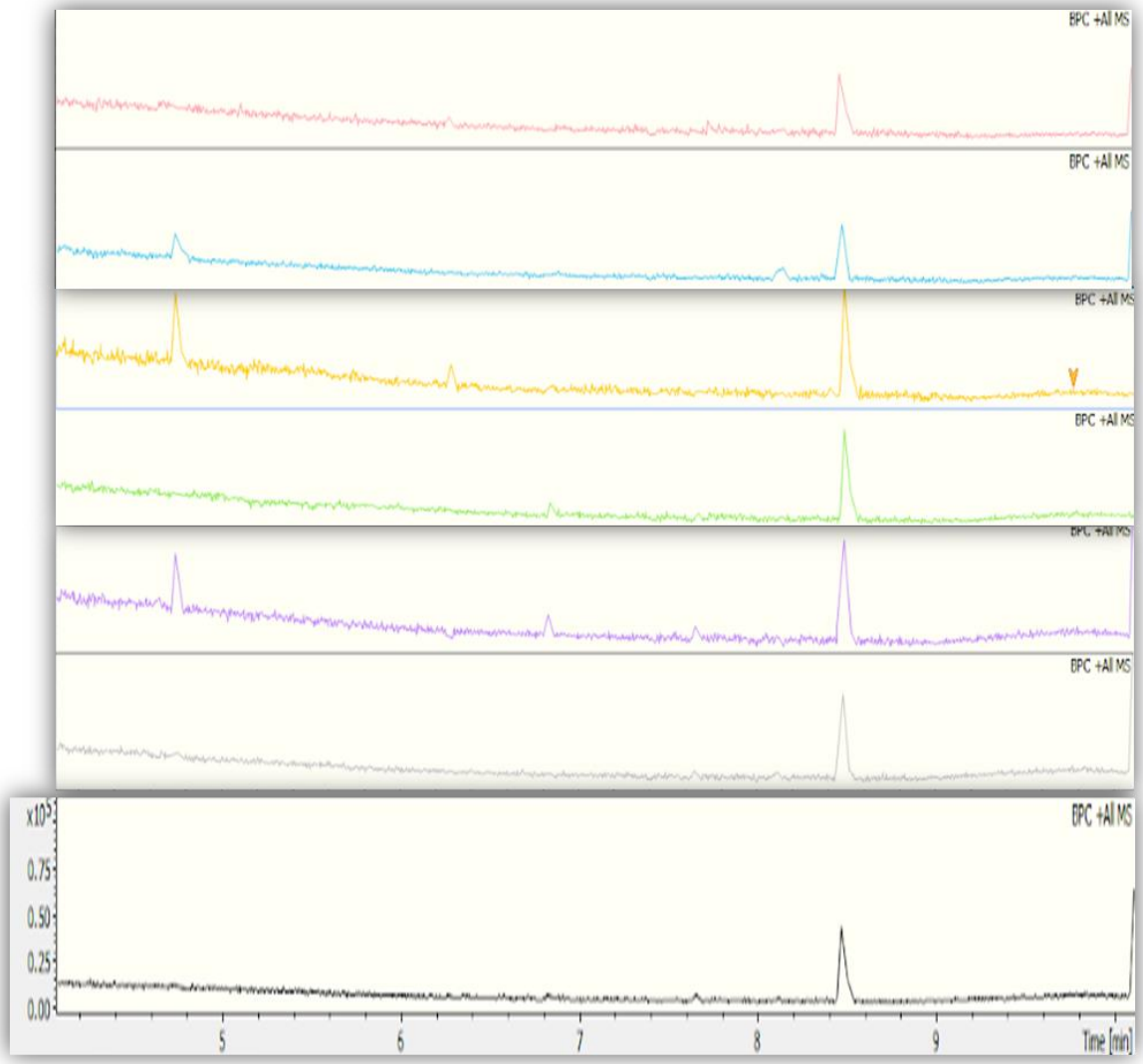
Supplementary Figure 2: Chromatograms from LC-MS analysis of the organic crude extract samples from SM3 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



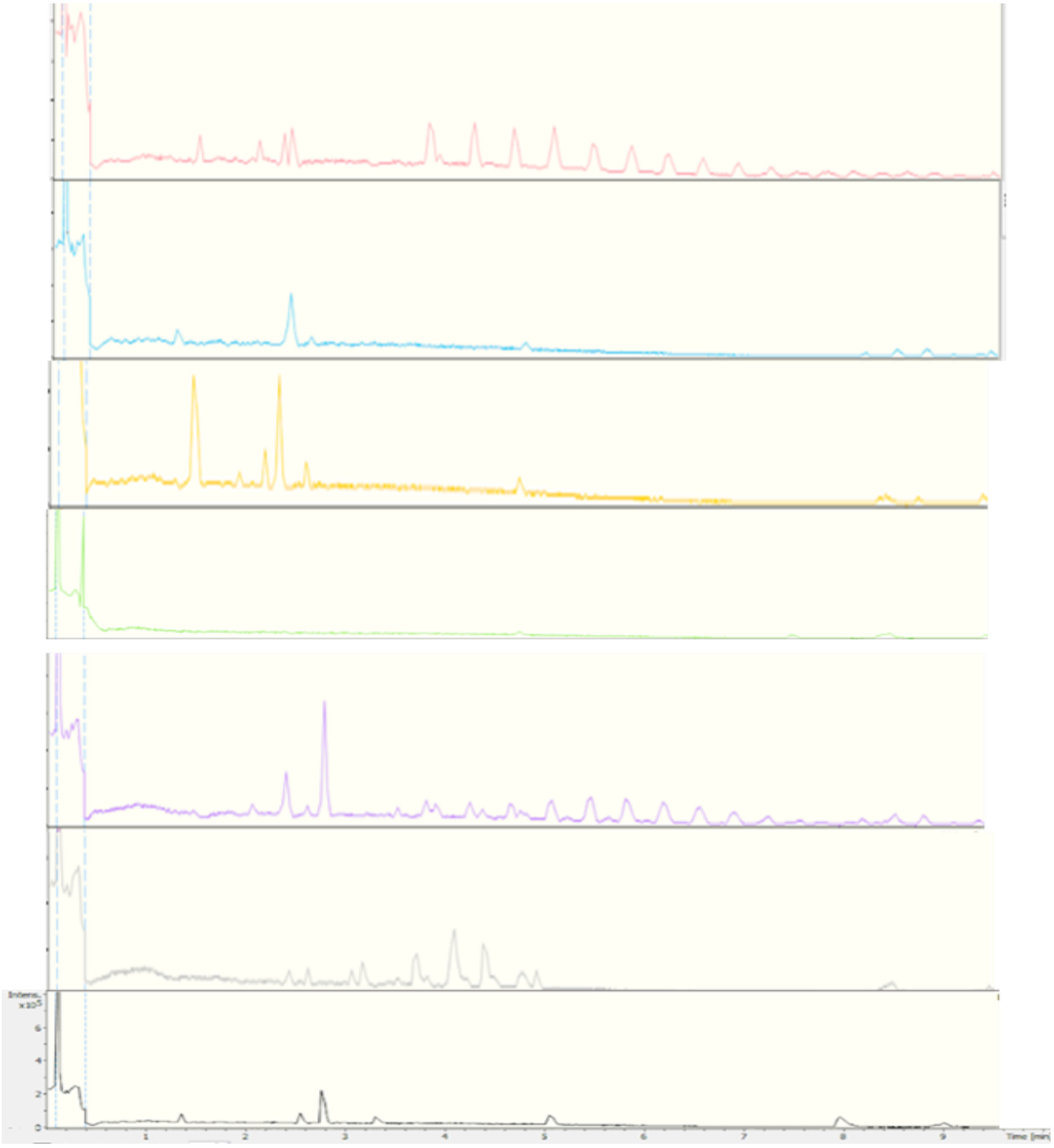
Supplementary Figure 3: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from SM3 when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



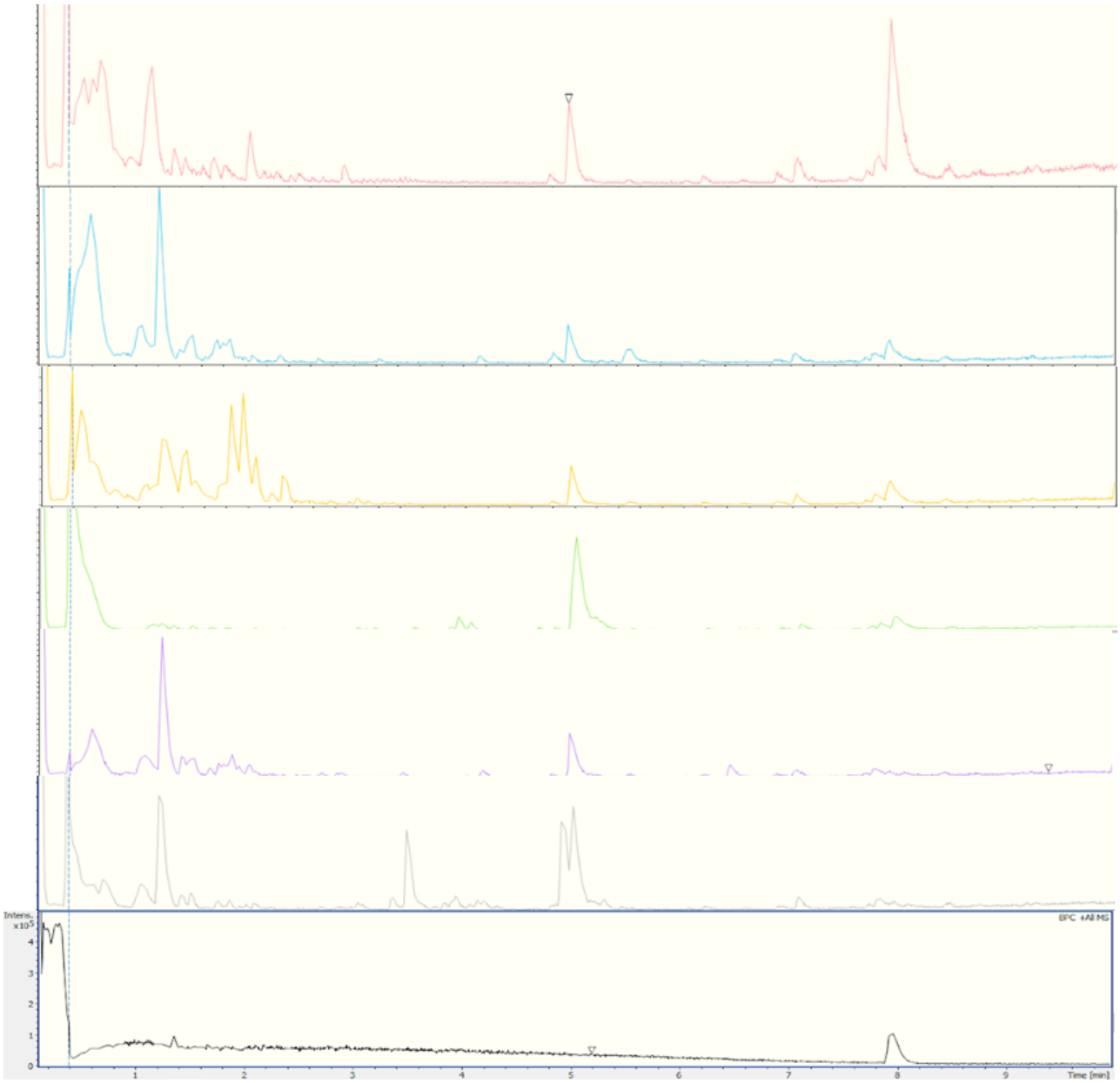
Supplementary Figure 4: Chromatograms from LC-MS analysis of the organic phase crude extract samples from SM3 when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



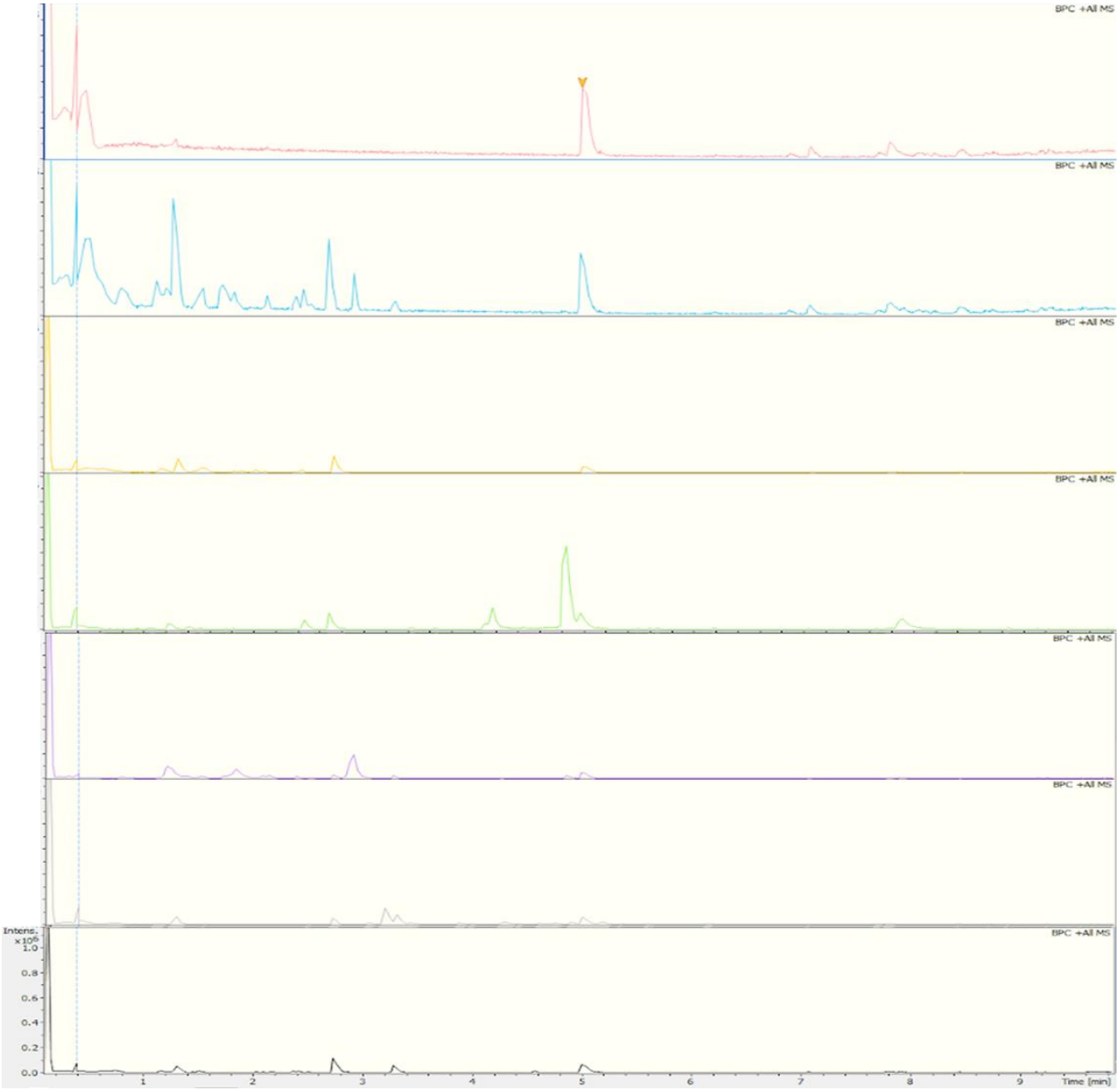
Supplementary Figure 5: Chromatograms from LC-MS analysis of the organic phase crude extract samples from B226SN104 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



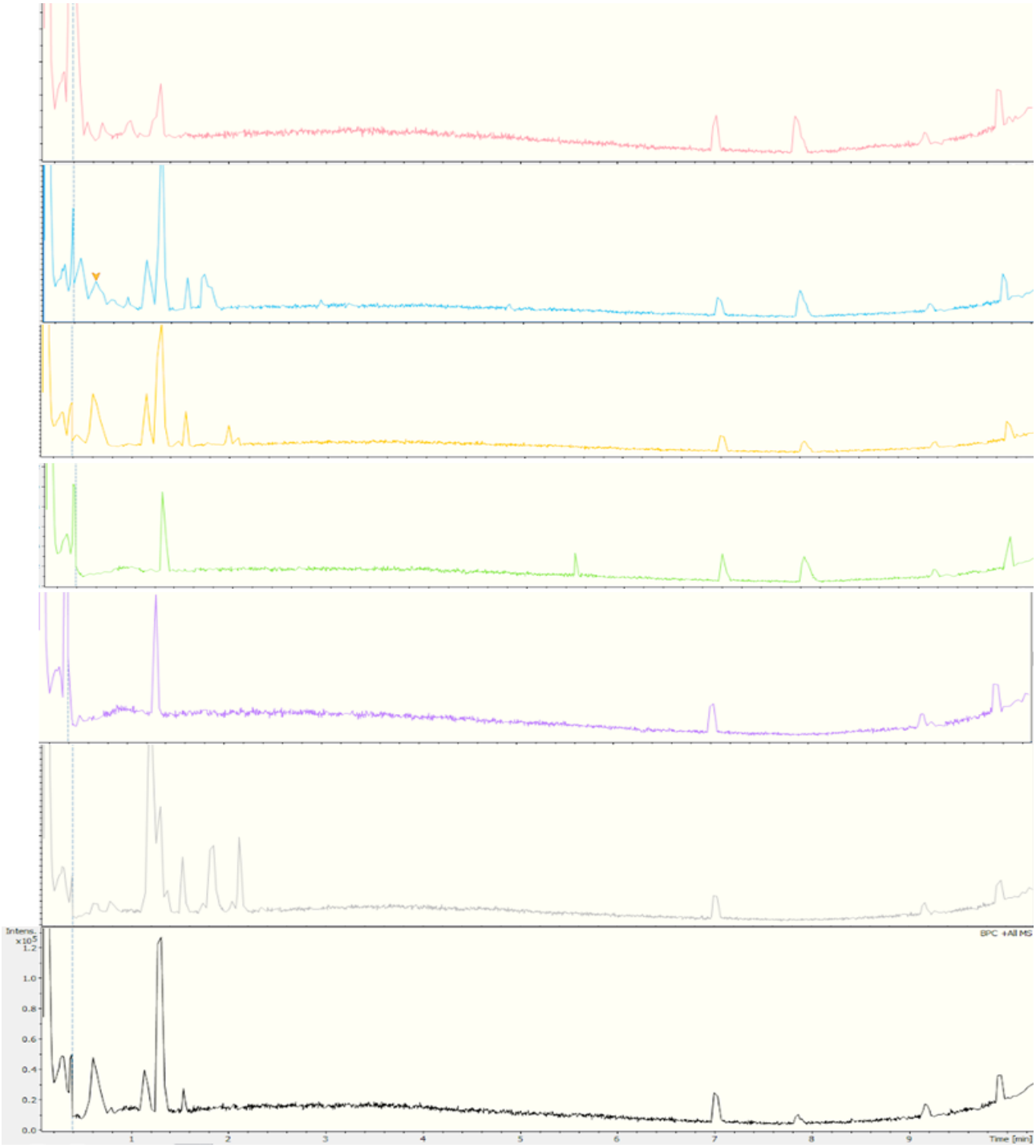
Supplementary Figure 6: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from B226SN104 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



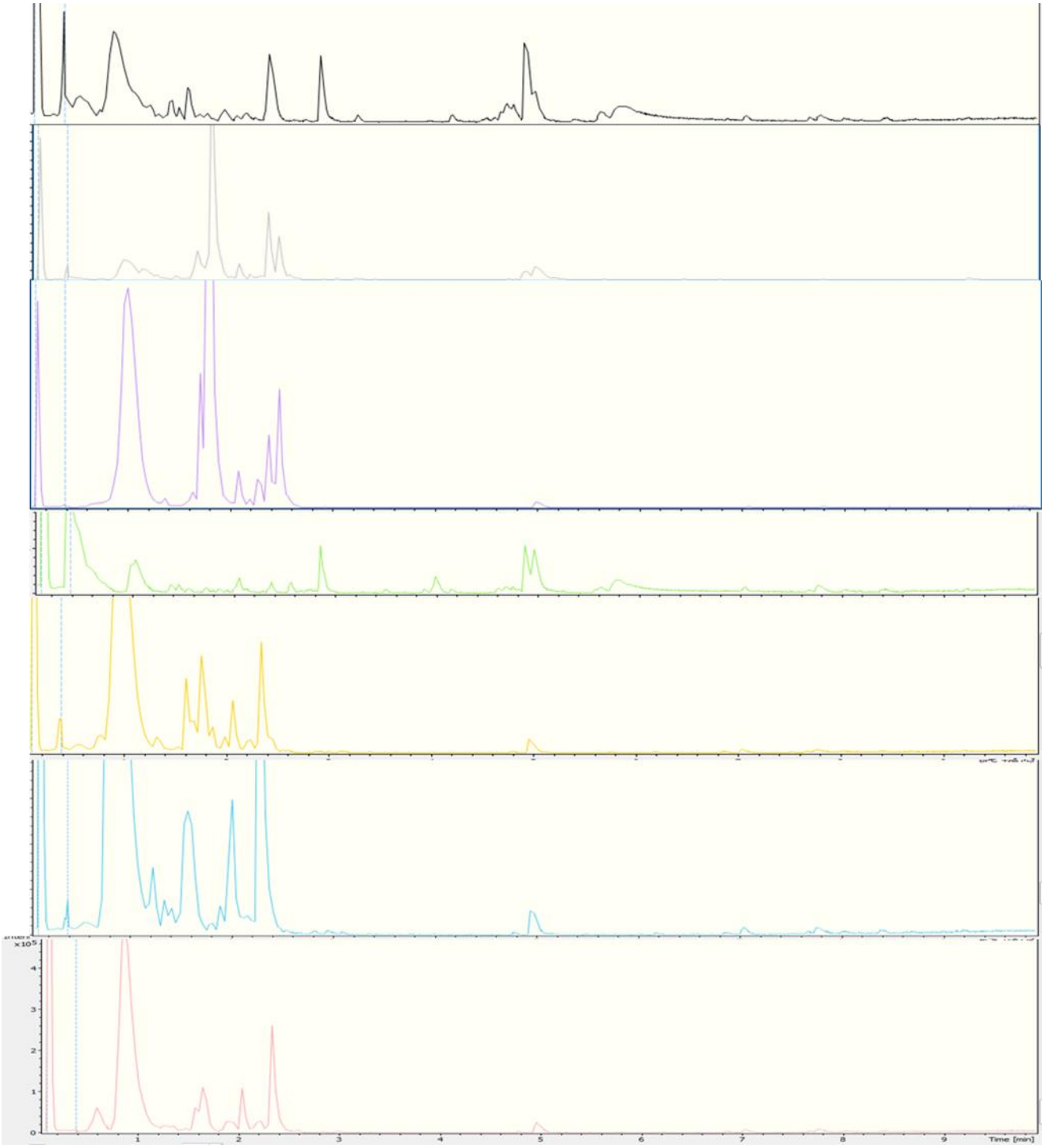
Supplementary Figure 7: Chromatograms from LC-MS analysis of the organic phase crude extract samples from B226SN104 when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



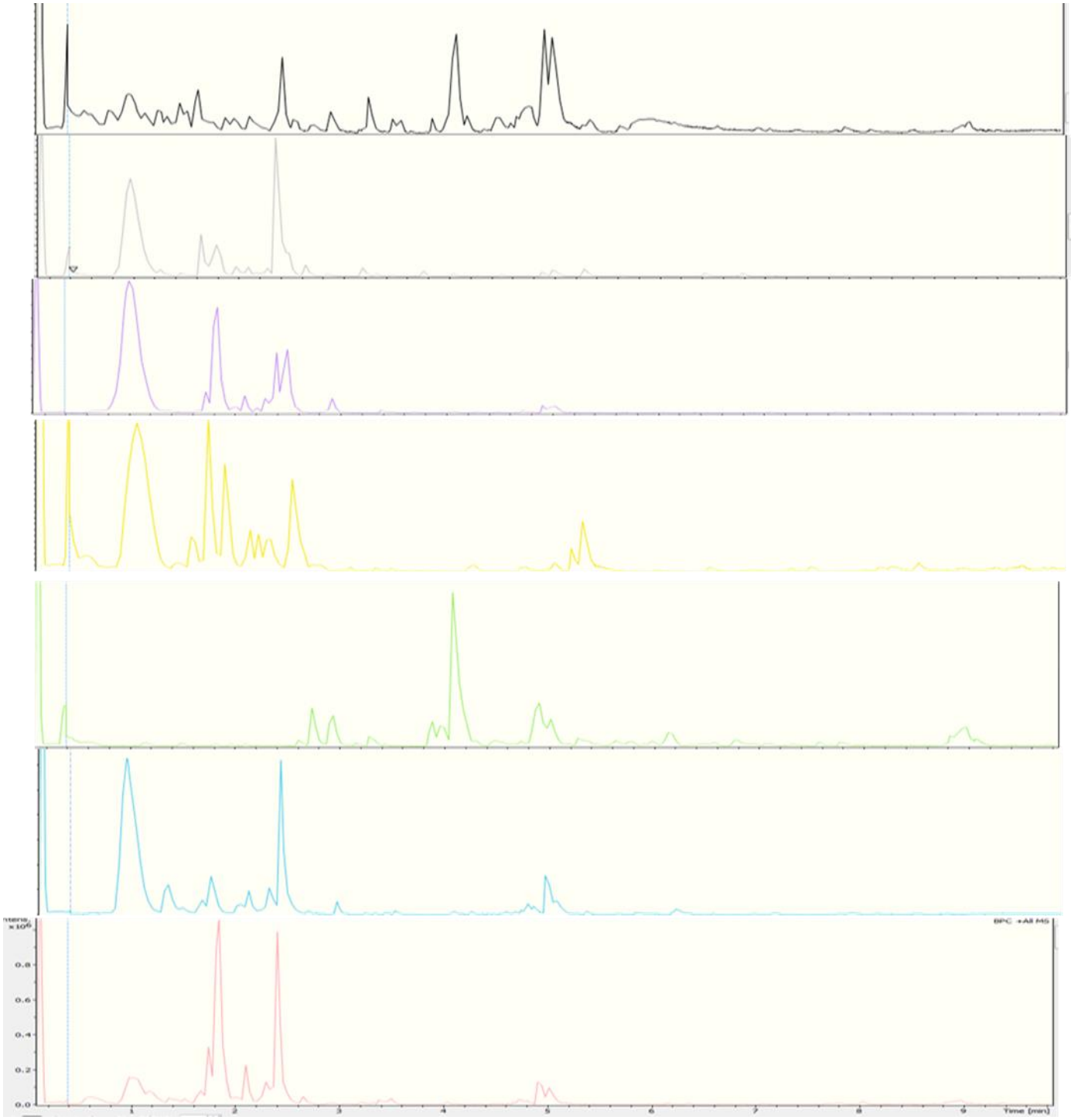
Supplementary Figure 8: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from B226SN104 – when fermented at 30°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP.

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



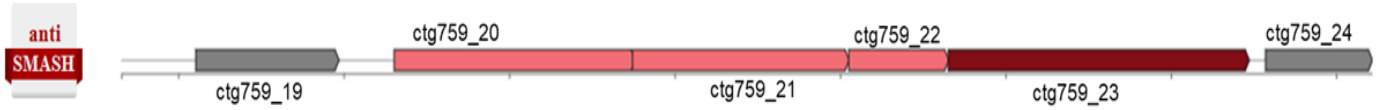
Supplementary Figure 9: Chromatograms from LC-MS analysis of the aqueous phase crude extract samples from SM9 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



Supplementary Figure 10: Chromatograms from LC-MS analysis of the organic phase crude extract samples from SM9 when fermented at 28°C in seven media (M19, M400, MMM, OM, SGG, SM and SYP).

Polar compounds elute in the beginning of the run, while moving towards the end of the run (10 min) we observe more hydrophobic compounds. Anything after 10min comes from flushing the column, so these are not considered real peaks. Color code: M19 pink, M400 blue, MMM yellow, OM gree, SGG purple, SM grey and SYP black.



Gene details

ctg759_19

Locus tag: ctg759_19
Protein ID: None
Gene: None
Location: 15,105 - 15,971, (total: 867 nt)

other (smcogs) SMCOG1200:vibriobactin utilization protein ViuB (Score: 256.8; E-value: 3.2e-78)

ctg759_20

Locus tag: ctg759_20
Protein ID: None
Gene: None
Location: 16,305 - 17,762, (total: 1458 nt)

biosynthetic-additional (smcogs) SMCOG1180:Decarboxylase, pyridoxal-dependent (Score: 370.9; E-value: 1.4e-112)

ctg759_21

Locus tag: ctg759_21
Protein ID: None
Gene: None
Location: 17,746 - 19,056, (total: 1311 nt)

biosynthetic-additional (smcogs) SMCOG1080:lysine/ornithine N-monooxygenase (Score: 423.7; E-value: 1.1e-128)

ctg759_22

Locus tag: ctg759_22
Protein ID: None
Gene: None
Location: 19,053 - 19,658, (total: 606 nt)

biosynthetic-additional (smcogs) SMCOG1203:putative siderophore biosynthesis protein (Score: 127; E-value: 2.1e-38)

ctg759_23

Locus tag: ctg759_23
Protein ID: None
Gene: None
Location: 19,655 - 21,475, (total: 1821 nt)

biosynthetic (rule-based-clusters) siderophore: lucA_lucC

ctg759_24

Locus tag: ctg759_24
Protein ID: None
Gene: None
Location: 21,573 - 22,220, (total: 648 nt)

ctg759_19

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
MULTISPECIES: siderophore-interacting protein [Streptomyces]	566	566	100%	0.0	100.00%	WP_037844137.1
MULTISPECIES: siderophore-interacting protein [Streptomyces]	563	563	100%	0.0	99.31%	WP_003946185.1
MULTISPECIES: siderophore-interacting protein [Streptomyces]	563	563	100%	0.0	99.65%	WP_008407734.1
MULTISPECIES: siderophore-interacting protein [Streptomyces]	562	562	100%	0.0	99.31%	WP_018894986.1
MULTISPECIES: siderophore-interacting protein [Streptomyces]	562	562	100%	0.0	99.31%	WP_103487675.1
NADPH-dependent ferric siderophore reductase [Streptomyces albidoflavus]	562	562	100%	0.0	99.31%	PJT48323.1
siderophore-interacting protein [Streptomyces albidoflavus]	561	561	100%	0.0	99.31%	WP_095710889.1
siderophore-interacting protein [Streptomyces albidoflavus]	561	561	100%	0.0	99.96%	WP_146583071.1

ctg759_20

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
MULTISPECIES: aspartate aminotransferase family protein [Streptomyces]	954	954	100%	0.0	96.49%	WP_100462630.1
aspartate aminotransferase family protein [Streptomyces sp. S6]	926	926	94%	0.0	99.34%	WP_121843041.1
aspartate aminotransferase family protein [Streptomyces sp. GSSD-12]	828	828	98%	0.0	84.13%	WP_114659356.1
aspartate aminotransferase family protein [Streptomyces sp. AC230]	827	827	98%	0.0	83.72%	WP_114622395.1
aspartate aminotransferase family protein [Streptomyces scopuliridis]	818	818	98%	0.0	83.75%	WP_030354991.1
lysine decarboxylase DesA [Streptomyces glaucescens]	817	817	98%	0.0	82.67%	WP_088738470.1
lysine decarboxylase DesA [Streptomyces pharetrae]	816	816	98%	0.0	82.88%	WP_095167599.1
lysine decarboxylase DesA [Streptomyces glaucescens]	815	815	98%	0.0	82.88%	WP_043501075.1

ctg759_21

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
MULTISPECIES: SidA/IucD/PvdA family monooxygenase [Streptomyces]	893	893	100%	0.0	99.77%	WP_003946187.1
SidA/IucD/PvdA family monooxygenase [Streptomyces sampsonii]	892	892	100%	0.0	99.54%	WP_067412180.1
MULTISPECIES: SidA/IucD/PvdA family monooxygenase [Streptomyces]	892	892	100%	0.0	99.54%	WP_008407736.1
MULTISPECIES: SidA/IucD/PvdA family monooxygenase [Streptomyces]	890	890	100%	0.0	99.54%	WP_103487676.1
SidA/IucD/PvdA family monooxygenase [Streptomyces albidoflavus]	890	890	100%	0.0	99.31%	WP_129529902.1
MULTISPECIES: SidA/IucD/PvdA family monooxygenase [Streptomyces]	889	889	100%	0.0	99.31%	WP_031178094.1
SidA/IucD/PvdA family monooxygenase [Streptomyces albidoflavus]	889	889	100%	0.0	99.31%	WP_129558805.1

ctg759_22

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
MULTISPECIES: acetyltransferase [Streptomyces]	409	409	100%	2e-144	100.00%	WP_003946188.1
acetyltransferase [Streptomyces albidoflavus]	408	408	100%	5e-144	99.50%	WP_033239093.1
MULTISPECIES: acetyltransferase [Streptomyces]	407	407	100%	1e-143	99.50%	WP_103487677.1
acetyltransferase [Streptomyces albidoflavus]	407	407	100%	2e-143	99.50%	WP_129817813.1
acetyltransferase [Streptomyces albidoflavus]	407	407	100%	2e-143	99.50%	WP_129826784.1
acetyltransferase [Streptomyces sp. CS227]	406	406	100%	2e-143	99.50%	WP_087775382.1
MULTISPECIES: acetyltransferase [Streptomyces]	406	406	100%	4e-143	99.00%	WP_067412183.1
MULTISPECIES: acetyltransferase [Streptomyces]	405	405	100%	7e-143	99.00%	WP_008407740.1
acetyltransferase [Streptomyces albidoflavus]	405	405	100%	1e-142	98.51%	WP_071337532.1

ctg759_23

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
MULTISPECIES: IucA/IucC family siderophore biosynthesis protein [Streptomyces]	1238	1238	100%	0.0	100.00%	WP_103506654.1
IucA/IucC family siderophore biosynthesis protein [Streptomyces albidoflavus]	1232	1232	100%	0.0	99.67%	WP_129839830.1
MULTISPECIES: IucA/IucC family siderophore biosynthesis protein [Streptomyces]	1231	1231	100%	0.0	99.67%	WP_018894988.1
IucA/IucC family siderophore biosynthesis protein [Streptomyces wadayamensis]	1230	1230	100%	0.0	99.50%	WP_049078146.1
MULTISPECIES: IucA/IucC family siderophore biosynthesis protein [Streptomyces]	1229	1229	100%	0.0	99.50%	WP_031178095.1
IucA/IucC family siderophore biosynthesis protein [Streptomyces albidoflavus]	1229	1229	100%	0.0	99.50%	WP_129819757.1
IucA/IucC family siderophore biosynthesis protein [Streptomyces sp. S6]	1229	1229	100%	0.0	99.50%	WP_121843040.1

ctg759_24

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
DUF4429 domain-containing protein [Streptomyces sp. IgraMP-1]	428	428	100%	5e-151	99.54%	WP_129826056.1
DUF4429 domain-containing protein [Streptomyces sp. SM9]	427	427	100%	7e-151	100.00%	WP_103539945.1
Tat pathway signal sequence domain protein [Streptomyces albidoflavus]	428	428	100%	8e-151	100.00%	PJT48318.1
DUF4429 domain-containing protein [Streptomyces sp. S6]	428	428	100%	9e-151	100.00%	WP_121843039.1
DUF4429 domain-containing protein [Streptomyces albidoflavus]	429	429	100%	1e-150	100.00%	WP_129836233.1
MULTISPECIES: DUF4429 domain-containing protein [Streptomyces]	429	429	100%	1e-150	100.00%	WP_031178096.1
MULTISPECIES: DUF4429 domain-containing protein [Streptomyces]	429	429	100%	1e-150	100.00%	WP_049078145.1

Supplementary Figure 11: Gene details from the desferrioxamine gene cluster identified by AntiSMASH, with the related Blastp results from NCBI.

Supplementary tables

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
SM3.1 M19 Water Phase	0.6	625.1592	$C_{21}H_{24}N_{10}O_{13}$	no hit
	1.0	553.2621	$C_{24}H_{36}N_6O_9$	23194934
	1.1	585.3181	$C_{20}H_{48}N_4O_{15}$	no hit
			$C_{20}H_{48}N_4O_{16}$	no hit
	1.3	428.1229	$C_{17}H_{35}NO_{11}$	no hit
	1.4	401.2030	$C_{17}H_{28}N_4O_7$	16600422
				16634341
				16577693
	1.8	627.3282	$C_{22}H_{51}N_4O_{16}$	no hit
	2.3	561.2786	$C_{24}H_{40}N_4O_{11}$	4470667
	5.1	387.1811	$C_{23}H_{22}N_4O_2$	12227
	7.9	403.2317	$C_{17}H_{26}N_{10}O_2$	no hit
	10	391.2844	$C_{24}H_{38}O_4$	6950
SM3.1 M400 Water Phase	0.6	456.2078	$C_{18}H_{33}NO_{12}$	8632101
	1.1	585.3181	$C_{20}H_{48}N_4O_{16}$	no hit
	1.6	513.3761	$C_{25}H_{48}N_6O_5$	no hit
	1.8	625.3142	$C_{14}H_{31}N_{27}O_3$	no hit
	5.1	387.1808	$C_{22}H_{26}O_6$	54
	7.8	403.2325	$C_{20}H_{34}O_8$	8891446
	10	391.2844	$C_{24}H_{38}O_4$	6950
SM3.1 MMM Water Phase	0.6	456.2078	$C_{18}H_{33}NO_{12}$	8632101
	1.1	585.3186	$C_{21}H_{44}N_8O_{11}$	no hit
	1.6	611.2979	$C_{22}H_{42}N_8O_{12}$	no hit
	1.7	455.2143	$C_{20}H_{30}N_{4O8}$	116946
	1.8	625.3131	$C_{22}H_{48}N_4O_{16}$	no hit
	2.1	397.2086	$C_{12}H_{23}N_{13}O_3$	no hit
	2.5	496.2767	$C_{23}H_{37}N_5O_7$	16700487
	5.1	387.1808	$C_{22}H_{26}O_6$	54
	7.8	403.2336	$C_{21}H_{30}N_4O_4$	no hit

	10	391.2844	$C_{24}H_{38}O_4$	6950
SM3.1 OM Water Phase	0.4	505.1756	$C_{15}H_{24}N_{10}O_{10}$	no hit
			$C_{18}H_{32}O_{16}$	164804
	3.0	821.4400	$C_{36}H_{48}N_{22}O_2$	no hit
			$C_{53}H_{68}N_{28}O_8$	no hit
	3.9	1225.5876	$C_{59}H_{84}N_8O_{20}$	no hit
	4.0	1063.5327	$C_{50}H_{66}N_{18}O_9$	no hit
	5.1	387.1808	$C_{22}H_{26}O_6$	54 8322
	7.8	403.2325	$C_{20}H_{34}O_8$	8891446
	9.1	683.5551	$C_{35}H_{70}N_8O_5$	no hit
	10	391.2844	$C_{24}H_{38}O_4$	6950
SM3.1 SGG Water Phase	1.1	585.3179	$C_{32}H_{44}N_2O_8$	10 substances possible
	1.6	611.2975	$C_{21}H_{46}N_4O_{16}$	no hit
	1.8	625.3131	$C_{22}H_{48}N_4O_{16}$	no hit
	5.1	387.1811	$C_{23}H_{22}N_4O_2$	12227
	7.8	403.2325	$C_{20}H_{34}O_8$	8891446
	10	391.2844	$C_{24}H_{38}O_4$	6950
SM3.1 SM Water Phase	0.4	373.1174	$C_6H_{16}N_{10}O_9$	no hit
	2.3	417.1185	$C_{22}H_{16}N_4O_5$	2627683
				2655868
				3118415
	5.1	387.1811	$C_{23}H_{22}N_4O_2$	12227
	7.8	403.2325	$C_{20}H_{34}O_8$	8891446
	9.2	683.5574	$C_{39}H_{74}N_2O_7$	8007098
	10	391.2844	$C_{24}H_{38}O_4$	6950
SM3.1 SYP Water Phase	0.5	436.1441	$C_{14}H_{17}N_{11}O_6$	no hit
	1.1	585.2506	$C_{24}H_{36}N_6O_{11}$	no hit
	5.1	387.1811	$C_{23}H_{22}N_4O_2$	12227
	7.8	403.2325	$C_{20}H_{34}O_8$	8891446
	10	391.2844	$C_{24}H_{38}O_4$	6950

Supplementary Table 1: MS data for SM3.1 aqueous phase crude extracts.

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest. SM3.1 samples represent the SM3 samples which were fermented at 28°**C**, while SM3.2 represent the SM3 samples which were fermented at 30°**C**. For the interpretation of the MS data a lot of different libraries were used, like Chemspider, Antibase, Dictionary of Natural Products, Reaxys etc. Some masses didn't give any hits when searched on these databases, which means that they may be new.

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
SM3.1 M19 Organic Phase	1.6	261.1235	$C_{14}H_{16}N_2O_3$	179735
	2.2	211.1446	$C_{11}H_{18}N_2O_2$	994
		421.2812	$C_{22}H_{36}N_4O_4$	7338 39374
	2.5	245.1286	$C_{14}H_{16}N_2O_2$	401243 401259 401242
	3.2	202.1808	$C_{11}H_{23}NO_2$	35106
	3.5	216.196	$C_{12}H_{25}NO_2$	907502
	3.7	249.0703	$C_{20}H_8$	no hit
			$C_5H_4N_{12}O$	no hit
	4.1	425.3114	$C_{21}H_{44}O_8$	no hit
	4.6	483.3531	$C_{24}H_{50}O_9$	no hit
			$C_{25}H_{46}N_4O_5$	no hit
	5.1	387.1811	$C_{22}H_{26}O_6$	54 8322
	5.4	616.4635	$C_{31}H_{61}N_5O_7$	no hit
			$C_{30}H_{65}NO_{11}$	no hit
	5.9	674.5038	$C_{34}H_{67}N_5O_8$	9171191
			$C_{33}H_{71}NO_{12}$	no hit
	6.2	732.5464	$C_{34}H_{65}N_{15}O_3$	no hit
			$C_{36}H_{77}NO_{13}$	no hit
	6.7	790.5875	$C_{40}H_{79}N_5O_{10}$	no hit
	7.0	Common impurities		
	7.8	403.2321	$C_{24}H_{38}O_4$	11575
			$C_{13}H_{33}N_5O_9$	no hit

SM3.1 M400 Organic Phase	0.6	456.2078	$C_{18}H_{33}NO_{12}$	9628696
	1.1	585.3190	$C_{21}H_{44}N_8O_{11}$	no hit
	1.8	625.3137	$C_{23}H_{44}N_8O_{12}$	no hit
			$C_{22}H_{48}N_4O_{16}$	no hit
	5.1	387.1808	$C_{22}H_{26}O_6$	54
			$C_{23}H_{22}N_4O_2$	12227
	7.8	403.2336	$C_{21}H_{30}N_4O_4$	no hit
			$C_{24}H_{38}O_4$	11575
SM3.1 MMM Organic Phase	0.6	474.2183	$C_{19}H_{31}N_5O_9$	16605691
	1.1	585. 3187	$C_{21}H_{44}N_8O_{11}$	no hit
	1.6	261.1237	$C_{14}H_{16}N_2O_3$	179735
	1.7	455.2141	$C_{20}H_{30}N_4O_8$	116946
	1.8	625.3139	$C_{20}H_{36}N_{18}O_6$	no hit
	5.1	387.1811	$C_{22}H_{26}O_6$	54
			$C_{23}H_{22}N_4O_2$	12227
	7.8	403.2336	$C_{24}H_{38}O_4$	11575
SM3.1 OM Organic Phase	0.4	667.2284	$C_{24}H_{42}O_{21}$	82582
	5.1	387.1804	$C_{22}H_{26}O_6$	8322
			$C_{23}H_{22}N_4O_2$	12227
	7.8	403.2323	$C_{24}H_{38}O_4$	11575
SM3.1 SGG Organic Phase	1.1	585.3189	$C_{21}H_{44}N_8O_{11}$	no hit
	1.8	625.3132	$C_{23}H_{44}N_8O_{12}$	no hit
			$C_{22}H_{48}N_4O_{16}$	no hit
	3.0	445.2322	$C_{24}H_{32}N_2O_6$	2269889
	4.9	655.2764	$C_{36}H_{38}N_4O_8$	411552
	5.1	387.1799	$C_{22}H_{26}O_6$	54
			$C_{23}H_{22}N_4O_2$	12227
	7.8	403.2336	$C_{21}H_{30}N_4O_4$	no hit
			$C_{24}H_{38}O_4$	11575
SM3.1	2.2	417.1181	$C_{22}H_{16}N_4O_5$	2627683

SM Organic Phase	2.6	433.1132	$C_{21}H_{20}O_{10}$	4444098
	4.3	353.2305	$C_{20}H_{32}O_5$	4557
		683.4704	$C_{34}H_{62}N_6O_8$	no hit
	4.9	943.5263	$C_{48}H_{78}O_{18}$	108898
	7.8	403.2323	$C_{24}H_{38}O_4$	11575
SM3.1 SYP Organic Phase	0.6	456.2078	$C_{18}H_{33}NO_{12}$	9628696
	0.8	490.1922	$C_{17}H_{27}N_7O_{10}$	16593356
			$C_{21}H_{31}NO_{12}$	no hit
			$C_{22}H_{27}N_5O_8$	no hit
	1.8	462.1606	$C_{25}H_{18}N_8O_2$	2065860
	5.0	409.1623	$C_{20}H_{20}N_6O_4$	6170831
	7.8	Common impurities		

Supplementary Table 2: MS data for SM3.1 organic phase crude extracts.

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest.

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
SM3.2 M19 Water Phase	1.1	439.2960	$C_{27}H_{38}N_2O_3$	407513
	2.7	579.2935	$C_{30}H_{38}N_6O_6$	16599024
	3.0	421.2238	$C_9H_{24}N_{16}O_4$	No hit
		438.3793	$C_{23}H_{51}NO_6$	no hit
	3.2	403.2327	$C_{20}H_{34}O_8$	11575

			$C_{21}H_{30}N_4O_4$	No hit
	3.9	424.2828	$C_{19}H_{24}N_{10}O_2$	2395726
			$C_{22}H_{32}O_8$	9157986
	4.0	384.1947	$C_{23}H_{45}NO_3$	no hit
			$C_{10}H_{25}N_9O_7$	no hit
			$C_{26}H_{25}NO_2$	no hit
	4.5	465.2767	$C_{22}H_{41}NO_9$	18738112
			$C_{28}H_{36}N_2O_4$	no hit
			$C_{29}H_{32}N_6$	no hit
	4.9	391.2844	$C_{24}H_{38}O_4$	6950
	6.5	663.4521	$C_{24}H_{65}N_5O_{15}$	no hit
			$C_{44}H_{58}N_2O_3$	No hit
	6.9	750.4056	$C_8H_3NO_{40}$	no hit
			$C_{52}H_{51}N_3O_2$	No hit
	7.5	430.9133	$C_4H_2N_2O_{22}$	No hit
		498.9008	$C_2H_2N_4O_{26}$	no hit
SM3.2 M400 Water Phase	4.7	387.1799	$C_{22}H_{26}O_6$	54/8322/8324
	6.8	403.3237	$C_{20}H_{34}O_8$	2849784
	7.6	384.1923	$C_{21}H_{25}N_3O_4$	36313
	8.5	391.2844	$C_{24}H_{38}O_4$	6950
SM3.2 MMM Water Phase	1.7	547.2975	$C_{24}H_{42}N_4O_{10}$	17462342
		836.4030	$C_{38}H_{57}N_7O_{14}$	6382
	1.9	814.4551	$C_{37}H_{63}N_7O_{13}$	no hit

	2.6	692.3500	$C_{33}H_{49}N_5O_{11}$	9357029
	4.7	387.1799	$C_{22}H_{26}O_6$	54
	8.1	465.2757	$C_{20}H_{49}N_7O_5$	no hit
	8.5	391.2844	$C_{24}H_{38}O_4$	6950
SM3.2 OM Water Phase	4.8	387.1805	$C_{23}H_{22}N_4O_2$	2248451
	6.8	403.3237	$C_{20}H_{34}O_8$	2849784
	7.6	384.1923	$C_{21}H_{25}N_3O_4$	36313
	8.1	465.2765	C29H32N6	no hit
			$C_{11}H_{29}N_{17}O_4$	no hit
	8.5	391.2844	$C_{24}H_{38}O_4$	6950
SM3.2 SGG Water Phase	4.7	387.1799	$C_{22}H_{26}O_6$	54/8322/8324
	6.8	403.3237	$C_{20}H_{34}O_8$	2849784
	7.6	384.1923	$C_{21}H_{25}N_3O_4$	36313
	8.5	391.2844	$C_{24}H_{38}O_4$	6950
SM3.2 SM Water Phase	8.1	465.2778	$C_{29}H_{32}N_6$	no hit
			$C_{33}H_{36}O_2$	no hit
	8.5	391.2844	$C_{24}H_{38}O_4$	6950

SM3.2 SYP Water Phase	4.8	387.1805	$C_{23}H_{22}N_4O_2$	2248451
	6.8	403.3237	$C_{20}H_{34}O_8$	2849784
	7.6	384.1923	$C_{21}H_{25}N_3O_4$	36313
	8.1	465.2765	$C_{29}H_{32}N_6$	no hit
			$C_{33}H_{36}O_2$	no hit
	8.5	391.2844	$C_{24}H_{38}O_4$	6950
	7.6	424.282	$C_{23}H_{39}NO_6$	no hit

Supplementary Table 3: MS data for SM3.2 aqueous phase crude extracts.

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest. SM3.1 samples represent the SM3 samples which were fermented at 28°**C**, while SM3.2 represent the SM3 samples which were fermented at 30°**C**. For the interpretation of the MS data a lot of different libraries were used, like Chemspider, Antibase, Dictionary of Natural Products, Reaxys etc. Some masses didn't give any hits when searched on these databases, which means that they may be new.

Sample	**Retention time (min)**	**$[M+H]^+$**	**Predicted Molecular formula**	**ChemSpider ID, AntiBase ID, PubChem and Reaxys ID**
SM3.2 M19 Organic Phase	6.3	579.2927	$C_{30}H_{38}N_6O_6$	no hit
SM3.2 M400 Organic Phase	0.8	582.3246	$C_{27}H_{40}N_{11}O_4$	no hit
	1.7	547.2977	$C_{24}H_{42}N_4O_{10}$	2794750

	2.6	530.2975	$C_{27}H_{39}N_5O_6$	65259724
	4.7	387.1807	$C_{22}H_{26}O_6$	54
	8.2	665.5816	$C_{40}H_{76}N_2O_5$	no hit
	8.5	391.2839	$C_{24}H_{38}O_4$	6950
SM3.2 MMM Organic Phase	0.5	409.1815	$C_{16}H_{28}N_2O_{10}$	4451032
	1.4	490.2765	$C_{22}H_{39}N_3O_9$	21387440
	1.5	261.1236	$C_{14}H_{16}N_2O_3$	179735
	4.7	387.1807	$C_{22}H_{26}O_6$	54
	8.5	391.2839	$C_{24}H_{38}O_4$	6950
SM3.2 OM Organic Phase	6.8	403.2331	$C_{20}H_{34}O_8$	11575
SM3.2 SGG Organic Phase	4.7	387.1807	$C_{22}H_{26}O_6$	54
	6.8	403.2331	$C_{20}H_{34}O_8$	11575
	8.5	391.2839	$C_{24}H_{38}O_4$	6950

SM3.2 SM Organic Phase	2.1	417.1185	$C_{21}H_{20}O_9$	4445119
	2.5	433.113	$C_{21}H_{20}O_{10}$	129878
SM3.2 SYP Organic Phase	2.8	821.4384	$C_{37}H_{64}N_4O_{16}$	no hit

Supplementary Table 4: MS data for SM3.2 organic phase crude extracts

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest. Chemspider, Antibase, Dictionary of Natural Products, Reaxys etc.

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
B226.1 M19 Organic Phase	1.3	395.1365	$C_{20}H_{18}N_2O_3$	no hit
	1.4	443.2143	$C_{19}H_{30}N_4O_8$	23169586
	1.5	261.1243	$C_{14}H_{16}N_2O_3$	179735
	1.6	521.2398	$C_{28}H_{32}N_4O_6$	899298 3510615
	2.0	284.1965	$C_{14}H_{24}N_3O_3$	no hit
	2.1	211.1444	$C_{11}H_{18}N_2O_2$	499509
	2.2	421.2813	$C_{22}H_{36}N_4O_4$	8223339
	2.3	459.2796	$C_{20}H_{42}O_{11}$	71989

	2.5	245.1288	$C_{14}H_{16}N_2O_2$	283250
	2.6	425.1315	$C_{18}H_{16}N_8O_5$	no hit
			$C_{17}H_{20}N_4O_9$	no hit
	3.0	301.2489	$C_{16}H_{32}N_2O_3$	no hit
	3.4	201.1481	$C_{11}H_{20}O_3$	30297
	3.5	391.1867	$C_{20}H_{26}N_2O_6$	574365
	3.7	410.2759	$C_{19}H_{39}NO_8$	no hit
	4.0	425.3115	$C_{21}H_{44}O_8$	2546847
	4.1	485.3445	$C_{22}H_{48}N2O_9$	no hit
	4.5	483.3529	$C_{18}H_{45}N9O_6$	no hit
	4.7	498.364	$C_{24}H_{53}NO_9$	no hit
	4.9	541.395	$C_{27}H_{56}O_{10}$	21989361
			$C_{28}H_{52}N_4O_6$	68006955
	5.0	387.1804	$C_{22}H_{26}O_6$	1937152
			$C_{23}H_{22}N_4O_6$	1314876
	5.2	556.4053	$C_{20}H_{54}N_6O_{11}$	no hit
	5.4	376.2595	$C_{21}H_{33}N_3O_3$	6338385 3519209
	5.8	657.4771	$C_{33}H_{68}O_{12}$	no hit
	6.2	737.5013	$C_{34}H_{68}N_6O_{11}$	no hit

	6.4	730.5307	$C_{34}H_{71}N_9O_8$	no hit
			$C_{33}H_{75}N_5O_{12}$	no hit
	6.6	811.5167	$C_{39}H_{74}N_2O_{15}$	no hit
	6.8	788.5724	$C_{35}H_{79}N_7O_{12}$	no hit
	7.0	528.3931	$C_{18}H_{53}N_7O_{10}$	no hit
	7.3	906.6712	$C_{46}H_{91}N_5O_{12}$	no hit
	7.7	491.3744	$C_{16}H_{42}N_{16}O_2$	no hit
	8.0	546.4882	$C_{35}H_{63}NO_3$	59702300 9988241
	8.5	673.4958	$C_{31}H_{68}N_4O_{11}$	no hit
			$C_{29}H_{10}N_2O_{18}$	no hit
	8.8	702.5191	$C_{25}H_{59}N_{21}O_3$	no hit
	9.0	421.178	$C_{25}H_{20}N_6O$	no hit
	9.1	734.0166	$C_{34}H_{11}N_3O_{17}$	no hit
	9.4	760.5591	$C_{27}H_{69}N_{17}O_8$	no hit
	9.7	789.5803	$C_{38}H_{76}N_8O_9$	no hit
	9.9/1.0	821.0798	$C_{30}H_{12}N_{16}O_{14}$	no hit
			$C_{31}H_8N_{20}O_{10}$	no hit
B226.1 M400	1.0	450.1985	$C_{20}H_{27}N_5O_7$	16702834 16718017
Organic	1.2	373.2092	$C_{16}H_{28}N_4O_6$	23466

Phase	1.3	712.2767	$C_{28}H_{45}N_3O_{18}$	170050
	1.5	261.1243	C14H16N2 O_3	179735
	1.7	521.2394	$C_{28}H_{32}N_4O_6$	2111100 2141713
	1.9	561.2386	$C_{22}H_{40}O_{16}$	4450543
	2.0	418.1729	$C_{19}H_{23}N_5O_6$	29402336
	2.1	211.1444	$C_{11}H_{18}N_2O_2$	499509
	2.2	421.2813	$C_{22}H_{36}N_4O_4$	8223339
	2.6	425.1315	$C_{18}H_{16}N_8O_5$	no hit
			$C_{17}H_{20}N_4O_9$	no hit
	2.8	398.1498	$C_{24}H_{19}N_3O_3$	42172
	3.0	301.2489	$C_{16}H_{32}N_2O_3$	no hit
	3.3	852.5100	$C_{45}H_{73}NO_{14}$	6273
	5.0	387.1804	$C_{23}H_{22}N_4O_6$	1314876
	8.7	788.6101	$C_{40}H_{85}NO_{13}$	no hit
	9.0	846.6503	$C_{40}H_{83}N_{11}O_8$	no hit
	9.4	904.6909	$C_{42}H_{93}N_7O_{13}$	no hit
	9.6	492.8637	$C_{34}H_{50}O_2$	34949611
	9.9	521.8831	$C_{28}H_{52}N_6O_3$	no hit
B226.1 MMM	1.6	261.1232	$C_{14}H_{16}N_2O_3$	179735

Organic Phase	1.5-1.7	521.2394	$C_{28}H_{32}N_4O_6$	2111100 2141713
	2.0	427.2555	$C_{20}H_{34}N_4O_6$	16654556
	2.3	245.1286	$C_{14}H_{16}N_2O_2$	90257
	2.5	245.1286	$C_{14}H_{16}N_2O_3$	90257
	2.8	398.1497	$C_{17}H_{18}N_8O_4$	35007911
	2.8	398.1498	$C_{24}H_{19}N_3O_3$	42172
	5.0	387.1803	$C_{22}H_{26}O_6$	1937152
	8.7	788.6101	$C_{41}H_{81}N_5O_9$	no hit
	9.0	421.1792	$C_{29}H_{24}O_3$	no hit
	9.4	904.6909	$C_{42}H_{93}N_7O_{13}$	no hit
	9.9-10.0	519.4052	$C_{25}H_{53}N_5O_6$	no hit
B226.1 OM **Organic Phase**	5.0	187.1801	$C_{22}H_{26}O_6$	1937152
	8.0	546.4882	$C_{35}H_{63}NO_3$	59702300
	9.0	421.1792	$C_{29}H_{24}O_3$	no hit
B226.1 SGG **Organic Phase**	1.5	261.1232	$C_{14}H_{16}N_2O_3$	179735
	1.7	521.2394	$C_{28}H_{32}N_4O_6$	2111100 2141713
	2.2	211. 1441	$C_{11}H_{18}N_2O_2$	499509
	2.6	447.1139	$C_{10}H_{20}N_6O_{14}$	no hit
	2.8	398.1498	$C_{24}H_{19}N_3O_3$	42172

	3.0	445.2334	$C_{24}H_{32}N_2O_6$	2269889 3289647
	3.7	440.1603	$C_{19}H_{20}N_8O_5$	32237821
	4.0	425.3115	$C_{21}H_{44}O_8$	2546847
			$C_{22}H_{40}N_4O_4$	9445222
	4.2	485.3445	$C_{22}H_{48}N_2O_9$	no hit
	4.5	483.353	$C_{24}H_{50}O_9$	no hit
			$C_{25}H_{46}N_4O_5$	no hit
	4.7	498.364	$C_{24}H_{53}NO_9$	no hit
	4.9	541.395	$C_{27}H_{56}O_{10}$	21989361
			$C_{28}H_{52}N_4O_6$	68006955
	5.0	558.4214	$C_{27}H_{59}NO_{10}$	no hit
	5.1	387.1804	$C_{22}H_{26}O_6$	1937152
	5.4	511.2805	$C_{17}H_{42}N_4O_{13}$	no hit
	5.6	614.4474	$C_{26}H_{61}N_7O_9$	no hit
	5.8	657.4771	$C_{33}H_{68}O_{12}$	no hit
	6.0	600.4318	$C_{29}H_{61}NO_{11}$	no hit
	6.2	737.5013	$C_{34}H_{68}N_6O_{11}$	no hit
	6.4	730.5304	$C_{31}H_{77}N_3O_{15}$	no hit
	6.6	811.5167	$C_{39}H_{74}N_2O_{15}$	no hit

	7.0	848.6299	$C_{42}H_{89}NO_{15}$	no hit
		1038.7503	$C_{47}H_{103}N_7O_{17}$	no hit
	7.3	906.6712	$C_{46}H_{91}N_5O_{12}$	no hit
	7.7	491.3744	$C_{24}H_{45}N_9O_2$	no hit
	8.0	546.4882	$C_{35}H_{63}NO_3$	59702300 9988241
	8.5	673.4958	$C_{31}H_{68}N_4O_{11}$	no hit
			$C_{29}H_{10}N_2O_{18}$	no hit
	8.7	788.6101	$C_{41}H_{81}N_5O_9$	no hit
	8.8	702.5191	$C_{24}H_{63}N_{17}O_{17}$	no hit
	8.9	421.178	$C_9H_{26}N_8O_{11}$	no hit
	9.1	846.653	$C_{43}H_{91}NO_{14}$	no hit
	9.4	904.6909	$C_{42}H_{93}N_7O_{13}$	no hit
	9.9-10.0	521.8831	$C_{28}H_{50}N_4O_5$	8612212
B226.1 SM **Organic Phase**	2.6	273.1707	$C_{14}H_{24}O_5$	9175388
	2.8	398.1498	$C_{24}H_{19}N_3O_3$	42172
	3.3	531.1056	$C_{28}H_{14}N_6O_6$	no hit
	3.4	285.0762	$C_{16}H_{12}O_5$	26775 20493
		569.1439	$C_{32}H_{24}O_{10}$	8501123
	3.7	440.1604	$C_{22}H_{23}N_3O_7$	2684152

	3.9	507.1076	$C_{31}H_{14}N_4O_4$	no hit
	4.1	537.1182	$C_{31}H_{20}O_9$	4479074
	4.3	523.1023	$C_{30}H_{18}O_9$	8003853
	4.5	583.1227	$C_{32}H_{22}O_{11}$	4477713
	4.7	523.1027	$C_{30}H_{18}O_9$	26371101
	4.8	507.1081	$C_{31}H_{14}N_4O_4$	no hit
	5.0	387.1804	$C_{22}H_{26}O_6$	1937152
	5.2	523.1023	$C_{30}H_{18}O_9$	8003853
	9.0	421.1792	$C_{29}H_{24}O_3$	no hit
B226.1 SYP **Organic Phase**	1.4	734.2583	$C_{38}H_{35}N_7O_9$	no hit
	1.6	261.1235	$C_{14}H_{16}N_2O_3$	179735
	1.9	256.2395	$C_{13}H_{33}N_{15}O_{10}$	no hit
			$C_{28}H_{37}N_3O_9$	34973344
	2.5	245.1288	$C_{14}H_{16}N_2O_2$	283250
	2.6	213.0697	$C_5H_{12}N_2O_7$	no hit
	2.8	398.1498	$C_{24}H_{19}N_3O_3$	42172
	3.3	852.5100	$C_{45}H_{73}NO_{14}$	6273
	5.0	387.1804	$C_{22}H_{26}O_6$	1937152
			$C_{23}H_{22}N_4O_6$	1314876

	8.0	546.4882	$C_{35}H_{63}NO_3$	59702300 9988241

Supplementary Table 5: MS data for B226SN104.1 organic phase crude extracts

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest. B226SN104.1 samples represent the B226SN104 samples which were fermented at 28°C, while B226SN104.2 represent the B226SN104 samples which were fermented at 30°C.

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
B226.1 M19 Water Phase	1.1	712.2759	$C_{28}H_{45}N_3O_{18}$	170050
	1.2	712.2762	$C_{26}H_{33}N_{17}O_8$	No hit
	1.4	445.2312	$C_{19}H_{32}N_4O_8$	16603920 16603958
	1.5	413.1421	$C_{15}H_{20}N_6O_8$	141829
	1.7	443.2509	$C_{18}H_{35}NO_{11}$	8798386
	1.8	394.1714	$C_{19}H_{25}N_3O_6$	3972975
	2.1	414.1251	$C_{18}H_{22}O_{11}$	8963229
	2.9	574.3604	$C_{30}H_{47}N_5O_6$	410355
	5.0	387.1802	$C_{22}H_{26}O_6$	54 8324 8322
	7.2	301.1408	$C_{14}H_{16}N_6O_2$	No hit

	7.7	403.2300	$C_{17}H_{27}N_{10}O_2$	No hit
	7.9	546.4878	$C_{35}H_{63}NO_3$	59702300 48063180 9988241
B226.1 M400 **Water Phase**	1.1	712.2759	$C_{28}H_{45}N_3O_{18}$	170050
	1.5	413.1409	$C_{15}H_{20}N_6O_8$	141829
	1.8	434.213	$C_{18}H_{31}N_3O_9$	25167892
	1.9	394.1735	$C_{19}H_{32}N_4O_5$	16653742
	2.1	425.1254	$C_{20}H_{30}N_8O_2$	58248827
	2.3	457.229	$C_{20}H_{32}N_4O_8$	23219440
	4.1	471.3178	$C_{23}H_{42}N_4O_6$	28586695 57535380
	4.8	655.2753	$C_{35}H_{39}N_5O_8$	17252939
	5.0	387.1802	$C_{22}H_{26}O_6$	54
	5.5	1107.0504	$C_{29}H_{78}N_{20}O_{25}$	No hit
	7.1	579.2924	$C_{30}H_{38}N_6O_6$	166902
	7.2	301.1408	$C_{14}H_{16}N_6O_2$	No hit
	7.7	403.2333	$c_{20}h_{34}0_8$	11575 20091
	7.9	546.4878	$C_{35}H_{63}NO_3$	59702300 48063180 9988241
B226.1 MMM	1.1	712.2762	$C_{26}H_{33}N_{17}O_8$	No hit

Water Phase	1.2	505.2144	$C_{20}H_{32}N_4O_{11}$	16617301
	1.4	465.1979	$C_{21}H_{28}N_4O_8$	16722589 16720894 16721831
	1.5	500.2358	$C_{21}H_{33}N_5O_9$	61710562
	1.8	394.1714	$C_{19}H_{25}N_3O_6$	3972975
	1.9	578.2655	$C_{20}H_{31}N_{15}O_6$	No hit
	2.0	427.2553	$C_{20}H_{34}N_4O_6$	16693511
	2.2	402.2026	$C_{21}H_{27}N_3O_5$	1198691
	2.3	457.229	$C_{20}H_{32}N_4O_8$	23219440
	5.0	387.1802	$C_{22}H2_6O_6$	54 8324 8322
	7.2	301.1408	$C_{14}H_{16}N_6O_2$	No hit
	7.8	403.2333	$C_{21}H_{30}N_4O_4$	No hit
	7.9	546.4878	$C_{35}H_{63}NO_3$	59702300 48063180 9988241
B226.1 OM **Water Phase**	1.1	712.2760	$C_{24}H_{41}N_9O_{16}$	28573801
	1.2	412.1836	$C_{17}H_{25}N_5O_7$	8607171
	1.3	394.1717	$C_{16}H_{27}NO_{10}$	26585174
	1.5	413.1403	$C_{14}H_{24}N_2O_{12}$	No hit
	3.8	1242.6119	$C_{59}H_{87}N_9O_{20}$	No hit
	3.9	1063.5326	$C_{52}H_{78}N_4O_{19}$	No hit

	4.0	901.4784	$C_{42}H_{64}N_{10}O_{12}$	No hit
			$C_{45}H_{72}O_{18}$	354002
	5.0	387.1802	$C_{22}H_{26}O_6$	54 8324 8322
	7.2	301.1408	$C_{14}H_{16}N_6O_2$	No hit
	7.7	403.2333	$C_{20}H_{34}O_8$	1157 20091 42039
	7.9	546.4878	$C_{35}H_{63}NO_3$	59702300 48063180 9988241
B226.1 SGG **Water Phase**	1.1	712.2759	$C_{28}H_{45}N_3O_{18}$	170050
	1.1-1.2	712.2745	$C_{24}H_{41}N_9O_{16}$	28573801
	1.5	413.1403	$C_{14}H_{24}N_2O_{12}$	no hit
	1.8	434.213	$C_{18}H_{31}N_3O_9$	25167892
	1.9	394.1437	$C_{12}H_{25}N_7O_8$	no hit
	2.1	442.2699	$C_{25}H_{35}N_3O_4$	8496558
	4.1	471.3178	$C_{23}H_{42}N_4O_6$	28586695 57535380
	5.0	387.1802	$C_{22}H_{26}O_6$	54 8324
	6.5	507.2288	$C_{32}H_{30}N_2O_4$	1293 1800
	7.2	301.1408	$C_{14}H_{16}N_6O_2$	no hit
	7.8	403.2333	$C_{21}H_{30}N_4O_4$	no hit

B226.1 SM Water Phase	1.1	712.2759	$C_{28}H_{45}N_3O_{18}$	170050
	1.4	445.2304	$C_{19}H_{32}N_4O_8$	16603920 16603958
	1.5	413.1419	$C_8H_{19}N_{11}O_9$	no hit
	1.8	434.213	$C_{18}H_{31}N_3O_9$	25167892
	5.0	387.1802	$C_{22}H_{26}O_6$	54 8324 8322
	7.2	301.1408	$C_{14}H_{16}N_6O_2$	no hit
B226.1 SYP Water Phase	1.1	712.2759	$C_{28}H_{45}N_3O_{18}$	170050
	1.5	413.1421	$C_{15}H_{20}N_6O_8$	141829
	2.3	461.2769	$C_{24}H_{36}N_4O_5$	2598026
	5.0	387.1802	$C_{22}H_{26}O_6$	54 8324 8322
	7.2	301.1408	$C_{14}H_{16}N_6O_2$	no hit

Supplementary Table 6: MS data for B226SN104.1 aqueous phase crude extracts

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest.

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
B226.2 M19 Organic Phase	0.5	429.1572	$C_{12}H_{24}N_6O_{11}$	no hit
	1.3	712.2769	$C_{28}H_{45}N_3O_{18}$	196239
	5.0	387.1811	$C_{23}H_{22}N_4O_2$	12227
	7.1	317.1154	$C_{17}H_{12}N_6O$	no hit
		579.2922	$C_{30}H_{38}N_6O_6$	no hit
	7.8	403.2328	$C_{20}H_{34}O_8$	8222245 2849784
	9.2	384.8468	$C_{23}H_{45}NO_3$	41519
	9.6	465.2799	$C_{16}H_{34}N_8O_8$	no hit
	10.0	391.2855	$C_{24}H_{38}O_4$	no hit
B226.2 M400 Organic Phase	0.5	404.1926	$C_{19}H_{25}N_5O_5$	3215137
	0.8	533.1472	$C_{26}H_{34}N_2O_{10}$	9845760
	1.1	712.2769	$C_{28}H_{45}N_3O_{18}$	196239
	1.2	412.1833	$C_{17}H_{25}N_5O_7$	8607171
	1.3	712.2757	$C_{25}H_{37}N_{13}O_{12}$	no hit
	1.6	413.1416	$C_{15}H_{20}N_6O_8$	19989181
	1.7	625.3131	$C_{22}H_{48}N_4O_{16}$	no hit
			$C_{27}H_{43}N_7O_{10}$	no hit

	1.8	403.178	$C_{16}H_{21}N_9O_4$	34512811
	2.1	211.1444	$C_{11}H_{18}N_2O_2$	2298456
	2.4	245.1288	$C_{14}H_{16}N_2O_2$	422943
	2.5	235.0516	$C_{11}H_9NO_5$	31422
	2.7	398.1502	$C_{24}H_{19}N_3O_3$	42172
	2.9	301.2482	$C_{16}H_{32}N_2O_3$	no hit
	3.3	852.5106	$C_{45}H_{73}NO_{14}$	no hit
	5.0	387.1811	$C_{23}H_{22}N_4O_2$	12227
	7.1	579.2922	$C_{30}H_{38}N_6O_6$	no hit
	7.8	403.2328	$C_{20}H_{34}O_8$	8222245 2849784
	10.0	391.2855	$C_{24}H_{38}O_4$	no hit
B226.2 MMM Organic Phase	0.5	404.1926	$C_{19}H_{25}N_5O_5$	3215137
	1.2	712.2769	$C_{28}H_{45}N_3O_{18}$	196239
	1.3	712.2757	$C_{25}H_{37}N_{13}O_{12}$	no hit
	1.5	261.1237	$C_{14}H_{16}N_2O_3$	179735
	1.8	547.298	$C_{24}H_{42}N_4O_{10}$	2794750
	1.9	578.2652	$C_{22}H_{43}NO_{16}$	no hit
	2.0	211.1443	$C_{11}H_{18}N_2O_2$	2298456

	2.1	427.2543	$C_{20}H_{34}N_4O_6$	16654556 35143235
	2.4	245.1288	$C_{14}H_{16}N_2O_2$	422943
	2.7	398.1496	$C_{24}H_{19}N_3O_3$	42172
	5.0	387.1801	$C_{22}H_{26}O_6$	54 8324 8322
	7.1	317.1154	$C_{17}H_{12}N_6O$	no hit
	7.8	403.2317	$C_{17}H_{26}N_{10}O_3$	no hit
	10.0	391.2855	$C_{24}H_{38}O_4$	no hit
B226.2 OM **Organic Phase**	0.4	505.1775	$C_{18}H_{32}O_{16}$	168531
	1.2	712.2769	$C_{28}H_{45}N_3O_{18}$	196239
	1.4	445.2291	$C_{19}H_{32}N_4O_8$	16690125 16621760 16623896
	1.5	413.1416	$C_{15}H_{20}N_6O_8$	19989181
	2.7	398.1502	$C_{24}H_{19}N_3O_3$	42172
	4.2	739.4253	$C_{36}H_{54}N_{10}O_7$	57535864
	4.5	787.4269	$C_{43}H_{62}O_{13}$	23249159
	4.9	739.4265	$C_{39}H_{62}O_{13}$	9472139
	5.0	387.1801	$C_{22}H_{26}O_6$	54 8324 8322

	5.5	387.1946	$C_{22}H_{26}O_3$	no hit
	5.6	781.4366	$C_{39}H_{52}N_{14}O_4$	no hit
	5.7	770.4459	$C_{43}H_{63}NO_{11}$	no hit
	6.2	723.4309	$C_{31}H_{65}NO_{17}$	no hit
			$C_{29}H_{53}N_{15}O_7$	9518027
	7.1	317.1154	$C_{17}H_{12}N_6O$	no hit
	7.7	407.3356	$C_{21}H_{43}N_7O$	no hit
	7.8	403.2317	$C_{17}H_{26}N_{10}O_3$	no hit
	7.9	546.4877	$C_{35}H_{63}NO_3$	9988241
	8.9	654.3305	$C_{28}H_{41}N_{13}O_6$	no hit
	10.0	391.2855	$C_{24}H_{38}O_4$	no hit
B226.2 SGG Organic Phase	0.8	533.1463	$C_{17}H_{28}N_2O_7$	no hit
	1.2	412.1833	$C_{17}H_{25}N_5O_7$	8607171
	1.5	261.1237	$C_{14}H_{16}N_2O_3$	179735
	1.8	403.1782	$C_{11}H_{26}N_6O_{10}$	no hit
	2.1	211.1444	$C_{11}H_{18}N_2O_2$	2298456
	2.4	611.7316	$C_{41}H_{92}N_2$	no hit
	2.4-2.5	213.0708	$C_9H_{11}NO_5$	9260283

	2.7	398.1494	$C_{24}H_{19}N_3O_3$	42172
	2.9	445.2531	$C_{24}H_{32}N_2O_6$	2269889 3289647
	3.3	852.5099	$C_{26}H_{76}N_8O_{22}$	no hit
	4.3	414.3362	$C_{27}H_{43}NO_2$	161334 10181233 10181225
	5.0	387.1801	$C_{22}H_{26}O_6$	54 8324 8322
	7.1	317.1154	$C_{17}H_{12}N_6O$	no hit
	7.6	403.2328	$C_{20}H_{34}O_8$	8222245 2849784
	7.9	546.4877	$C_{35}H_{63}NO_3$	9988241
	10.0	391.2855	$C_{24}H_{38}O_4$	no hit
B226.2 SM **Organic Phase**	1.1	712.2793	$C_{29}H_{41}N_7O_{14}$	17893700
	1.3	712.2769	$C_{28}H_{45}N_3O_{18}$	196239
	2.7	398.1502	$C_{24}H_{19}N_3O_3$	42172
	3.1	1269.6127	$C_{60}H_{92}N_4O_{25}$	no hit
			$C_{56}H_{88}N_{10}O_{23}$	no hit
	3.2	531.1054	$C_{28}H_{14}N_6O_6$	no hit
	3.3	591.1256	$C_{30}H_{18}N_6O_8$	no hit
	3.9	507.1076	$C_{30}H_{18}O_8$	9852539

	4.3	523.1024	$C_{30}H_{18}O_9$	4572736
	4.6	523.1029	$C_{24}H_{13}N_9O_6$	no hit
	4.8	545.0612	$C_{30}H_{12}N_2O_9$	no hit
	4.9	943.5257	$C_{49}H_{74}N_4O_{14}$	9717317
	5.0	387.1801	$C_{22}H_{26}O_6$	54 8324 8322
	5.2	523.1033	$C_{30}H_{18}O_8$	9791
	5.6	539.0965	$C_{27}H_{10}N_{10}O_4$	no hit
	6.2	393.2614	$C_{19}H_{32}N_6O_3$	no hit
	7.1	317.1154	$C_{17}H_{12}N_6O$	no hit
	7.7	407.3354	$C_{23}H_{42}N_4O_2$	34404126
	8.7	413.3239	$C_{20}H_{37}N_5O_3$	64438363 64689869
	10.0	391.2855	$C_{24}H_{38}O_4$	no hit
B226.2 SYP **Organic Phase**	0.6	437.1767	$C_{17}H_{28}N_2O_{11}$	9628214
	0.8	188.0923	$C_8H_{13}NO_4$	16701
	1.2	712.2769	$C_{28}H_{45}N_3O_{18}$	196239
	1.3	712.2757	$C_{25}H_{37}N_{13}O_{12}$	no hit
	1.5	261.1237	$C_{14}H_{16}N_2O_3$	179735

	2.0	211.1443	$C_{11}H_{18}N_2O_2$	2298456
	2.1	211.1441	$C_7H_{11}N_7O$	54947356
	2.4	217.0974	$C_{12}H_{12}N_2O_2$	206710
	2.5	213.0697	$C_{12}H_{12}N_2O_3$	no hit
			$C_5H_{12}N_2O_7$	no hit
	2.7	398.1502	$C_{24}H_{19}N_3O_3$	42172
	3.0	623.4707	$C_{31}H_{60}N_8O_5$	no hit
	3.3	852.5106	$C_{45}H_{73}NO_{14}$	no hit
	4.3	414.3362	$C_{27}H_{43}NO_2$	161334 10181233 10181225
	4.6	419.1709	$C_{22}H_{26}O_8$	8326
	5.0	387.1801	$C_{22}H_{26}O_6$	54 8324 8322
	7.1	317.1154	$C_{17}H_{12}N_6O$	no hit
	7.8	403.2328	$C_{20}H_{34}O_8$	8222245 2849784
	7.9	546.4877	$C_{35}H_{63}NO_3$	9988241
	10.0	391.2855	$C_{24}H_{38}O_4$	no hit

Supplementary Table 7: MS data for B226SN104.2 organic phase crude extracts

B226.1 samples represent the B226SN104 samples which were fermented at 28°**C**, while B226.2 represent the B226SN104 samples which were fermented at 30°**C**.

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
B226.2 M19 Water Phase	0.6	294.1202	$C_{12}H_{16}N_5O_4$	26929
	0.7	276.1101	$C_{12}H_{13}N_5O_3$	7986
			$C_{11}H_{17}NO_7$	No hit
		441.1895	$C_{25}H_{28}O_7$	35800
	1.0	205.0984	$C_{11}H_{12}N_2O_2$	24808
				28079
	1.3	712.2738	$C_{24}H_{41}N_9O_{16}$	No hit
	7.0	279.1581	$C_{16}H_{22}O_4$	7589
	7.8	Common impurities		
	9.2	Common impurities		
B226.2 M400 Water Phase	0.5	387.0693	$C_{15}H_{10}N_6O_7$	No hit
	0.6	415.1817	$C_{16}H_{26}N_6O_7$	No hit
	1.0	450.1972	$C_{19}H_{31}NO_{11}$	No hit
	1.2	Common impurities		
	1.6	413.1408	$C_{15}H_{20}N_6O_8$	21113
	1.7	625.3141	$C_{23}H_{44}N_8O_{12}$	No hit
	2.9	574.3594	$C_{29}H_{51}NO_{10}$	No hit
	4.9	655.277	$C_{36}H_{38}N_4O_8$	24134

	7.0	279.1581	$C_{16}H_{22}O_4$	7589
	7.8	Common impurities		
	9.2	Common impurities		
B226.2 MMM Water Phase	0.6	437.179	$C_{18}H_{24}N_6O_7$	No hit
	1.3	Common impurities		
	1.6	413.1438	$C_{19}H_{24}O_{10}$	No hit
	1.8	588.2552	$C_{29}H_{37}N_3O_{10}$	No hit
	2.0	Common impurities		
	2.1	427.2577	$C_{25}H_{34}N_2O_4$	27796
	7.0	279.1581	$C_{16}H_{22}O_4$	7589
	9.2	Common impurities		
B226.2 OM Water Phase	1.3	Common impurities		
	5.5	413.2662	$C_{21}H_{36}N_2O_6$	14209
	7.0	279.1581	$C_{16}H_{22}O_4$	7589
	9.2	Common impurities		
B226.2 SGG Water	1.2	Common impurities		
	7.0	279.1581	$C_{16}H_{22}O_4$	7589

Phase	9.2	Common impurities		
B226.2 SM Water Phase	1.9	Common impurities		
	2.1	415.125	$C_{19}H_{18}N_4O_7$	24565
	7.0	279.1581	$C_{16}H_{22}O_4$	7589
	9.2	Common impurities		
B226.2 SYP Water Phase	0.6	415.1820	$C_{25}H_{26}N_4O_2$	No hit
	1.2	Common impurities		
	1.6	413.1408	$C_{15}H_{20}N_6O_8$	21113
	7.0	279.1581	$C_{16}H_{22}O_4$	7589
	9.2	Common impurities		

Supplementary Table 8: MS data for B226SN104.2 aqueous phase crude extracts

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest.

Sample	Retention time (min)	[M+H]+	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
SM9 M19 Water Phase	0.8	585.3200	$C_{21}H_{44}N_8O_{11}$	no hit
	0.9	585.3195	$C_{20}H_{48}N_4O_{15}$	no hit
	1.6	455.2140	$C_{20}H_{30}N_4O8$	116946
	1.7	413.2032	$C_{18}H_{28}N_4O_7$	16616289
		545.3658	$C_{25}H_{48}N_6O_7$	26389401
	1.9	639.3280	$C_{18}H_{30}N_{28}$	no hit
	2.1	461.2603	$C_{21}H_{32}N_8O_4$	8044569
			$C_{20}H_{36}N_4O_8$	16619974
	2.3	443.2502	$C_{20}H_{34}N_4O_7$	16621964
				16690128
	5.0	387.1813	$C_{22}H_{26}O_6$	319252
				2559952
			$C_{23}H_{22}N_4O_2$	1314876
				1315399
	7.0	579.2937	$C_{27}H_{41}N_5O_9$	8161241
	7.8	425.2152	$C_{19}H_{24}N_{10}O_2$	2544392
SM9 M400 Water Phase	0.8	585.3188	$C_{20}H_{48}N_4O_{15}$	no hit
	1.6	455.2134	$C_{21}H_{26}N_8O_4$	27443598
	1.7	518.2134	$C_{23}H_{43}N_5O_8$	(PubChem)
	1.9	401.2104	$C_{18}H_{32}N_4O_6$	16694122
	2	461.2603	$C_{20}H_{36}N_4O_8$	16619974
	2.3	443.2502	$C_{20}H_{34}N_4O_7$	16621964
				16690128
	2.9	557.3647	$C_{27}H_{44}N_{10}O_3$	no hit
	5	387.1813	$C_{22}H_{26}O_6$	319252
			$C_{23}H_{22}N_4O_2$	1314876

Sample	Retention time (min)	[M+H]+	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
	7	579293	$C_{10}H_{25}N_{31}$	no hit
	7.8	425.2145	$C_{18}H_{28}N_6O_6$	57486676
SM9 MMM Water Phase	1	585.3191	$C_{21}H_{44}N_8O_{11}$	no hit
	1.6	455.2140	$C_{20}H_{30}N_4O_8$	116946
	1.8	723.4129	$C_{30}H_{62}N_2O_{17}$	no hit
	1.9	502.2399	$C_{22}H_{35}N_3O_{10}$	26574096
			$C_{23}H_{31}N_7O_6$	16642294
	2.0	461.2603	$C_{20}H_{36}N_4O_8$	16619974
	2.3	443.2505	$C_{20}H_{34}N_4O_7$	16729951
	5	387.1813	$C_{22}H_{26}O_6$	319252
				2559952
			$C_{23}H_{22}N_4O_2$	1314876
				1315399
	7	579.2943	$C_{31}H_{34}N_{10}O_2$	no hit
	7.8	425.2145	$C_{18}H_{28}N_6O_6$	57486676
SM9 OM Water Phase	1.4	486.3043	$C_{22}H_{35}N_{11}O_2$	58840443
	1.5	413.1423	$C_{15}H_{20}N_6O_8$	141829
	1.6	455.2144	$C_{14}H_{25}N_{13}O_5$	no hit
	2.0	576.2297	$C_{24}H_{21}N_{19}$	no hit
	2.4	443.2426	$C_{34}H_{55}N_9O_{10}$	13758052
	2.9	732.1032	$C_{33}H_{57}N_5O_{13}$	No hit
	4.0	901.4807	$C_{51}H_{68}N_2O_{12}$	3264848
	4.8	912.6273	$C_{48}H_{81}N_9O_8$	30770945
			$C_{47}H_{85}N_5O_{12}$	no hit

Sample	Retention time (min)	[M+H]+	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
	5.0	387.1813	$C_{22}H_{26}O_6$	319252
			$C_{23}H_{22}N_4O_2$	1314876
	5.6	1225.5382	$C_{58}H_{84}N_2O_{26}$	no hit
			$C_{56}H_{72}N_{16}O_{16}$	no hit
			$C_{59}H_{80}N_6O_{22}$	no hit
	5.9	1038.6309	$C_{47}H_{83}N_{13}O_{13}$	no hit
			$C_{50}H_{91}N_3O_{19}$	no hit
			$C_{51}H_{87}N_7O_{15}$	no hit
	7	579.2943	$C_{31}H_{34}N_{10}O_2$	no hit
	7.8	425.2145	$C_{18}H_{28}N_6O_6$	57486676
SM9 SGG Water Phase	1	585.3191	$C_{21}H_{44}N_8O_{11}$	no hit
	1.7	517.3346	$C_{23}H_{44}N_6O_7$	8133795
	1.8	561.3594	$C_{37}H_{44}N_4O$	2356339
	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619974
	2.3	587.3387	$C_{25}H_{50}N_2O_{13}$	no hit
	2.5	601.3548	$C_{27}H_{48}N_6O_9$	141883
	5	387.1813	$C_{22}H_{26}O_6$	319252
				2559952
			$C_{23}H_{22}N_4O_2$	1314876
				1315399
	7	579.2943	$C_{31}H_{34}N_{10}O_2$	no hit
	7.8	425.2152	$C_{19}H_{24}N_{10}O_2$	2544392
SM9	1.4	486.3043	$C_{22}H_{35}N_{11}O_2$	58840443
	1.8	561.3611	$C_{25}H_{48}N_6O_8$	2867

Sample	Retention time (min)	[M+H]+	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
SM **Water Phase**	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619974
	2.4	443.2502	$C_{20}H_{34}N_4O_7$	16621964
				16690128
	2.5	601.3548	$C_{26}H_{52}N_2O_{13}$	no hit
			$C_{27}H_{48}N_6O_9$	141883
	3	1269.608	$C_{55}H_{92}N_6O_{27}$	no hit
			$C_{56}H_{88}N_{10}O_{23}$	no hit
	3.2	1107.5614	$C_{34}H_{69}N_{29}O_{14}$	no hit
	3.4	1077.5479	$C_{49}H_{76}N_{10}O_{17}$	no hit
	3.8	959.5205	$C_{48}H_{78}O_{19}$	112 possible hit
			$C_{45}H_{70}N_{10}O_{13}$	13166493
			$C_{49}H_{74}N_4O_{15}$	no hit
	4.8	959.5193	$C_{47}H_{62}N_{18}O_5$	no hit
	5	387.1813	$C_{22}H_{26}O_6$	319252
				2559952
			$C_{23}H_{22}N_4O_2$	1314876
				1315399
	5.6	507.2345	$C_{26}H_{30}N_6O_5$	29405263
	6.4	535.2652	$C_{21}H_{39}N_5O_{11}$	no hit
	6.8	535.2652	$C_{28}H_{40}N_2O_9$	21171714
			$C_{29}H_{36}N_6O_5$	8362966
	7	579293	$C_{30}H_{38}N_6O_6$	no hit
	7.4	708.5399	$C_{37}H_{69}N_7O_6$	no hit
	7.8	425.2152	$C_{19}H_{24}N_{10}O_2$	2544392
	9.2	405.2403	$C_{17}H_{46}N_2O_8$	no hit

Sample	Retention time (min)	[M+H]+	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
SM9 SYP Water Phase	0.9	585.3195	$C_{21}H_{44}N_8O_{11}$	no hit
	1.6	455.2140	$C_{20}H_{30}N_4O_8$	116946
	2.4	750.4124	$C_{33}H_{59}N_5O_{14}$	no hit
			$C_{34}H_{55}N_9O_{10}$	8007677
			$C_{23}H_{37}N_5O_7$	16703145
				16725901
	2.9	732.4031	$C_{33}H_{57}N_5O_{13}$	no hit
	3.3	716.4078	$C_{42}H_{56}N_2O_8$	34531180
			$C_{37}H_{56}N_4O_{10}$	29214737
	4.2	946.5797	$C_{54}H_{79}N_3O_{11}$	2551896
	4.8	898.615	$C_{51}H_{83}N_3O_{10}$	no hit
	4.9	912.6273	$C_{45}H_{73}N_{19}O_2$	no hit
			$C_{48}H_{81}N_9O_8$	30770945
	5	387.1813	$C_{22}H_{26}O_6$	319252
				2559952
			$C_{23}H_{22}N_4O_2$	1314876
				1315399
	5.6	1225.5382	$C_{58}H_{84}N_2O_{26}$	no hit
			$C_{56}H_{72}N_{16}O_{16}$	no hit
			$C_{59}H_{80}N_6O_{22}$	no hit
	5.9	1038.6309	$C_{47}H_{83}N_{13}O_{13}$	no hit
	7	579.293	$C_{30}H_{38}N_6O_6$	no hit
	7.8	425.2145	$C_{18}H_{28}N_6O_6$	57486676

Supplementary Table 9: MS data for SM9 aqueous phase crude extracts

SM9 crude extract samples represent the samples which were fermented at 28°**C**, in the seven media above. Underlined= compounds only present to this extract, represent potential unique compound.

Sample	Retention time (min)	$[M+H]^+$	Predicted Molecular formula	ChemSpider ID, AntiBase ID, PubChem and Reaxys ID
SM9 M19 Organic Phase	1.0	585.3186	$C_{21}H_{44}N_8O_{11}$	no hit
	1.4	419.2500	$C_{18}H_{34}N_4O_7$	17588243
	1.5	261.1238	$C_{14}H_{16}N_2O_3$	106647
	1.7	455.2139	$C_{20}H_{30}N_4O_8$	62881823
	1.8	517.3352	$C_{23}H_{44}N_6O_7$	8133795
			$C_{22}H_{48}N_2O_{11}$	no hit
	1.9	561.3603	$C_{24}H_{52}N_2O_{12}$	no hit
	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619887
	2.3	399.2235	$C_{18}H_{30}N_4O_6$	16733937
	2.4	443.2496	$C_{20}H_{34}N_4O_7$	16690128
	2.6	457.2660	$C_{21}H36_6N_4O_7$	16621764
	2.9	493.264	$C_{19}H_{36}N_6O_9$	no hit
	3.2	303.1708	$C_{26}H_{38}N_4O_6$	no hit
	3.4	503.2844	$C_{26}H_{38}N_4O_6$	7634
	3.5	519.2814	$C_{26}H_{38}N_4O_7$	4979463
				4979406
				4979449
	4.0	454.3032	$C_{22}H_{39}N_5O_5$	21375778
	4.9	912.6273	$C_{47}H_{85}N_5O_{12}$	no hit
	5.0	387.1803	$C_{22}H_{26}O_6$	54
			$C_{23}H_{22}N_4O_2$	12227
	6.2	954.6373	$C_{50}H_{83}N_9O_9$	no hit
	6.4	535.2656	$C_{27}H_{38}N_2O_9$	7982903
	7.8	403.2334	$C_{20}H_{34}O_8$	8891446

SM9 M400 Oraganic Phase	1	585.3186	$C_{21}H_{44}N_8O_{11}$	no hit
	1.4	419.25	$C_{18}H_{34}N_4O_7$	17588243
		467.2084	$C_{16}H_{30}N_6O_{10}$	no hit
			$C_{15}H_{34}N_2O_{14}$	no hit
	1.5	261.1238	$C_{14}H_{16}N_2O_3$	106647
	1.7	455.2139	$C_{19}H_{34}O_{12}$	8679289
		518.3179	$C_{23}H_{43}N_5O_8$	no hit
	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619887
	2.3	399.2235	$C_{18}H_{30}N_4O_6$	16733937
	2.4	443.2496	$C_{20}H_{34}N_4O_7$	16690128
	2.6	479.2477	$C_{19}H_{30}N_{10}O_5$	no hit
			$C_{22}H_{38}O_{11}$	9039028
		599.3402	$C_{27}H_{46}N_6O_9$	59052244
	2.8	472.1714	$C_{23}H_{25}N_3O_8$	11777969
	2.9	732.4026	$C_{31}H_{45}N_{19}O_3$	no hit
		493.264	$C_{19}H_{36}N_6O_9$	no hit
	3.4	503.2859	$C_{16}H_{40}N_8O_{10}$	no hit
	3.5	519.2814	$C_{23}H_{41}N_3O_{10}$	21387441
	4.0	454.3032	$C_{22}H_{39}N_5O_5$	21375778
	4.9	912.6259	$C_{46}H_{89}NO_{16}$	no hit
	5.0	387.1811	$C_{23}H_{22}N_4O_2$	12227
	5.3	376.259	$C_{21}H_{33}N_3O_3$	11199111
	6.1	954.6384	$C_{49}H_{87}N_5O_{13}$	no hit
	7.8	403.2334	$C_{20}H_{34}O_8$	8891446
SM9 MMM	1.0	585.3186	$C_{21}H_{44}N_8O_{11}$	no hit
	1.5	261.1238	$C_{14}H_{16}N_2O_3$	179735
				106647

Organic Phase	1.7	455.2139	$C_{20}H_{30}N_4O_8$	62881823
	1.8	547.2960	$C_{21}H_{34}N_{14}O_4$	no hit
	2.0	397.2083	$C_{18}H_{28}N_4O_6$	15460220
	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619887
	2.2	397.2088	$C_{18}H_{28}N_4O_6$	16694007
	2.4	443.2496	$C_{20}H_{34}N_4O_7$	16690128
	2.9	732.4026	$C_{31}H_{45}N_{19}O_3$	no hit
	4.9	912.6273	$C_{47}H_{85}N_5O_{12}$	no hit
	5.0	387.1803	$C_{22}H_{26}O_6$	54
	6.2	954.6405	$C_{49}H_{87}N_5O_{13}$	no hit
SM9 SGG Organic Phase	1.0	585.3186	$C_{21}H_{44}N_8O_{11}$	no hit
	1.4	419.2500	$C_{18}H_{34}N_4O_7$	17588243
	1.8	517.3352	$C_{23}H44N6O7$	8133795
			$C_{22}H_{48}N_2O_{11}$	no hit
	1.8	561.3610	$C_{25}H_{48}N_6O_8$	2867
	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619887
	2.2	573.3200	$C_{32}H_{42}N_6O_4$	10592905
	2.3	587.3386	$C_{22}H_{44}N_{12}O_7$	no hit
	2.4	443.2496	$C_{20}H_{34}N_4O_7$	16690128
	2.5	601.3546	$C_{27}H_{48}N_6O_9$	141883
	2.9	445.2348	$C_{24}H_{32}N_2O_6$	2269889
				3289647
	3.4	503.2844	$C_{26}H_{38}N_4O_6$	16529241
				16529256
				16529271
	4.9	912.6273	$C_{48}H_{81}N_9O_8$	30770945
	5.0	387.1811	$C_{23}H_{22}N_4O_2$	12227

SM9 OM Organic Phase	1.5	223.0942	$C_8H_{10}N_6O_2$	21374769
	2.0	462.2916	$C_{19}H_{43}NO_{11}$	no hit
	2.6	265.1408	$C_{11}H_{16}N_6O_2$	336159
				22153576
				22348155
	2.7	243.1592	$C_{13}H_{23}O_4$	no hit
	2.9	241.1441	$C_{13}H_{21}O_4$	no hit
	3.1	265.1410	$C_{11}H_{14}N_6O_2$	1078806
	3.3	213.1494	$C_{12}H_{20}O_3$	Reaxys: 65 substances possible
	3.9	453.3205	$C_{26}H_{44}O_6$	34241116
	4.1	247.1302	$C_{11}H_{14}N_6O$	no hit
	5.3	376.2590	$C_{21}H_{33}N_3O_3$	11199111
	5.6	696.3436	$C_{31}H_{53}NO_{16}$	no hit
		717.2733	$C_{32}H_{40}N_6O_{13}$	no hit
	6.1	954.6371	$C_{49}H_{87}N_5O_{13}$	no hit
	6.6	790.5885	Impurity	
	7.6	780.4001	$C_{33}H_{53}N_{11}O_{11}$	17351786
	7.8	403.2329	$C_{20}H_{34}O_8$	9048392
SM9 SM Organic Phase	1.0	585.3186	$C_{21}H_{44}N_8O_{11}$	no hit
	1.3	712.277	**$C_{28}H_{45}N_3O_{18}$**	170050
	1.7	518.3179	$C_{23}H_{43}N_5O_8$	no hit
	1.8	561.3619	$C_{25}H_{48}N_6O_8$	2867
	2.0	401.2391	$C_{18}H_{32}N_4O_6$	16694122
	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619887
	2.3	399.2235	$C_{18}H_{30}N_4O_6$	16733937
	2.4	443.2496	$C_{20}H_{34}N_4O_7$	16690128

	2.5	601.3546	$C_{27}H_{48}N_6O_9$	141883
	2.6	599.3402	$C_{27}H_{46}N_6O_9$	59052244
	2.8	472.1714	$C_{23}H_{25}N_3O_8$	11777969
	3.2	255.0652	$C_{15}H_{10}O_4$	Reaxys : 60 substances possible
	3.7	271.0601	$C_{15}H_{10}O_5$	29831
	5.0	387.1803	$C_{22}H_{26}O_6$	54
	5.3	376.259	$C_{21}H_{33}N_3O_3$	11199111
	5.6	507.2353	$C_{25}H_{34}N_2O_9$	8566920
	6.4	535.2664	$C_{28}H_{34}N_6O_5$	8362445
			$C_{27}H_{38}N_2O_9$	7982903
	6.8	549.282	$C_{28}H_{40}N_2O_9$	21171714
SM9 SYP Organic Phase	1.3	712.277	**$C_{28}H_{45}N_3O_{18}$**	170050
	1.4	467.2084	$C_{16}H_{30}N_6O_{10}$	no hit
			$C_{15}H_{34}N_2O_{14}$	no hit
	1.4	585.7333	$C_{22}H_{46}N_8O_{11}$	no hit
			$C_{38}H_{46}O_6$	5 substances possible
	1.5	261.1238	$C_{14}H_{16}N_2O_3$	106647
	1.6	455.2134	$C_{20}H_{30}N_4O_8$	116946
	1.7	455.2139	$C_{20}H_{30}N_4O_8$	62881823
	2.1	461.2603	$C_{20}H_{36}N_4O_8$	16619887
	2.4	750.4137	$C_{34}H_{55}N_9O_{10}$	8007677
			$C_{33}H_{59}N_5O_{14}$	no hit
	2.5	392.2534	$C_{13}H_{36}N_4O_9$	no hit
	2.9	445.2338	$C_{23}H_{29}N_3O_6$	3685208
	3.3	852.5073	$C_{45}H_{73}NO_{14}$	6273
			$C_{52}H_{69}NO_9$	no hit
	3.5	415.2224	$C_{23}H_{30}N_2O_5$	2144568

	3.9	453.3205	$C_{26}H_{44}O_6$	34241116
	4.1	247.1302	$C_{11}H_{14}N_6O$	no hit
	4.9	912.6288	$C_{49}H_{77}N_{13}O_4$	no hit
	5.0	387.1811	$C_{23}H_{22}N_4O_2$	12227
	5.4	599.4374	$C_{21}H_{33}N_3O_3$	417389
	5.6	1225.5426	$C_{48}H_{80}N_{12}O_{25}$	no hit
			$C_{47}H_{84}N_8O_{29}$	no hit
	6.6	790.5895	$C_{40}H_{79N5}O_{10}$	no hit
	7.8	403.2329	$C_{20}H_{34}O_8$	9048392

Supplementary Table 10: MS data for SM9 aqueous phase crude extracts

Possible molecular formulas as predicted from the Bruker software used for Data Analysis based on two criteria: mass error (ppm) and isotope pattern matching (mSigma). In some cases two or more possible molecular formulas or molecules came up with the same possibility, so they are all included on the table. Compounds labelled as 'common impurities' are often present in various samples, therefore even though their structure is not defined, they are of no real interest. SM9 crude extract samples represent the samples which were fermented at 28°C, in the seven media above. Underlined= compounds only present to this extract, represent potential unique compound.