

**University of Zululand**



**Comprehensive comparative analysis of cytochrome P450s and those associated with  
secondary metabolism in *Salinispora* species**

**Nsikelelo Allison Malinga**

**201123095**

A dissertation submitted to the Department of Biochemistry and Microbiology, Faculty of  
Science and Agriculture, University of Zululand, KwaDlangezwa, in fulfilment of the  
requirement for the degree of Master of Science

Faculty of Science and Agriculture  
Department of Biochemistry and Microbiology  
University of Zululand  
Private Bag X 1001  
KwaDlangezwa  
3886

**Supervisor: Prof Khajamohiddin Syed**

August 2022

## DECLARATION

I, Nsikelelo Allison Malinga, declare that this dissertation is entirely my work and has not been taken from the work of others, except where I have appropriately acknowledged and referenced the source. This dissertation has never been submitted for any degree for examination at any university. Considering that quality is more important than quantity (Gould, 2016), care has been taken to present the dissertation in a publication format to enable the presentation of data in a concise manner and for easy understanding of the work. I state that the work presented in this dissertation has been published as indicated in the research outputs section.



Signed on the 24<sup>th</sup> day of August 2022

Reference: Gould, J., 2016. Future of the thesis. Nature, 535(7610), pp.26-29.

## ABSTRACT

Cytochrome P450 monooxygenases (CYPs/P450s) are heme thiolate proteins present in species across the biological kingdoms. By virtue of their broad substrate promiscuity and regio- and stereo-selectivity, these enzymes enhance or attribute diversity to the secondary metabolites. Actinomycetes species are well-known producers of secondary metabolites, especially *Salinispora* species. Despite the importance of P450s, a comprehensive comparative analysis of P450s and their role in secondary metabolism in *Salinispora* species is not reported. Thus, P450s analysis in 126 strains from three different species *Salinispora arenicola*, *S. pacifica*, and *S. tropica* is carried out in this study. The study revealed the presence of 2643 P450s that can be grouped into 45 families and 103 subfamilies. CYP107 and CYP125 families are conserved, and CYP105 and CYP107 families are bloomed (a P450 family with many members) across *Salinispora* species. Analysis of P450s that are part of secondary metabolite biosynthetic gene clusters (smBGCs) revealed *Salinispora* species have an unprecedented number of P450s (1236 P450s-47%) part of smBGCs compared to other bacterial species belonging to the genera *Streptomyces* (23%) and *Mycobacterium* (11%), phyla *Cyanobacteria* (8%) and *Firmicutes* (18%) and the classes *Alphaproteobacteria* (2%) and *Gammaproteobacteria* (18%). A peculiar characteristic of up to six P450s in smBGCs was observed in *Salinispora* species. Future characterization *Salinispora* species P450s and their smBGCs have the potential for discovering novel secondary metabolites.

## **DEDICATION**

I would like to dedicate this work to my fiancée (Miss Siyanakekelwa Hope Mthethwa) and our Son (Masilakhe Sibanisethu Malinga), as she has had my back from the day we met and has also played an instrumental role in encouraging me to complete this study, also the support I have received from her has been nothing but amazing and so I am very appreciative. To my Son, who is always my constant reminder that I should do better and always strive to achieve more in life.

## **ACKNOWLEDGEMENTS**

Firstly, I would like to thank God Almighty for giving me the wisdom to do this work and for his undying love and support.

Secondly, I would like to express my deepest gratitude to the person who made all this possible, Prof Syed, through his greatest leadership and sincerest supervision. None of this would have been possible without his guidance and support. I would also like to extend my appreciation to the following people, Prof David R Nelson (University of Tennessee, Memphis, USA), Prof Dominik Gront (Biological and Chemical Research Center, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland), Puleng Rosinah Syed and Rajshekhar Karpoormath (College of Health Sciences, University of KwaZulu-Natal, Durban 4000, South Africa) for their collaboration.

Thirdly and most importantly, I would like to thank my P450 group members (Nomfundo Nzuza and Tiara Padayachee) for their continued support and help in the completion of this work.

Lastly, I would like to sincerely extend my gratitude to my family and friends for their love and moral support.

## RESEARCH OUTPUTS

- **Article**

**Malinga, N.A.**, Nzuzi, N., Padayachee, T., Syed, P.R., Karpoormath, R., Gront, D., Nelson, D.R. and Syed, K., 2022. An Unprecedented Number of Cytochrome P450s Are Involved in Secondary Metabolism in *Salinispora* Species. *Microorganisms*, 10(5), p.871. <https://doi.org/10.3390/microorganisms10050871>

- **Conference attendance**

1. Faculty of Science, Agriculture and Engineering Symposium. 11 October 2022  
University of Zululand
2. 5<sup>th</sup> Bio Africa Convention 27 – 31 August 2022, at Durban International Convention Centre (ICC), South Africa.
3. International Conference on Chemistry Education (ICCE) 2022, 18-22 July 2022, Lagoon beach Hotel, Cape Town, South Africa.
4. Conference on Genomics, Proteomics and Metabolomics: All in the Bioinformatics (CGPMB-2019), 27-28 July 2019, Umfolozi Hotel Casino, and Convention Centre, Empangeni, KwaZulu-Natal.

## ABBREVIATIONS

Anti-SMASH	Antibiotics & Secondary Metabolite Analysis Shell
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
CDPS	tRNA-dependent cyclodipeptide synthases
CYPs/P450s	Cytochrome P450 monooxygenase
iTOL	Interactive Tree of Life
KEGG	Kyoto Encyclopedia of Genes and Genomes
MAFFT	Multiple alignment using fast Fourier transform
NAD	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
NRPS	Non-ribosomal peptide synthases
TK1PS	Type I Polyketide synthase
Trex	Tree and Reticulogram Reconstruction

## TABLE OF CONTENTS

### Contents

DECLARATION .....	2
ABSTRACT.....	3
DEDICATION .....	4
ACKNOWLEDGEMENTS .....	5
RESEARCH OUTPUTS.....	6
ABBREVIATIONS .....	7
TABLE OF CONTENTS.....	8
LIST OF FIGURES .....	10
LIST OF TABLES .....	11
CHAPTER 1: INTRODUCTION .....	12
1.1. Background and problem statement.....	12
1.2. Value to the body of knowledge .....	14
1.2.1. Novelty and Scientific contribution to new knowledge .....	14
1.2.2. Socio-economic impact .....	14
1.2.3. Scientific outcomes.....	14
1.3. Aim and Objectives .....	14
1.3.1 Aim of the Study.....	14
1.3.2 Objectives of the Study.....	14
1.4. Dissertation overview.....	15
CHAPTER 2: LITERATURE REVIEW .....	16
2.1. Cytochrome P450 monooxygenases.....	16
2.1.1. Nomenclature.....	16
2.1.2. Structure.....	17
2.1.3. Catalytic cycle .....	18
2.1.4. Applications.....	20
2.2 Secondary Metabolites .....	21
2.3. <i>Salinispora</i> species .....	22
2.4. Genome data mining .....	23
CHAPTER 3: METHODOLOGY .....	26
3.1. Species and database information .....	26

3.2. Genome data mining and identification of P450s .....	26
3.3. Assigning family and subfamily to P450s.....	26
3.4. Phylogenetic analysis of P450s .....	26
3.5. <i>Salinispora</i> species P450s profile heat-maps .....	27
3.6. Identification of P450s part of smBGCs .....	27
3.7. Data analysis .....	27
3.8. Comparative analysis of P450s and smBGCs data .....	27
CHAPTER 4: RESULTS AND DISCUSSION.....	29
4.1. <i>Salinispora</i> species P450 profiles .....	29
4.2. CYP015 and CYP107 families are bloomed in <i>Salinispora</i> species.....	36
4.3. CYP107 and CYP125 are conserved in <i>Salinispora</i> species .....	41
4.4. Unprecedented number of P450s involved in smBGCs.....	43
4.5. Functional prediction of <i>Salinispora</i> species P450s .....	51
CHAPTER 5: CONCLUSIONS AND FUTURE PERSPECTIVES.....	53
REFERENCES .....	54
ANNEXURE.....	63

## LIST OF FIGURES

<b>Figure 2. 1:</b> Cytochrome P450 absorption spectra (Bak et al., 2011).....	16
<b>Figure 2. 2:</b> Schematic diagram representing P450 nomenclature (Bak et al., 2011) .....	17
<b>Figure 2. 3:</b> Typical structure of a cytochrome P450 (Greule et al., 2018) .....	18
<b>Figure 2. 4:</b> A generic P450 active cycle schematic representation resulting in substrate hydroxylation, taken from (Greule et al., 2018). .....	19
<b>Figure 2. 5:</b> Reactions catalyzed by P450s (Lamb et al., 2007). .....	20
<b>Figure 2. 6:</b> Applications of P450s and their catalytic reactions (Kelly and Kelly, 2013). ....	21
<b>Figure 2. 7:</b> Secondary metabolites produced by <i>Salinispora</i> strains (Jensen et al., 2007)....	22
<b>Figure 2. 8:</b> Geographical distribution of <i>Salinispora</i> strains (Millan-Aguinaga et al., 2017). .....	23
<b>Figure 2. 9:</b> The classical approach and the genome-mining approach to natural product screening (Choi et al., 2015). .....	25
<b>Figure 4. 1</b> Phylogenetic analysis of <i>Salinispora</i> species P450s. 2643 P450s were used to construct the tree, and the members of the eight most abundant P450 families are highlighted in different colors and indicated in the figure. P450 protein sequences used to build the tree are listed in Table S2. A high-resolution phylogenetic tree is provided in Figure S1.....	30
<b>Figure 4. 2</b> Heat-map of P450 family conservation or co-presence analysis in <i>Salinispora</i> species. In the heat-map, the presence and absence of P450 families are indicated in red and green colors. The horizontal axis represents P450 families, and the vertical axis represents <i>Salinispora</i> species.....	42
<b>Figure 4. 3</b> Comparative analysis of P450s associated with secondary metabolism in <i>Salinispora</i> species. The P450 family name, number of P450s, and the percentage of the total number of P450s that are part of secondary metabolite biosynthetic gene clusters (smBGCs) are presented in the figure. Detailed information on secondary metabolite clusters, species, and P450s are shown in Table S1. ....	44

## LIST OF TABLES

<b>Table 4. 1</b> Comparative analysis of key features of P450s and their association with secondary metabolism between <i>Salinispora</i> species and different bacterial species. Abbreviation: No., number of; BGCs: biosynthetic gene clusters.....	31
<b>Table 4. 2</b> Genome-wide data mining and annotation of P450s in 126 <i>Salinispora</i> species. Abbreviation, No. indicates the number in the table. ....	32
<b>Table 4. 3</b> Comparative analysis of P450 families and subfamilies in <i>Salinispora</i> species. ..	37
<b>Table 4. 4</b> Secondary metabolite biosynthetic gene cluster (smBGC) types and P450s are part of the cluster in <i>Salinispora</i> species. smBGC types were again classified into different varieties based on the P450s. The smBGCs type count and the total number of P450s in the cluster variety are also presented. The same smBGCs type names listed in the antibiotics & secondary metabolite analysis shell (anti-SMASH) database (Blin et al., 2021) were used in the table. Detailed information on secondary metabolite clusters, species, and P450s are shown in Table S1.....	44

## CHAPTER 1: INTRODUCTION

### 1.1. Background and problem statement

Cytochrome P450 monooxygenases (CYPs/P450s) comprise a superfamily of heme-thiolate proteins. P450s are present in all species of different biological kingdoms, including in viruses considered non-living entities (Nelson, 2018, Lamb et al., 2019). This suggests that these enzymes play an important role in species' primary and secondary metabolism. These enzymes were initially identified as monooxygenases due to their ability to introduce one oxygen atom into a substrate (White and Coon, 1980). Subsequent research revealed that P450s are catalytically diverse enzymes performing some unusual enzymatic reactions (Sono et al., 1996, Bernhardt, 2006, Kelly and Kelly, 2013, Guengerich and Munro, 2013, Lamb and Waterman, 2013). The regio- and stereo-specific oxidation of many substrates by P450s caught the attention of researchers for biotechnological exploration of these enzymes (Girvan and Munro, 2016, Urlacher and Eiben, 2006, Bernhardt and Urlacher, 2014, Li et al., 2020). P450s reactions are essential in designing drugs such that drug toxicity of prodrugs is primarily assessed against these enzymes (Guengerich, 2020). Also, P450s play a vital role in detoxifying xenobiotic compounds (Esteves et al., 2021). Microbial P450s, especially from lower eukaryotes such as fungal CYP51, have been used as an azole drug target (Debnath et al., 2017, Lepesheva et al., 2018). The study also suggested that fungal CYP53 can be a potential alternative drug target (Jawallapersand et al., 2014). One of the best examples of P450s biotechnological applications is synthesizing antibiotics and anticancer drugs (Andersen et al., 1993, Bischoff et al., 2005, Jennewein et al., 2005, van Beilen et al., 2005).

The utilization of P450s in the generation of secondary metabolites or natural products, organic compounds not directly involved in an organism's normal growth, development, or reproduction, is gaining momentum as reactions catalyzed by these enzymes contribute to the secondary metabolite diversity (Podust and Sherman, 2012, Greule et al., 2018). Secondary metabolites, their structural diversity, bioactivity, and ecological functions, including their application in almost all areas of biology, have been thoroughly reviewed (Vaishnav and Demain, 2011, Demain and Fang, 2000, Thirumurugan et al., 2018, Sharma et al., 2014, Abegaz and Kinfe, 2019, Katz and Baltz, 2016). For example, secondary metabolites are widely used in human and veterinary medicine, agriculture, and manufacturing .

Secondary metabolites in organisms are produced by a set of genes usually located next to each other as a cluster known as secondary metabolite biosynthetic gene cluster (smBGCs)

(Medema et al., 2015). Earlier, researchers used to clone and sequence smBGCs to identify the genes/proteins involved in producing a particular secondary metabolite. The onset of genome sequencing and the advancement of science, especially in bioinformatics, led to the development of software programs that can automatically detect smBGCs (Weber and Kim, 2016). Due to this advancement, many smBGCs were reported in species belonging to different biological kingdoms .

In the bacterial kingdom, species belonging to the phylum *Actinobacteria* are well-known for producing secondary metabolites (Nair and Abraham, 2020, Jose et al., 2021, Berdy, 2005, Barka et al., 2016), especially species of the genus *Streptomyces* (de Lima Procópio et al., 2012). It is a well-known fact that two-thirds of the clinically valuable antibiotics come from *Streptomyces* species (de Lima Procópio et al., 2012). Actinomycetes belonging to the genus *Salinispora* produce biotechnologically valuable secondary metabolites (Bonet et al., 2015, Tyc et al., 2017, Penn et al., 2009, Asolkar et al., 2010, Eustaquio et al., 2011, Jensen et al., 2015, Jensen et al., 2007, Matsuda et al., 2009, Ziemert et al., 2014). Salinosporamide A, a secondary metabolite, is one of the best examples, now under clinical trials as an anticancer drug (Fenical et al., 2009).

*Salinispora* is the first genus of *Actinobacteria* identified for its requirement of seawater for growth . This genus includes three distinct but closely related species *Salinispora arenicola*, *S. pacifica*, and *S. tropica* (Barka et al., 2016, Maldonado et al., 2005, Ahmed et al., 2013). *Salinispora* species are widely distributed in tropical and subtropical marine environments with distinct geographical patterns (Jensen and Mafnas, 2006, Contador et al., 2019). The genome sequence of *S. tropica* revealed a large percentage of its genome (~9.9%) is dedicated to natural products biosynthesis, which was greater than any other natural product producing actinomycetes . The genome sequencing analysis revealed that P450s were also part of smBGCs . CYP107 from *S. arenicola* CNS-205 is involved in the biosynthesis of secondary metabolites, saliniketal, and rifampicin (Wilson et al., 2010). Apart from these notable mentions, no information is available on *Salinispora* species P450s.

Despite knowing that *Salinispora* species produce different types of human valuable secondary metabolites/natural products and the role of P450s in attributing diversity to these compounds, to date, comparative analysis of P450s and their role in secondary metabolism in *Salinispora* species is not reported. This study is aimed to address this research gap by performing genome-wide data mining, identification, annotation (assigning family and

subfamily), and phylogenetic analysis of P450s in *Salinispora* species. The study also encompasses identification of P450s part of smBGCs, and comparative analysis of *Salinispora* P450 features with other bacterial species belonging to the genera, *Streptomyces* and *Mycobacterium*, phyla *Firmicutes* and *Bacteroidetes*, and the classes *Alpha-* and *Gamma-proteobacteria*.

## **1.2. Value to the body of knowledge**

### **1.2.1. Novelty and Scientific contribution to new knowledge**

This study is the first report on a comparative analysis of P450s and their role in secondary metabolite production in *Salinispora* species.

### **1.2.2. Socio-economic impact**

- The study has the potential to identify novel secondary metabolite biosynthetic gene clusters. These gene clusters can be used to synthesize antibiotics useful for humans.
- Human capacity development – black male candidate training in Bioinformatics, which is a scarce skilled research area.

### **1.2.3. Scientific outcomes**

- MSc dissertation
- One publication in internationally reputed accredited journals

## **1.3. Aim and Objectives**

### **1.3.1 Aim of the Study**

Comparative analysis of cytochrome P450s and those associated with secondary metabolism in *Salinispora* species.

### **1.3.2 Objectives of the Study**

- To perform genome data mining and identification of P450s in selected *Salinispora* species.
- To annotate P450s i.e., assignment of P450 families and P450 subfamilies.
- To perform a phylogenetic analysis of *Salinispora* species P450s.
- To identify and classify biosynthetic gene clusters in *Salinispora* species.
- To identify P450s involved in secondary metabolite synthesis in *Salinispora* species.
- To perform comparative and diversity analysis of P450s of *Salinispora* species with other bacterial species.
- To predict P450 functions in *Salinispora* species.

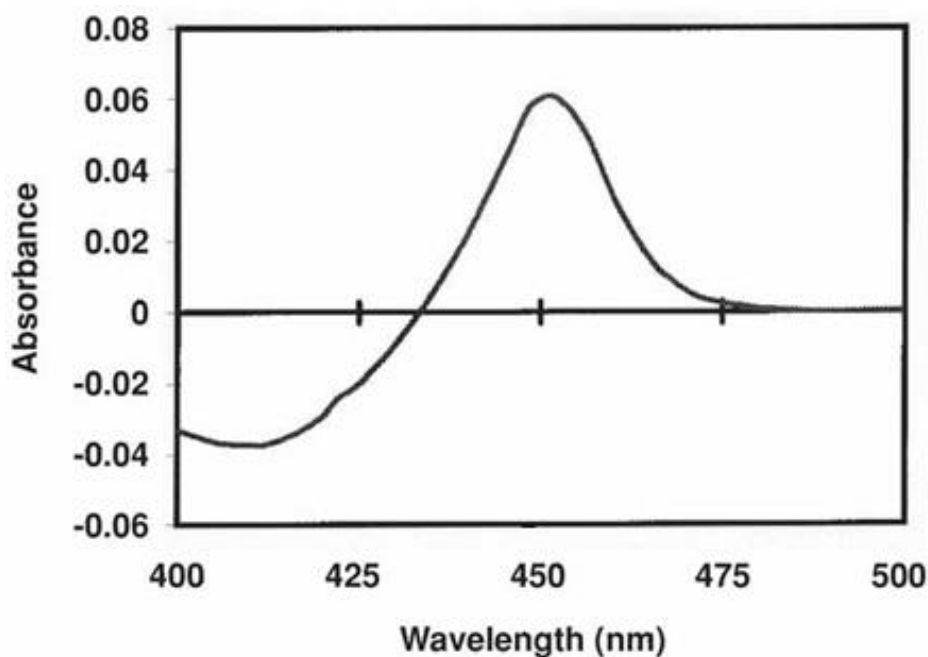
#### **1.4. Dissertation overview**

This dissertation is divided into five chapters. Chapter 1 is an introduction that provides an overview of the background and problem statement, value to the body of knowledge, and study aim and objectives. Chapter 2 consists of a literature review. Chapters 3-5 consist of results & discussion, methodology and conclusion & future perspectives. Chapters 3-5 are presented in research article format, the same as the published article in the journal *Microorganisms* (impact factor 4.92). Furthermore, the information provided under sections abstract and background & problem statement is the same as the published article.

## CHAPTER 2: LITERATURE REVIEW

### 2.1. Cytochrome P450 monooxygenases

Cytochrome P450 monooxygenases (CYPs or P450s), also known as mixed-function oxidases, belong to a superfamily of heme-thiolate proteins that catalyze many enzymatic reactions (Sono et al., 1996, Bernhardt, 2006). They are known for inserting one oxygen into a substrate (White and Coon, 1980). P450 proteins are named due to their absorption of light at the 450 nm wavelength of their carbon-monoxide-bound form (Figure 2.1) (Omura and Sato, 1962, Klingenberg, 1958).



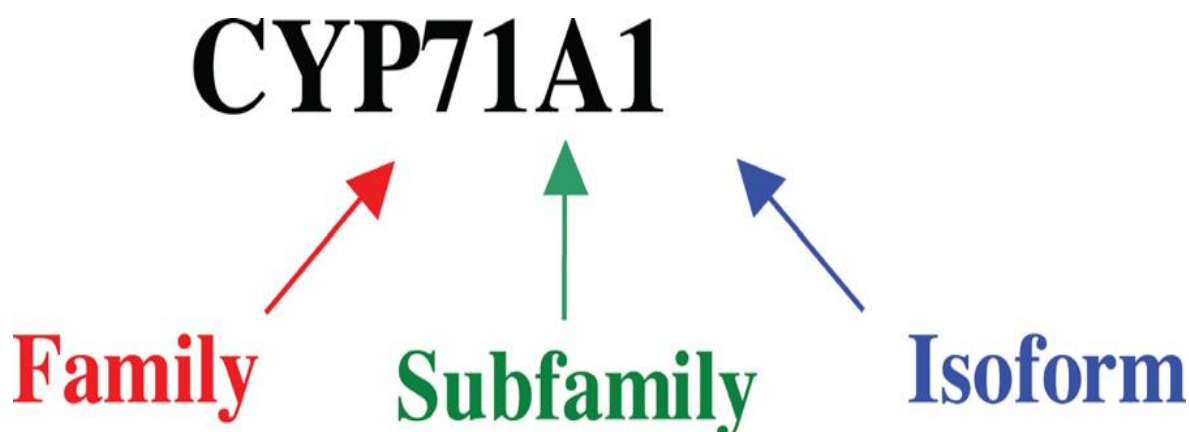
**Figure 2. 1:** Cytochrome P450 absorption spectra (Bak et al., 2011).

#### 2.1.1. Nomenclature

A method for classifying P450s has been developed based on the percentage sequence identity. The family and subfamily concept has been proposed where families and subfamilies must share more than 40% and 55% of sequence identity, respectively (Figure 2.2) (Nelson, 1998, Nelson, 2006, Nelson, 2013).

Interestingly, this method of nomenclature has nothing to do with the functional role, where substrate discrimination and catalytic role can vary extensively within P450 families – classic

examples of P450 families that include members displaying highly varied functions include the CYP105, CYP107, and CYP109 families (Greule et al., 2018).



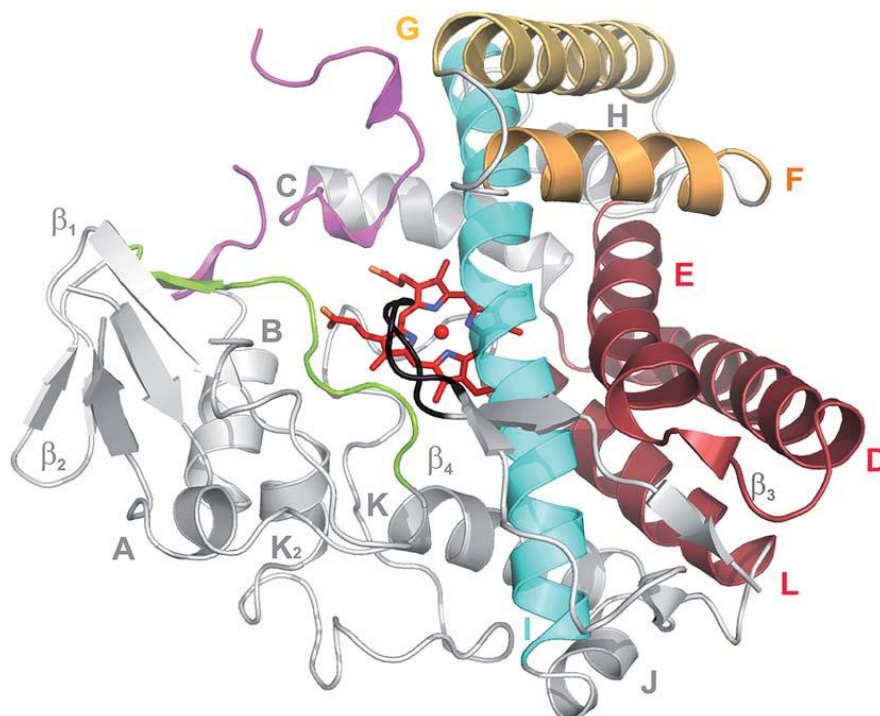
**Figure 2. 2:** Schematic diagram representing P450 nomenclature (Bak et al., 2011)

### 2.1.2. Structure

The structure of P450s is extraordinarily conserved, typically comprising a helical secondary structure (standard helices are labeled A–L) complemented by one  $\beta$ -sheet section (Figure 2.3). Forming the essential center of a P450 is a four-helix package containing the D, E, I, and L-helices, together with a prosthetic heme group in the middle of the P450 active site packed in amongst the I- and L-helices (Sirim et al., 2010, Dodani et al., 2014). The I-helix, which drives through the upper face of the heme group, comprises a classically well-preserved acid/alcohol pair typically imperative aimed at the P450 active cycle, while the loop just before the beginning of the L-helix comprises a cysteine remains that offers the critical proximal axial thiolate ligand for the heme iron of the P450 (Greule et al., 2018). The proximal face of a P450 is also the site where interactions with redox partner proteins happen, with these progressions mostly seeming to be compelled by electrostatic interactions (Dodani et al., 2014, Xu and Du, 2018).

P450s active site contains areas that border the distal face of the heme moiety and contain the alleged BC ring section in the middle of the B- and C-helices, the C-terminal section of the F-helix, the N-terminal part of the G-helix, the midpoint of the I-helix, the b-strand succeeding the K-helix and the C-terminal ring that impinges of the P450 active site (such areas are also denoted to as substrate recognition sites, SRSs) (Bak et al., 2011, Shumyantseva et al., 2018). The capability of variations within these areas leads to the capability of P450s to receive a

wide range of different substrates. At the same time, the preservation of the overall structure surrounding the active site heme group and axial cysteine ligand support the extraordinary uniformity of the P450 catalytic mechanism (Urlacher and Girhard, 2012).



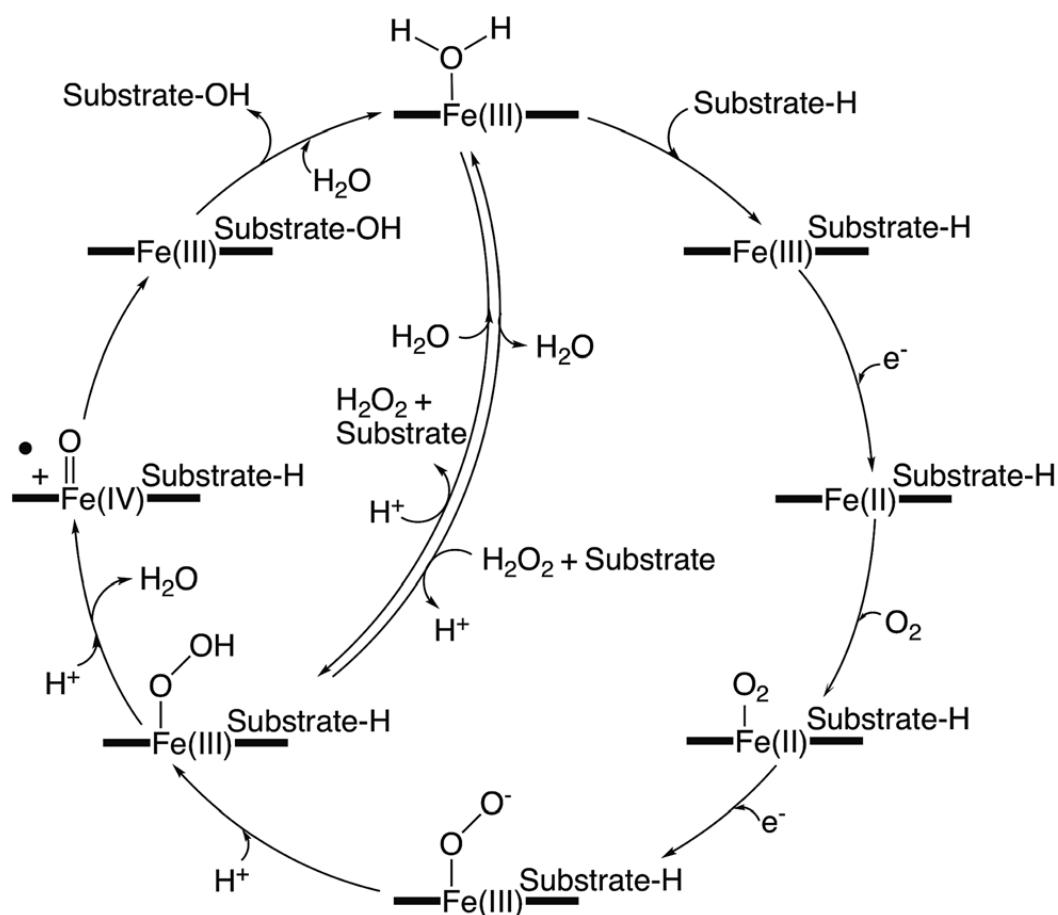
**Figure 2. 3:** Typical structure of a cytochrome P450 (Greule et al., 2018)

### 2.1.3. Catalytic cycle

Though xenobiotic metabolism in higher eukaryotes highlights an amazing capability that P450s can portray for substrate manipulability, it is in the kingdom of bacterial P450s – mainly biosynthetic ones – where the extraordinary variety of biochemical conversions that P450s can facilitate is discovered (Chigorimbo-Murefu et al., 2013, Moody and Loveridge, 2014, Greule et al., 2018). Also, this is combined with an extraordinary variety of diverse, multifaceted substrates that points out the astonishing abilities of these robust biocatalysts (Shumyantseva et al., 2018).

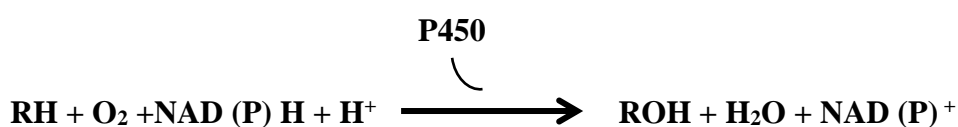
While the typical P450-catalysed conversion, hydroxylation of inactivated C–H bonds stays the powerful feature of P450-catalysed reaction, the continued reports of various chemical reactions that can be executed by P450s are forever going on the rise. Numerous instances of these are established in bacterial biosynthetic pathways: such reactions vary from aromatic hydroxylation, alkene epoxidation, and heteroatom oxidation to processes such as aromatic

coupling and multistep transformations (Dodani et al., 2014, Rudolf et al., 2017). One of the unique properties of P450s is that these reactions are all performed within a conserved active cycle (Figure 2. 4). In this process, a carefully designed transfer of two electrons and two protons to molecular oxygen is performed by the P450. This extremely electrophilic oxidant, which has been recognized as an iron- porphyrin cation radical, is accountable for the massive majority of the oxidation chemistry exhibited by P450s (Gricman et al., 2016).

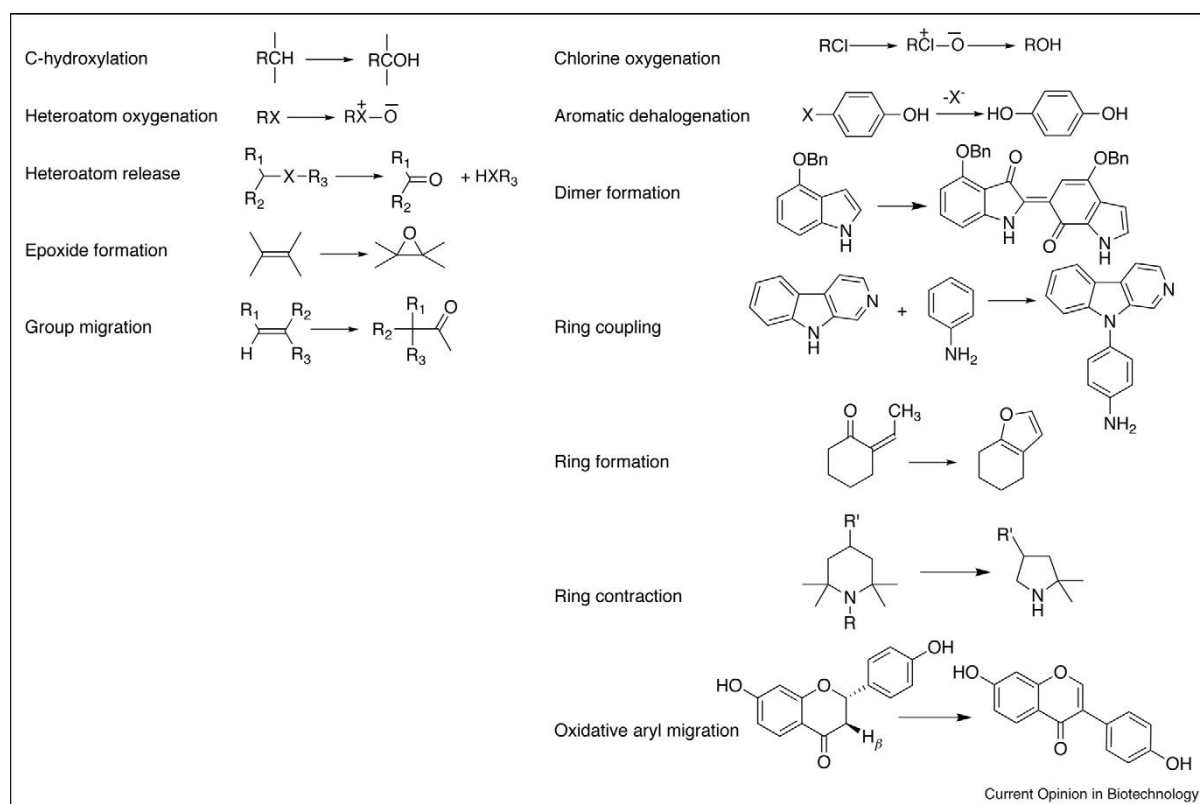


**Figure 2. 4:** A generic P450 active cycle schematic representation resulting in substrate hydroxylation, taken from (Greule et al., 2018).

P450 have been well known for their generic monooxygenases reaction also represented as follows:



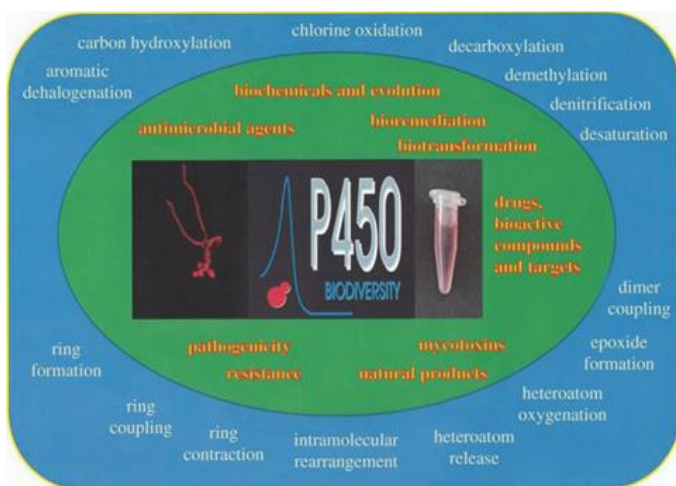
The RH and ROH represent the substrate and its hydroxylated form in this reaction. (Bernhardt, 2006). Apart from the reaction mentioned above, many other reactions performed by P450s are illustrated in figure 2.5.



**Figure 2. 5:** Reactions catalyzed by P450s (Lamb et al., 2007).

### 2.1.4. Applications

Due to their ability to catalyze a variety of reactions on various substrates, P450s have been of key interest to many chemists and biochemists as they play an important role in drug metabolism (Figure 2.6) (Bernhardt, 2006, Kelly and Kelly, 2013, Rudolf et al., 2017, Naveed et al., 2018). Moreover, P450s have also been found to be very crucial in the biosynthesis of secondary metabolites (Figure 2.6) (Greule et al., 2018, Kelly and Kelly, 2013).

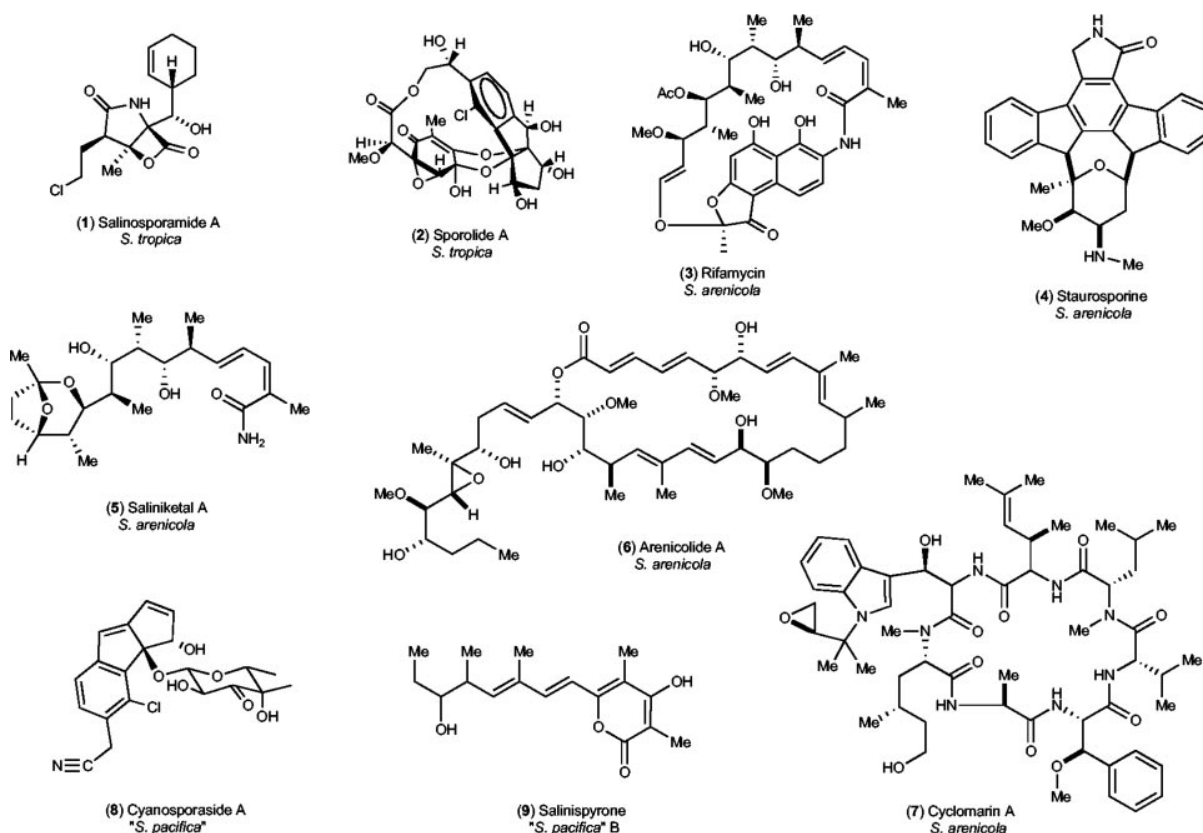


**Figure 2. 6:** Applications of P450s and their catalytic reactions (Kelly and Kelly, 2013).

## 2.2 Secondary Metabolites

Natural products from marine microorganisms offer a vast number of poorly explored bioactive molecules that are of prospective importance in the production of new pharmaceutical agents (Subramani and Aalbersberg, 2012, Fiedler et al., 2005, Bose et al., 2014, Bose et al., 2015, Peters et al., 2003, Hwang et al., 2014). Marine bacteria have indeed been demonstrated to be a reputable source of molecules. An enormous bulk of compounds is revealed from chemically prolific Gram-positive bacteria belonging to the order *Actinomycetales* (*Actinomycetes*) (Lam, 2006, Hassan et al., 2017, Newman and Hill, 2006). Within the marine actinomycetes, where many novel natural products continue being reported. The genus *Salinispora* has generated a remarkable collection of structurally varied and pharmaceutically relevant compounds (Jensen et al., 2015), such as Salinosporamide A, which is presently under medical examination for the treatment of cancer (Bose et al., 2015). These compounds are manufactured in species-specific designs and also include effective antibacterial antibiotics (rifamycins) class and anti-inflammatory (cyclomarin D) and antimalarial (salinosporamide A) products, which are constantly detected in sources of *Salinispora arenicola* (Asolkar et al., 2010, Manivasagan et al., 2014). Rifamycins are a collection of polyketide antibiotics that belong to the family of ansamycins manufactured by a collection of soil-derived actinomycetes of the species *Amycolatopsis mediterranei* (Bose et al., 2015). These compounds stimulate their antibacterial action through a direct inhibition of RNA production, binding to the  $\beta$ -subunit of RNA polymerase. Rifamycin-related mixtures/compounds have also been reported and known to be produced by *S. arenicola*

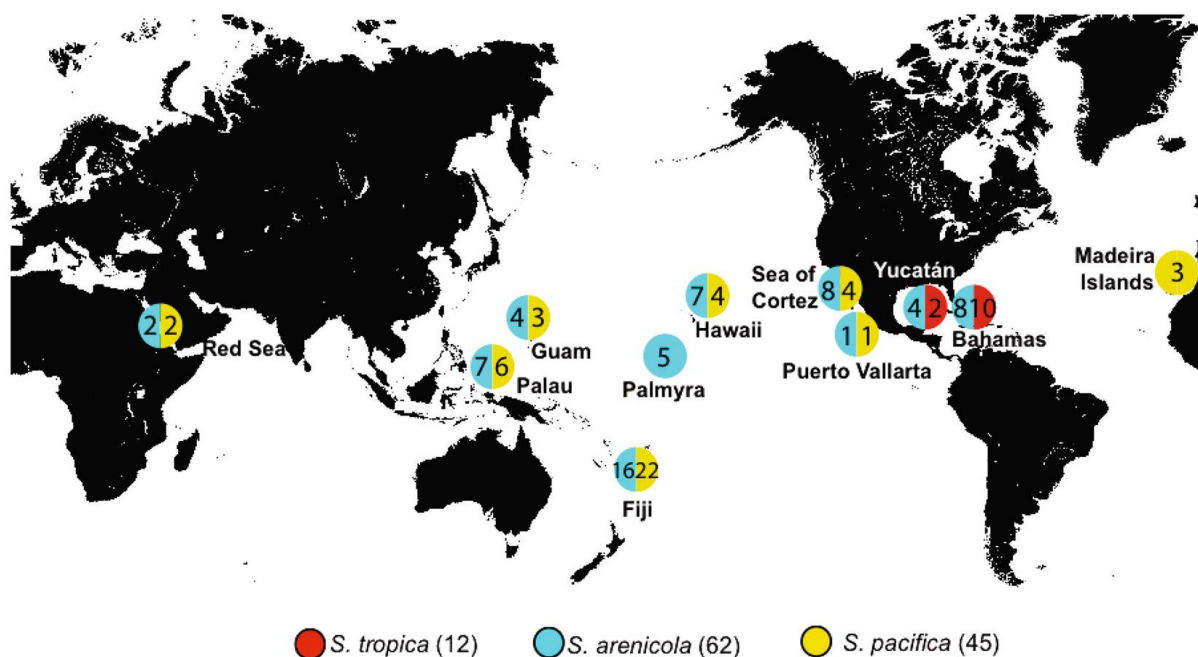
extracted from Great Barrier Reef aquatic sponges (Bose et al., 2014, Bose et al., 2017). Different natural products produced by *Salinispora* species are presented in Figure 2.7.



**Figure 2. 7:** Secondary metabolites produced by *Salinispora* strains (Jensen et al., 2007).

### 2.3. *Salinispora* species

The genus *Salinispora* belongs to the phylum *Actinobacteria*. Species in this phylum are an inexhaustible source of structurally diverse secondary metabolites (Letzel et al., 2017, Udvary et al., 2007, Penn et al., 2009). The genus *Salinispora* is widely distributed (Figure 2.8) and found in tropical and subtropical ocean sediments. At the present moment, there are two officially defined species, *S. tropica* and *S. arenicola*, and a third species for which the name *S. pacifica* has been suggested (Millan-Aguinaga et al., 2017).



**Figure 2. 8:** Geographical distribution of *Salinispora* strains (Millan-Aguinaga et al., 2017).

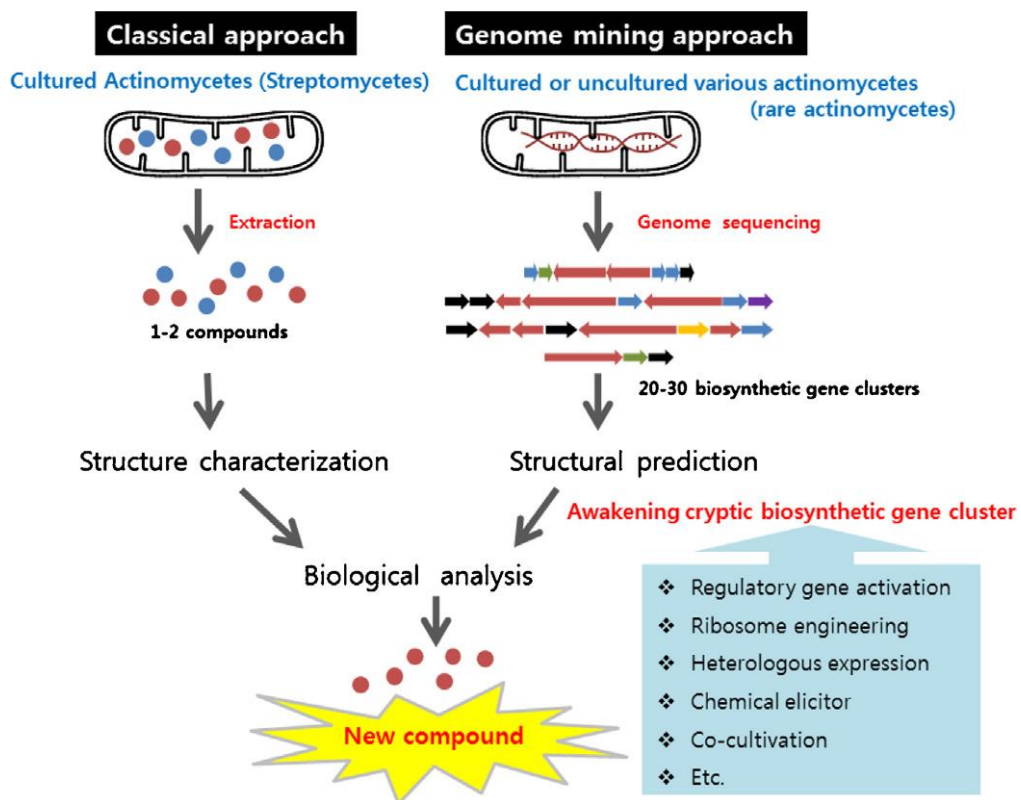
#### 2.4. Genome data mining

Ever since Penicillin was discovered in the twentieth century, microorganisms have been of paramount importance in producing over 23,000 active compounds (i.e., antimicrobials, antivirals, cytotoxic and immunosuppressive compounds) (Choi et al., 2015). As much as efforts are being put into discovering new metabolites, less than 1% of all currently known microorganisms have been identified and characterized (Duncan et al., 2015). There is still quite a large pool of uncultured microorganisms and numerous natural products that are dispersed in natural environments (Frasch et al., 2013). Over the past few years, the manifestation of antibiotic resistance and new diseases has increased the importance of finding new bioactive compounds for pharmaceutical drug leads (Li et al., 2017a). Research on the outstanding 99% of new microbes from unmapped habitats plays a vital role in finding new antibiotics and other therapeutic agents against drug-resistant bacteria (Jensen et al., 2014).

Actinomycetes have become a current focus in the hunt for new natural products with a broad variety of biological activities and pharmacological properties (Kersten et al., 2013, Qhanya et al., 2015, Udvary et al., 2007). Furthermore, the progression of genome mining technologies has lately resulted in the discovery of various natural products from rare actinomycetes (Liu, 2016, Li et al., 2017b). Genome mining is a powerful example for detecting and describing natural product biosynthetic genes (Syed et al., 2013, Letzel et al.,

2017). As shown in Figure 2.9 the advantages of the genome mining approach over the classical approach denote excellent potential for obtaining uncultivable actinomycetes and an enormous amount of genomic data (cryptic gene clusters) (Choi et al., 2015). The genes responsible for synthesizing different secondary metabolites are part of gene clusters included on the P450 gene sequence called secondary metabolite biosynthetic gene clusters (smBGCs) (Cimermancic et al., 2014).

Since the earliest reports of complete sequences of bacterial species genomes, various genome sequences, and studies have discovered many cryptic (silent) gene clusters that encode secondary metabolism (Choi et al., 2015). The genomes of actinomycetes hide an enormous degree of chemical diversity encrypted in silent biosynthetic gene clusters (Li et al., 2017b, Letzel et al., 2017, Penn et al., 2009). It is remarkably challenging to isolate and characterize these cryptic gene clusters that code for natural products within organisms using conventional methods. Thus, many researchers have concentrated on genome mining to address silent gene clusters encoding potentially valuable new natural products (Liu, 2016, Khater et al., 2016). Hence the *in silico* screening of annotated sequences from different publicly available databases has been considered one of the most promising strategies to date. The wide-ranging or draft sequencing of the whole genomes of actinomycetes is beginning to make contributions to an age of more rapid discovery of new compounds (Eustaquio et al., 2011, Jensen et al., 2014, Jensen et al., 2015, Kersten et al., 2013, Li et al., 2017b).



**Figure 2. 9:** The classical approach and the genome-mining approach to natural product screening (Choi et al., 2015).

## CHAPTER 3: METHODOLOGY

### 3.1. Species and database information

A total of 126 *Salinispora* species genomes (permanent and finished draft genomes) are available for public use at the Joint Genome Institute Integrated Microbial Genomes and Microbiomes (JGI IMG/M) (Chen et al., 2021, Mukherjee et al., 2021) were used in this study (last accessed on 2nd February 2022). Information on the species and their genome IDs used in the study is provided in Table S1.

### 3.2. Genome data mining and identification of P450s

Genome data mining and identification of P450s in *Salinispora* species were carried out following the protocol described elsewhere (Mnguni et al., 2020, Syed et al., 2019). Each *Salinispora* species genome available at JGI IMG/M (Chen et al., 2021, Mukherjee et al., 2021) was searched for P450s using the InterPro code "IPR001128". The hit protein sequences were then searched for the presence of P450 characteristic motifs such as EXXR and CXG (Syed and Mashele, 2014, Gotoh, 1992). Proteins with one of these motifs or short amino acid length are considered P450-fragments. P450 fragments were not considered for the final P450 family and subfamily count.

### 3.3. Assigning family and subfamily to P450s

Above selected P450s were assigned to different families and subfamilies based on the International P450 Nomenclature Committee rule (Nelson et al., 1993, Nelson, 2006, Nelson, 1998), proteins with a percentage identity greater than 40% were assigned to the same family as named homolog P450s, and those that had greater than 55% identity were assigned to the same subfamily as named homolog P450s. Proteins with a percentage identity of less than 40% were assigned to a new family. *Salinispora* species P450s, along with P450-fragments, are presented in Table S2.

### 3.4. Phylogenetic analysis of P450s

Phylogenetic analysis of P450s was carried out following the procedure described elsewhere (Nzuza et al., 2021a, Nzuza et al., 2021b). The phylogenetic tree of P450s was constructed using protein sequences. Firstly, the MAFFT v6.864 (Kato et al., 2005) was used to align the Trex web server's protein sequences (Boc et al., 2012). The alignments were then used to interpret the best tree by the Trex web server (Boc et al., 2012). Finally, the best-inferred tree was visualized, colored, and generated by a web-based tool, VisuaLife (Kryś and Gront, 2021).

### 3.5. *Salinispora* species P450s profile heat-maps

P450 profile heat-maps were generated following a method described elsewhere (Msomi et al., 2021, Nzuza et al., 2021b) to check the presence and absence of or co-presence of or conserved nature of P450 families in *Salinispora* species. Briefly, a tab-delimited file was imported into Multi-Experiment Viewer (Mev) (Howe et al., 2011), and hierarchical clustering using a Euclidean distance metric was used to cluster the data. 126 *Salinispora* species formed the vertical axis, and P450 families formed the horizontal axis. Data were presented as -3 for family absence (green) and 3 for family presence (red).

### 3.6. Identification of P450s part of smBGCs

P450s that are part of smBGCs were identified following the method described elsewhere (Mnguni et al., 2020, Syed et al., 2019). Briefly, for each *Salinispora* species genome available at JGI IMG/M (Chen et al., 2021, Mukherjee et al., 2021), the smBGCs were searched for the presence of P450s using the P450 gene ID. The cluster type is noted if a P450 is found as part of the cluster. Results were recorded on Excel spread-sheets and represented species-wise smBGCs, smBGC type, and P450s part of specific smBGCs. Among 126, only 103 *Salinispora* species smBGCs information is available at JGI IMG/M (Chen et al., 2021, Mukherjee et al., 2021). Thus the same 103 *Salinispora* species smBGCs were analyzed for the presence of P450s (Table S1).

### 3.7. Data analysis

All calculations were carried out following the procedure reported previously by our laboratory (Msomi et al., 2021). The average number of P450s was calculated using the formula: Average number of P450s = Number of P450s/Number of species. The P450 diversity percentage was calculated using the formula: P450 diversity percentage =  $100 \times \text{Total number of P450 families} / \text{Total number of P450s} \times \text{Number of species with P450s}$ . The percentage of P450s that formed part of BGCs was calculated using the formula: Percentage of P450s part of BGCs =  $100 \times \text{Number of P450s part of BGCs} / \text{Total number of P450s present in species}$ .

### 3.8. Comparative analysis of P450s and smBGCs data

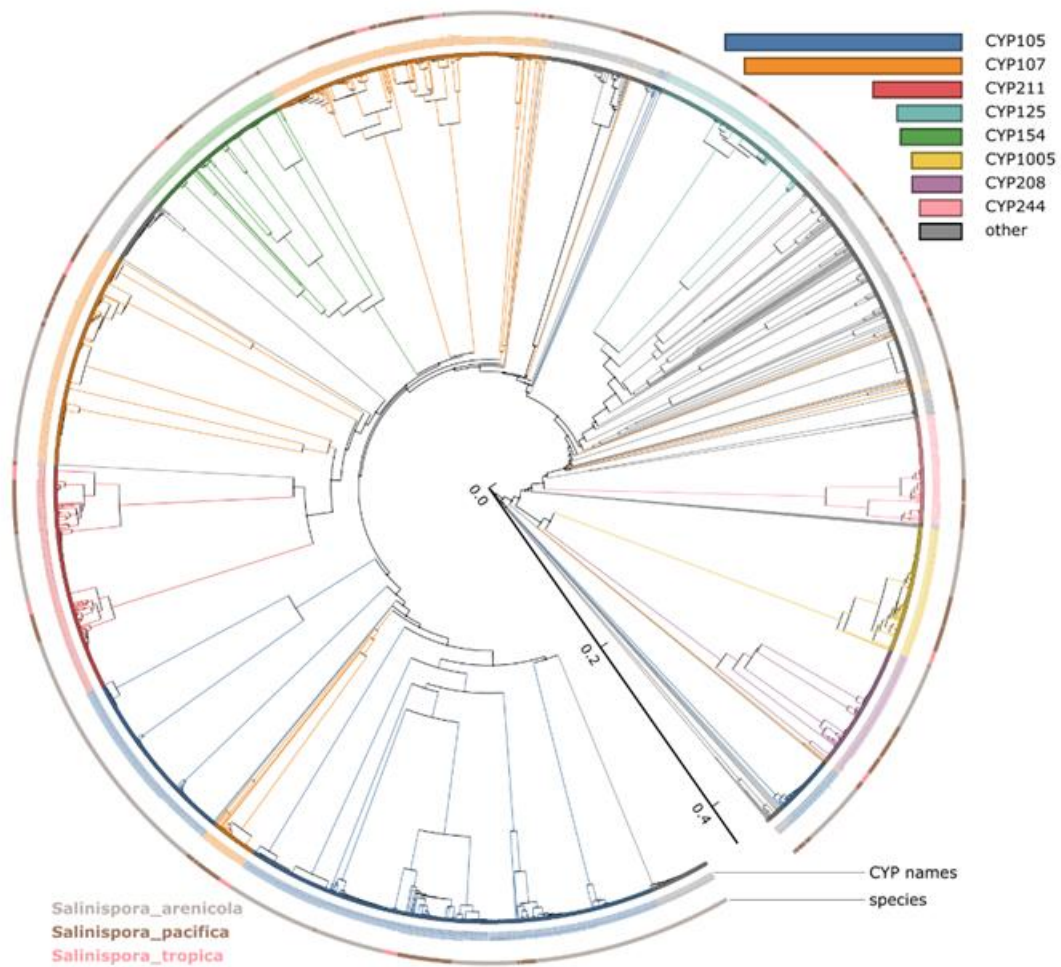
For comparative analysis of P450s and smBGCs, information for bacterial species belonging to different groups such as classes, Alpha- and *Gamma-proteobacteria* (Msomi et al., 2021, Nzuza et al., 2021b), phyla, *Firmicutes* (Padayachee et al., 2020) and Cyano-bacteria (Khumalo, 2020), and the genera, *Streptomyces* (Senate et al., 2019, Mnguni et al., 2020),

*Mycobacterium* (Parvez et al., 2016, Senate et al., 2019), was resourced from published articles.

## CHAPTER 4: RESULTS AND DISCUSSION

### 4.1. *Salinispora* species P450 profiles

Genome-wide data mining and annotation of P450s in 126 *Salinispora* species revealed the presence of 2643 P450s in their genomes (Figure 4.1, Tables 4.1 and 4.2). The P450 count in *Salinispora* species ranged from 10 to 35 P450s, with an average of 21 P450s (Tables 4.1 and 4.2). Apart from the complete P450 sequences, 129 P450 fragments were also found in some *Salinispora* species (Table 4.2). P450 fragments in species are natural (Mnguni et al., 2020, Msomi et al., 2021, Senate et al., 2019), and thus, these were excluded from further analysis. Among *Salinispora* species, *S. arenicola* CNY280 has the highest number of P450s (35 P450s), and *S. pacifica* CNS801 and *S. pacifica* CNT148 have the lowest number of P450s (10 P450s each) (Table 4.2). Comparative analysis revealed that *Salinispora* species have the highest average number of P450s than species belonging to Cyanobacteria, *Firmicutes*, *Alphaproteobacteria*, and *Gammaproteobacteria* (Table 4.1). However, *Salinispora* species had the lowest average number of P450s compared to species belonging to *Streptomyces* and *Mycobacterium* (Table 4.1). A point to be noted is that, among bacterial species, species belonging to the phylum Actinobacteria have the highest average number of P450s (Table 4.1). This indicates selective enrichment of P450s in these species due to their adaptation to ecological niches vis a vis P450s, helping them adapt to diverse ecological niches described elsewhere (Senate et al., 2019, Parvez et al., 2016, Mnguni et al., 2020). *Salinispora* species P450s, along with P450-fragments, are presented in Table S2.



**Figure 4. 1** Phylogenetic analysis of *Salinispora* species P450s. 2643 P450s were used to construct the tree, and the members of the eight most abundant P450 families are highlighted in different colors and indicated in the figure. P450 protein sequences used to build the tree are listed in Table S2. A high-resolution phylogenetic tree is provided in Figure S1.

**Table 4. 1** Comparative analysis of key features of P450s and their association with secondary metabolism between *Salinispora* species and different bacterial species.

Abbreviation: No., number of; BGCs: biosynthetic gene clusters.

Category	<i>Salinispora</i> species	<i>Streptomyces</i> species	Mycobacterial species	Cyanobacterial species	<i>Firmicutes</i> species	Alphaproteobacterial species	Gammaproteobacterial species
Species analysed	126	203	60	114	972	599	1261
Species without P450s	0	0	0	0	743	370	1091
Species with P450s	126	203	60	114	229	229	169
Percentage of species with P450s	100	100	100	100	24	38	13
No. of P450s	2643	5460	1784	341	712	873	277
No. of families	45	253	77	36	14	143	81
No. of subfamilies	103	698	132	79	53	214	102
Dominant P450 family	CYP105	CYP107	CYP125	CYP110	CYP107	CYP202	CYP133 & CYP107
Average No. of P450s	21	27	30	3	3	4	2
P450 diversity percentage	0,01	0,02	0,07	0,09	0,01	0,07	0,17
No. of P450s part of BGCs	1236	1231	204	27	126	21	49
No. of P450	35	135	31	6	10	16	22

families part of BGCs							
Percentage of P450s part of BGCs	47	23	11	8	18	2	18
Reference(s)	This study	(Senate et al., 2019, Mnguni et al., 2020)	(Senate et al., 2019, Parvez et al., 2016)	(Khumalo, 2020)	(Padayachee et al., 2020)	(Nzuzi et al., 2021)	(Msomi et al., 2021)

**Table 4. 2** Genome-wide data mining and annotation of P450s in 126 *Salinispora* species.

Abbreviation, No. indicates the number in the table.

Species name	No. of P450s	No. of P450 fragments	No. of P450 families	No. of Subfamilies
<i>Salinispora arenicola</i> CNH996	26	6	14	25
<i>Salinispora arenicola</i> CNH996B	27		14	25
<i>Salinispora arenicola</i> CNY280	35		18	32
<i>Salinispora arenicola</i> CNH877	34		15	30
<i>Salinispora arenicola</i> CNS848	32		16	29
<i>Salinispora arenicola</i> CNT798	31		14	27
<i>Salinispora arenicola</i> CNH643	31	1	14	28
<i>Salinispora arenicola</i> CNS-991	31		15	28
<i>Salinispora arenicola</i> CNT799	31		14	28
<i>Salinispora arenicola</i> CNY679	31	1	14	27
<i>Salinispora arenicola</i> CNT850	31		13	27
<i>Salinispora arenicola</i> CNT800	31		14	28
<i>Salinispora arenicola</i> CNY011	31		14	26
<i>Salinispora arenicola</i> CNY230	30		17	30
<i>Salinispora arenicola</i> CNH713	30		14	27
<i>Salinispora arenicola</i> CNH905	31	1	14	28
<i>Salinispora arenicola</i> CNT857	30		14	28
<i>Salinispora arenicola</i> CNY281	29	1	17	29

<i>Salinispora arenicola</i> CNH941	29		14	26
<i>Salinispora arenicola</i> CNB527	29	4	15	27
<i>Salinispora arenicola</i> CNT859	29		13	26
<i>Salinispora arenicola</i> CNT005	28		16	28
<i>Salinispora arenicola</i> CNH964	28	1	14	24
<i>Salinispora arenicola</i> CNP193	28		14	26
<i>Salinispora arenicola</i> CNP105	28	2	14	25
<i>Salinispora arenicola</i> CNH646	28		14	26
<i>Salinispora arenicola</i> CNR425	28		15	28
<i>Salinispora arenicola</i> CNS-205	28	1	15	28
<i>Salinispora arenicola</i> ATCC BAA-917	27	13	11	21
<i>Salinispora arenicola</i> CNY685	26	6	14	26
<i>Salinispora arenicola</i> CNS325	26		13	26
<i>Salinispora arenicola</i> CNS744	26		13	26
<i>Salinispora arenicola</i> CNY694	26	6	13	26
<i>Salinispora arenicola</i> CNY260	26		14	26
<i>Salinispora arenicola</i> CNT-088	26	1	13	24
<i>Salinispora arenicola</i> CNB458	26	4	13	26
<i>Salinispora arenicola</i> CNS296	25	1	14	25
<i>Salinispora arenicola</i> CNY231	25		14	26
<i>Salinispora arenicola</i> CNY282	25		13	25
<i>Salinispora arenicola</i> CNS299	25	1	14	25
<i>Salinispora arenicola</i> CNQ748	25		13	25
<i>Salinispora arenicola</i> CNY244	25		13	25
<i>Salinispora arenicola</i> CNS820	25		13	25
<i>Salinispora arenicola</i> CNS673	25		13	25
<i>Salinispora arenicola</i> CNY237	24		12	24
<i>Salinispora arenicola</i> CNS342	24	1	13	24
<i>Salinispora arenicola</i> CNH718	24	1	13	24
<i>Salinispora arenicola</i> CNX891	24	3	15	24
<i>Salinispora arenicola</i> CNY256	24		13	25

<i>Salinispora arenicola</i> CNS243	24	1	13	24
<i>Salinispora arenicola</i> CNY234	24	1	13	24
<i>Salinispora arenicola</i> CNY690	24	4	13	24
<i>Salinispora arenicola</i> CNQ884	23	1	13	25
<i>Salinispora arenicola</i> CNR107	22		12	22
<i>Salinispora arenicola</i> CNR921	22		12	22
<i>Salinispora arenicola</i> CNH962	22	1	12	22
<i>Salinispora arenicola</i> CNX481	22	2	12	22
<i>Salinispora arenicola</i> CNH963	22	1	12	22
<i>Salinispora arenicola</i> CNX814	22	1	12	21
<i>Salinispora arenicola</i> CNY486	22	1	13	24
<i>Salinispora arenicola</i> CNX508	21	1	12	21
<i>Salinispora arenicola</i> CNX482	21	1	12	21
<i>Salinispora pacifica</i> CNS996	21	1	15	21
<i>Salinispora pacifica</i> CNS237	20	1	12	19
<i>Salinispora pacifica</i> CNY646	20	1	13	19
<i>Salinispora tropica</i> CNT261	20	2	10	18
<i>Salinispora pacifica</i> DSM 45548	19		7	10
<i>Salinispora pacifica</i> CNT045	19	1	13	19
<i>Salinispora pacifica</i> CNT124	19		13	19
<i>Salinispora pacifica</i> DSM 45543	19		12	18
<i>Salinispora tropica</i> CNB536	19		11	19
<i>Salinispora tropica</i> CNH898	18	1	11	20
<i>Salinispora pacifica</i> CNT403	18	1	12	17
<i>Salinispora pacifica</i> CNS860	18	2	11	16
<i>Salinispora pacifica</i> CNS863	18	2	12	17
<i>Salinispora tropica</i> CNY012	18	2	10	18
<i>Salinispora pacifica</i> CNT584	17		11	17
<i>Salinispora pacifica</i> DSM 45549	17	1	11	16
<i>Salinispora pacifica</i> CNR114	17	1	13	17
<i>Salinispora tropica</i> CNR699	17	2	10	16
<i>Salinispora pacifica</i> CNT854	18		13	18

<i>Salinispora pacifica</i> CNT150	17	1	11	16
<i>Salinispora pacifica</i> CNT131	17	1	11	15
<i>Salinispora pacifica</i> DSM 45544	16		11	16
<i>Salinispora pacifica</i> CNT003	16	1	10	15
<i>Salinispora tropica</i> CNY681	16	1	10	16
<i>Salinispora tropica</i> CNS197	16	1	10	16
<i>Salinispora tropica</i> CNY678	16	1	10	16
<i>Salinispora tropica</i> CNT250	16	1	10	16
<i>Salinispora tropica</i> CNB-440	16	1	10	16
<i>Salinispora tropica</i> CNS416	15		9	15
<i>Salinispora pacifica</i> CNT001	15	1	11	15
<i>Salinispora pacifica</i> CNY498	15	1	11	15
<i>Salinispora pacifica</i> CNR909	15	1	10	15
<i>Salinispora tropica</i> CNB476	15	1	9	15
<i>Salinispora pacifica</i> CNR894	15		11	15
<i>Salinispora pacifica</i> CNY363	15		11	15
<i>Salinispora pacifica</i> CNS055	15		9	15
<i>Salinispora pacifica</i> CNT603	15	1	11	15
<i>Salinispora pacifica</i> CNT138	14		10	14
<i>Salinispora pacifica</i> DSM 45547	14	1	10	14
<i>Salinispora pacifica</i> CNH732	14	1	10	14
<i>Salinispora pacifica</i> CNY703	14		9	13
<i>Salinispora pacifica</i> CNQ768	14	1	10	14
<i>Salinispora pacifica</i> CNY673	14		10	14
<i>Salinispora pacifica</i> CNT855	14		9	14
<i>Salinispora pacifica</i> CNY239	14	1	10	14
<i>Salinispora pacifica</i> CNR942	14	1	10	14
<i>Salinispora pacifica</i> DSM 45546	14	1	10	14
<i>Salinispora pacifica</i> CNT609	14	1	10	14
<i>Salinispora pacifica</i> CNY331	14		10	14
<i>Salinispora tropica</i> CNR416	14		9	14
<i>Salinispora pacifica</i> CNY330	13		9	13

<i>Salinispora pacifica</i> CNT851	13	1	9	13
<i>Salinispora pacifica</i> CNT796	13	1	9	13
<i>Salinispora pacifica</i> CNS103	13		9	13
<i>Salinispora pacifica</i> CNY202	13	1	9	13
<i>Salinispora pacifica</i> CNT133A	13		9	13
<i>Salinispora arenicola</i> CNY666	13	5	8	13
<i>Salinispora pacifica</i> CNT029	13	1	9	13
<i>Salinispora pacifica</i> CNT084	13		9	13
<i>Salinispora pacifica</i> CNR510	13		9	13
<i>Salinispora pacifica</i> CNT569	12	1	9	13
<i>Salinispora pacifica</i> CNT-133	11	11	7	9
<i>Salinispora pacifica</i> CNS801	10		7	10
<i>Salinispora pacifica</i> CNT148	10		7	10

#### 4.2. CYP015 and CYP107 families are bloomed in *Salinispora* species

Based on the International P450 Nomenclature Committee Rules (Nelson et al., 1993, Nelson, 2006, Nelson, 1998), all 2643 P450s can be grouped into 45 families and 103 subfamilies (Tables 4.1 and 4.3). Phylogenetic analysis revealed that large P450 families CYP105 and CYP107 were scattered across the evolutionary tree (Figure 4.1). Previously, this phenomenon was observed for these P450 families (Senate et al., 2019, Mnguni et al., 2020). Authors suggested that phylogenetic-based annotation of P450s could detect similarity cues beyond a simple percentage identity cutoff (Senate et al., 2019, Mnguni et al., 2020). Except for CYP105 and CYP107, the rest of the P450s are grouped as per their families (Figure 4.1). A point to be noted is that most of the P450s are orthologs considering the *Salinispora* species analyzed in this study are different strains of three species. Comparative analysis revealed that *Salinispora* species have the lowest number of P450 families and subfamilies compared to other actinomycetes such as *Streptomyces* and *Mycobacterium* (Table 4.1).

Among *Salinispora* species, *S. arenicola* CNY280 had the highest number of P450 families (18) and P450 subfamilies (32) in its genome (Table 4.2). This is quite an interesting observation where a species with the highest number of P450s also had the highest number of P450 families and subfamilies. This phenomenon was not found in other actinomycetes such

as *Streptomyces* (Mnguni et al., 2020) and *Mycobacterium* (Parvez et al., 2016, Senate et al., 2019). For example, in *Streptomyces* species, *Streptomyces albulus* ZPM had the highest number of P450s, but *Streptomyces rimosus rimosus* ATCC 10970, and *Streptomyces clavuligerus* had the highest number of P450 families and subfamilies, respectively (Mnguni et al., 2020). Among mycobacterial species, *Mycobacterium rhodesiae* NBB3 had the highest P450s and P450 families, but *M. marinum* had the highest P450 subfamilies (Parvez et al., 2016, Senate et al., 2019).

**Table 4. 3** Comparative analysis of P450 families and subfamilies in *Salinispora* species.

P450 Family	P450 Count	Percentage Count	Subfamily	Count	Percentage Count
CYP1004	34	1.29%	A	17	0.64
			B	17	0.64
CYP1005	127	4.81%	A	127	4.79
CYP1037	2	0.08%	B	2	0.08
CYP1051	60	2.27%	A	60	2.26
CYP1056	2	0.08%	B	2	0.08
CYP105	600	22.70%	AB	124	4.67
			AH	4	0.15
			B	1	0.04
			BL	78	2.94
			BN	1	0.04
			CH	44	1.66
			CN	62	2.34
			CP	62	2.34
			CT	41	1.55
			EJ	3	0.11
			G	62	2.34
			H	3	0.11
			J	52	1.96
			W	63	2.37

CYP107	551	20.85%	AW	65	2.45
			AX	75	2.83
			AY	116	4.37
			CL	3	0.11
			CT	6	0.23
			E	38	1.43
			EP	2	0.08
			EU	44	1.66
			FH	25	0.94
			FJ	20	0.75
			FS	61	2.30
			GU	1	0.04
			HF	2	0.08
			LA	6	0.23
			N	2	0.08
			NE	2	0.08
			NF	2	0.08
			NG	4	0.15
			NH	8	0.30
			Q	63	2.37
			Z	6	0.23
CYP1114	1	0.04%	C	1	0.04
CYP113	24	0.91%	B	6	0.23
			D	1	0.04
			E	10	0.38
			R	2	0.08
			S	2	0.08
			T	1	0.04
			X	2	0.08
CYP1197	1	0.04%	A	1	0.04
CYP1198	43	1.63%	B	43	1.62
CYP1207	4	0.15%	A	4	0.15

CYP1223	6	0.23%	D	2	0.08
			A	4	0.15
CYP1226	2	0.08%	A	2	0.08
CYP124	15	0.57%	M	15	0.57
CYP125	164	6.21%	A	128	4.82
			G	36	1.36
CYP1269	2	0.08%	A	2	0.08
CYP1278	11	0.42%	A	5	0.19
			B	6	0.23
CYP1437	1	0.04%	C	1	0.04
CYP146	1	0.04%	A	1	0.04
CYP1522	1	0.04%	A	1	0.04
CYP154	155	5.86%	AJ	4	0.15
			J	1	0.04
			M	150	5.65
CYP1611	1	0.04%	B	1	0.04
CYP161	28	1.06%	N	23	0.87
		0.00%	T	5	0.19
CYP162	39	1.48%	A	11	0.41
			B	2	0.08
			G	2	0.08
			H	1	0.04
			J	1	0.04
			K	1	0.04
			L	1	0.04
			M	1	0.04
			N	1	0.04
			P	18	0.68
CYP163	39	1.48%	A	2	0.08
			B	37	1.39
CYP164	4	0.15%	C	4	0.15
CYP166	62	2.35%	A	62	2.34

CYP173	1	0.04%	K	1	0.04
CYP1902	2	0.08%	A	2	0.08
CYP2054	22	0.83%	A	22	0.83
CYP205	1	0.04%	A	1	0.04
CYP208	126	4.77%	A	126	4.75
CYP2091	1	0.04%	A	1	0.04
CYP2098	2	0.08%	A	2	0.08
CYP211	225	8.51%	B	124	4.67
			C	101	3.81
CYP2296	1	0.04%	A	1	0.04
CYP244	107	4.05%	A	107	4.03
CYP245	83	3.14%	A	83	3.13
CYP247	21	0.79%	A	21	0.79
CYP248	63	2.38%	A	63	2.37
CYP2611	1	0.04%	B	1	0.04
CYP283	1	0.04%	A	1	0.04
CYP285	4	0.15%	A	2	0.08
			D	2	0.08
CYP294A4	2	0.08%	A	2	0.08

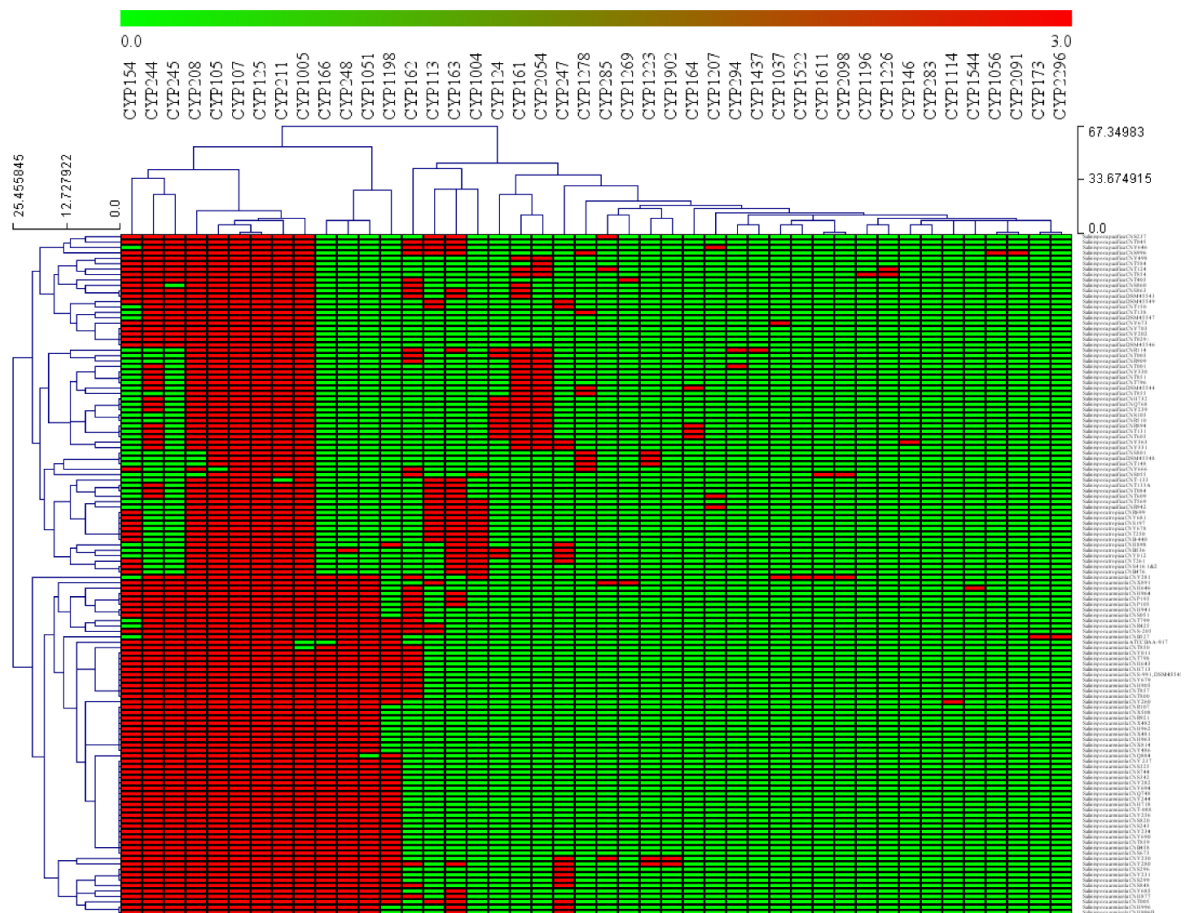
Analysis of P450 families and subfamilies suggested that P450s in *Salinispora* species bloomed (presence of more copies of the same P450 family in a species by duplication of an ancestral gene) (Table 4.3). Among P450 families, the CYP105 was dominant with 600 members, followed by CYP107 with 551 members, CYP211 with 225 members, CYP125 with 164 members, CYP154 with 155 members, CYP1005 with 127 members, and CYP208 with 126 members (Table 4.3). These P450 families contributed more than 70% to the total P450s (Table 4.3). This indicates that P450 families such as CYP105, CYP107, CYP211, CYP125, and CYP154 are bloomed, whereas CYP1005 and CYP208 families are expanded in these species. Comparing the dominant P450 families revealed that CYP105 is prevalent only in *Salinispora* species (Table 4.1), where this family was second most dominant in *Streptomyces* species (Table 4.1). Interestingly, the second most dominant P450 family of *Salinispora* species, CYP107, was dominant in species belonging to bacterial groups *Streptomyces*, *Firmicutes*, and *Gammaproteobacteria* (Table 4.1). The blooming was also

observed at the subfamily level, indicating these P450s are preferred by *Salinispora* species for a particular reason. For example, subfamily AB was dominant with 124 members in CYP105; Subfamily AY was dominant with 116 members in CYP107, subfamily A was dominant with 128 members in CYP125, Subfamily M was dominant with 150 members, subfamily A was dominant with 126 members in CYP208, and Subfamily B dominant with 124 members in CYP211 (Table 4.3). Due to the blooming of specific P450s at the family level, *Salinispora* species had the lowest P450 diversity percentage, the same as *Firmicutes* species (Table 4.1). The blooming or expansion of P450s is a common phenomenon in organisms and is observed in other bacterial species (Table 4.2). It has been hypothesized that species enrich specific P450s in their genomes that are beneficial to them, particularly to adapt to ecological niches (Senate et al., 2019, Mnguni et al., 2020).

#### **4.3. CYP107 and CYP125 are conserved in *Salinispora* species**

P450 family conservation analysis revealed that CYP107 and CYP125 families are conserved in 126 *Salinispora* species (Figure 4.2). Except for a few species, CYP208 (4 species), CYP105 (one species), CYP211 (one species), and CYP1005 (2 species), the rest of the *Salinispora* species have these families (Figure 4.2). In addition to this, P450 families such as CYP154, CYP244, CYP245, CYP166, CYP248, and CYP1056 are co-present in many species (Figure 4.2). This suggests a prominent role of these P450 families in these species, possibly in secondary metabolism as observed in other bacterial species (Senate et al., 2019, Mnguni et al., 2020, Padayachee et al., 2020). Conservation or co-presence of specific P450s in other bacterial species was also reported. The CYP107 family is conserved in all 203 *Streptomyces* species, and P450 families such as CYP156, CYP105, CYP154, and CYP157 are also present in the majority of the *Streptomyces* species (Mnguni et al., 2020). Ten P450 families, CYP51, CYP123, CYP125, CYP130, CYP135, CYP136, CYP138, CYP140, CYP144, and CYP1128, were conserved in mycobacterial species (Parvez et al., 2016). Analysis of conservation of P450 families in 229 *Firmicutes* species and 114 cyanobacterial species revealed no conservation of the P450 family (Padayachee et al., 2020, Khumalo et al., 2020). Still, some of the P450 families were co-present in most of the species. The P450 families CYP152, CYP107, CYP012, and CYP109, were found to be a co-presence in most *Firmicutes* species (Padayachee et al., 2020), and the P450 families CYP110 and CYP120 were found to be a co-presence in most cyanobacterial species (Khumalo et al., 2020).

If a P450 family is conserved or few P450 families are co-presence, these families play an important role in a species's primary or secondary metabolism. Previous studies showed that this type of P450s prominently plays a role in secondary metabolism, helping species adapt to diverse ecological niches (Senate et al., 2019, Parvez et al., 2016, Syed et al., 2019, Padayachee et al., 2020, Mnguni et al., 2020). The importance of P450 families that are conserved and co-presence in *Salinispora* species is discussed in detail in the next section.



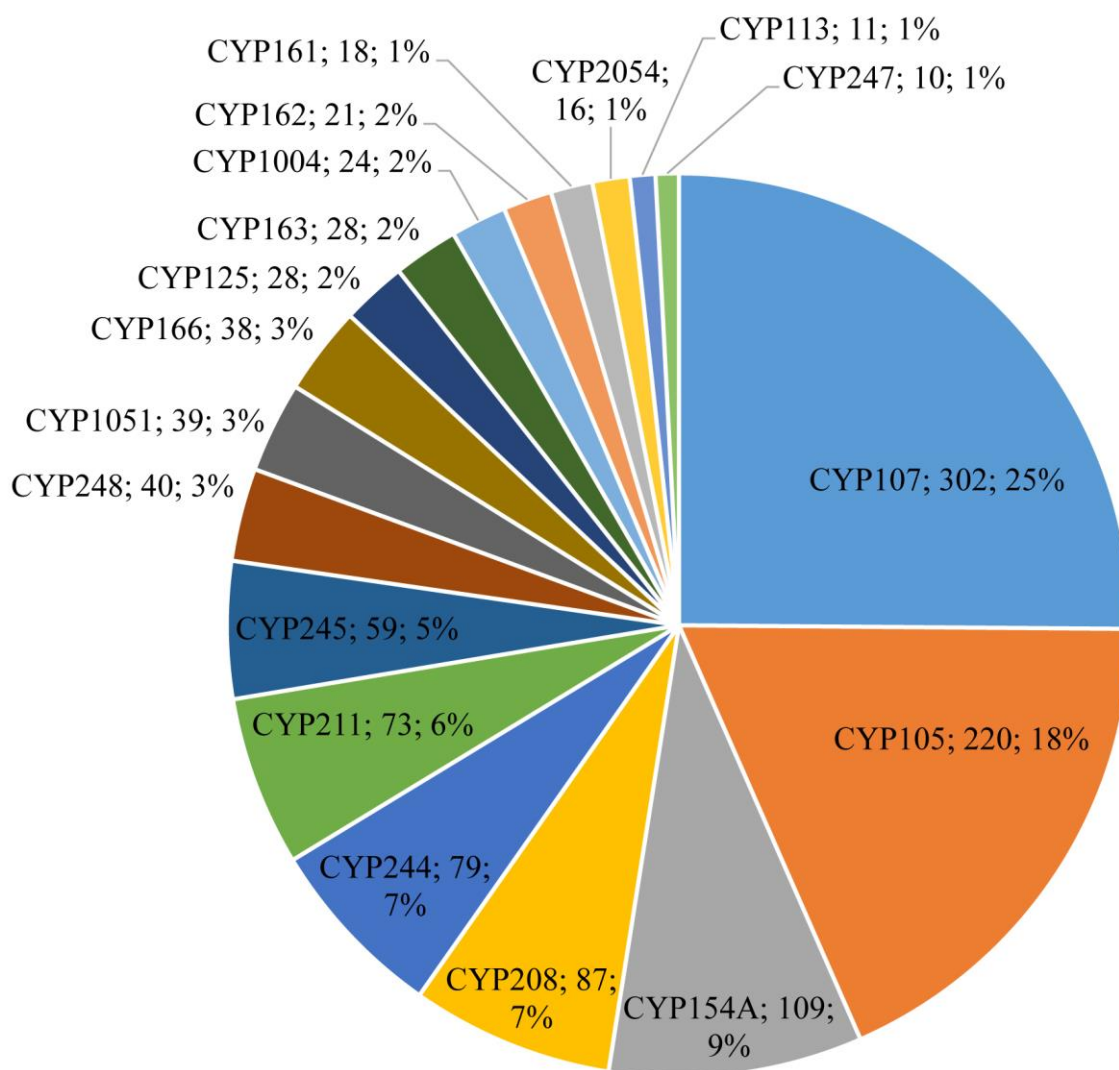
**Figure 4. 2** Heat-map of P450 family conservation or co-presence analysis in *Salinispora* species. In the heat-map, the presence and absence of P450 families are indicated in red and

green colors. The horizontal axis represents P450 families, and the vertical axis represents *Salinispora* species.

#### **4.4. Unprecedented number of P450s involved in smBGCs**

Analysis of the P450s part of smBGCs revealed that many P450s (47%) are part of these clusters, indicating their involvement in producing different secondary metabolites in *Salinispora* species (Tables 4.4 and S1). The percentage of P450s part of smBGCs in *Salinispora* species was found to be unprecedented compared to other bacterial species, including other actinomycetes *Streptomyces* species and mycobacterial species that had 30% and 27% of P450s as part of smBGCs (Table 4.1). This suggests that *Salinispora* species dedicated half of their P450s to the production of secondary metabolites.

Among 2643 P450s, 1236 P450s belonging to the 35 P450 families were part of smBGCs (Figure 4.3 and Tables 4.4 and S1). This means almost 78% of P450 families of *Salinispora* species are involved in secondary metabolism. Among the families that are part of smBGCs, CYP107 is dominant with 302 members (25%), followed by CYP105 with 220 members (18%), CYP208 with 87 members (7%), CYP244 with 79 members (7%), and CYP211 with 73 members (6%) (Figure 4.3 and Table S1). Analysis of the P450s part of smBGCs revealed a strong correlation between the dominant P450 families (Table 4.3) being dominant in smBGCs (Figure 4.3). This suggests that *Salinispora* species are enriched by blooming or expanding these P450 families (as discussed in the previous section) in their genome to produce secondary metabolites.



**Figure 4. 3** Comparative analysis of P450s associated with secondary metabolism in *Salinispora* species. The P450 family name, number of P450s, and the percentage of the total number of P450s that are part of secondary metabolite biosynthetic gene clusters (smBGCs) are presented in the figure. Detailed information on secondary metabolite clusters, species, and P450s are shown in Table S1.

**Table 4. 4** Secondary metabolite biosynthetic gene cluster (smBGC) types and P450s are part of the cluster in *Salinispora* species. smBGC types were again classified into different varieties based on the P450s. The smBGCs type count and the total number of P450s in the cluster variety are also presented. The same smBGCs type names listed in the antibiotics & secondary metabolite analysis shell (anti-SMASH) database (Blin et al., 2021) were used in

the table. Detailed information on secondary metabolite clusters, species, and P450s are shown in Table S1.

smBGC type	SmBGC type count	smBGC type variety	P450s	P450 Count
Bacteriocin	47	46	CYP107AW	46
		1	CYP283A	1
betalactone	2	1	CYP162A6,CYP107HF1	2
		1	CYP113S1	1
butyrolactone	1	1	CYP105CT1,CYP154M5	2
Indole	54	51	CYP244A,CYP245A	102
		3	CYP244A	3
ladderane	18	4	CYP154M15,CYP125G6,CYP107FS2,CYP105CN1,CYP105CP2	20
		8	CYP107AX-fragment	8
		6	CYP107AX	6
lanthipeptide	2	1	CYP1223A5	1
		1	CYP105CP2,CYP105CN1,CYP107FS2,CYP248A2,CYP105W2	5
LAP	1	1	CYP154AJ2	1
lipolanthine	2	2	CYP1223A5	2
NRPS	205	1	CYP1004B1,CYP1004A1	2
		8	CYP1004B,CYP1004A,CYP125G	24
		1	CYP105CH2-fragment,CYP105CH1-fragment	2
		1	CYP105CN1	1

		1	CYP105CN1,CYP105CP2	2
		1	CYP105CN1,CYP107FS2,CYP125G6,CYP154M15	4
		1	CYP105CN1,CYP107FS2,CYP247A7	3
		1	CYP105CP2	1
		7	CYP105CP2,CYP105CN1,CYP107FS2	21
		10	CYP105CP2,CYP105CN1,CYP107FS2,CYP125G6,CYP154M15	50
		2	CYP105CP2,CYP105CN1,CYP107FS2,CYP248A2	8
		1	CYP105CP2,CYP105CN1,CYP107FS2,CYP248A2,CYP105W2	5
		3	CYP105W	3
		38	CYP107AY	38
		1	CYP107AY14,CYP244A-fragment2	2
		6	CYP107AY2,CYP105CT1,CYP154M5	18
		1	CYP107AY2,CYP163B16	2
		1	CYP107AY7,CYP244A5,CYP245A11	3
		1	CYP107AY9,CYP162B3	2
		1	CYP107AY9,CYP244A10	2
		2	CYP107CL2,CYP1056B2	4
		5	CYP107CT3	5
		1	CYP107CT3,CYP107AY7	2
		1	CYP107FS2	1
		6	CYP107FS2,CYP105CN1,CYP105CP2	18
		2	CYP107NH1,CYP247A8,CYP107Z27	6
		1	CYP107Z27,CYP247A8,CYP107NH1	3
		1	CYP113D13,CYP163B22	2
		1	CYP1196A2	1
		1	CYP1198B1	1
		1	CYP1198B1,CYP107AY2	2
		2	CYP1207A12	2
		1	CYP125G1,CYP1004A1,CYP1004B1	3
		1	CYP125G6,CYP154M15	2

		4	CYP1278A4	4
		1	CYP1437C1	1
		1	CYP154AJ3	1
		1	CYP154J2,CYP244A5,CYP245A11	3
		5	CYP154M1,CYP208A4	10
		6	CYP154M	6
		3	CYP154M,CYP208A	6
		1	CYP154M16,CYP211C6	2
		1	CYP154M21,CYP154M13	2
		1	CYP154M21,CYP154M13,CYP105W2,CYP248A2	4
		3	CYP154M21,CYP154M13,CYP105W2,CYP248A2,CYP154M20	15
		1	CYP154M21,CYP154M13,CYP105W2,CYP248A2,CYP154M20,CYP162P1	6
		12	CYP162	12
		1	CYP163A10,CYP162K1	2
		15	CYP163B	15
		3	CYP164C2	3
		5	CYP208A21,CYP154M16	10
		2	CYP208A4,CYP154M1	4
		8	CYP244A,CYP107AY	16
		5	CYP244A5,CYP245A11	10
		2	CYP244A,CYP107AY	4
		2	CYP244A	2
		1	CYP245A11	1
		3	CYP247A	3
		1	CYP247A8,CYP107NH1	2
		1	CYP247A8,CYP107Z27	2
		1	CYP248A2	1
		1	CYP248A2,CYP105W2	2
		1	CYP285D2	1
NRPS-like	30	5	CYP107EU	5

		1	CYP107EU1,CYP1198B1,CYP105CH1	3
		1	CYP107FH3,CYP161N4,CYP107AY9	3
		7	CYP107FH3,CYP2054A3,CYP161N4	21
		2	CYP161N4,CYP2054A3,CYP107FH3	6
		6	CYP162A8	6
		6	CYP166A4	6
		1	CYP166A4,CYP107Q4,CYP105G5	3
		1	CYP285A9-fragment,CYP285A9-fragment	2
oligosacch aride	35	1	CYP105CP2	1
		1	CYP105W2,CYP107FS2,CYP105CN1,CYP105CP2	4
		1	CYP105W2,CYP107NH1	2
		1	CYP105W2,CYP154M20,CYP154M13,CYP154M21, CYP248A2	5
		5	CYP105W2,CYP248A2	10
		1	CYP105W2,CYP248A2,CYP107FS2	3
		9	CYP105W2/3,CYP248A2,CYP107FS2,CYP105CN1, CYP105CP2	45
		1	CYP1269A2	1
		8	CYP154M20,CYP248A2,CYP105W2,CYP154M13,C YP154M21	40
		1	CYP2091A1	1
		3	CYP248A2	3
		3	CYP248A2,CYP105W2/3	6
other	4	2	CYP247A7	2
		2	CYP105AH4	2
		1	CYP1004A3,CYP1004B4,CYP113E2,CYP163B18	4
T1PKS	223	1	CYP105AH4	1
		1	CYP105BN4	1
		17	CYP105CH1/2	17
		1	CYP105CN1	1
		4	CYP105G5	4

		16	CYP105G5,CYP107Q4	32
		2	CYP105H11	2
		1	CYP107AY13	1
		24	CYP107E	24
		1	CYP107E3,CYP125G1,CYP1004A1,CYP1004B1	4
		8	CYP107EU1	8
		2	CYP107FH4	2
		1	CYP107NE1	1
		3	CYP107Q4	3
		17	CYP107Q4,CYP105G5	34
		6	CYP113E1/2	6
		1	CYP113E2,CYP107EP2	2
		2	CYP1198B2	2
		1	CYP125G1	1
		1	CYP1278B-fragment2	1
		5	CYP154M5,CYP105CT1	10
		1	CYP154M5,CYP105CT1,CYP105G5,CYP105CP2	4
		1	CYP154M5,CYP105CT1,CYP107AY2-fragment	3
		1	CYP154M5,CYP105CT2	2
		1	CYP1611B1,CYP2098A1	2
		29	CYP166A4	29
		1	CYP166A4,CYP107Q4,CYP105G5	3
		70	CYP208A	70
		1	CYP208A28,CYP154M18	2
		1	CYP211C5	1
		2	CYP294A4	2
T2PKS	76	2	CYP107NG1	2
		1	CYP107NH1	1
		1	CYP125G4	1
		1	CYP161T1	1
		69	CYP211C	69
		1	CYP2296A2,CYP166A4,CYP173K1	3

		1	CYP244A5,CYP211C6	2
T3PKS	8	7	CYP161N4,CYP2054A3,CYP107FH3	21
		1	CYP107FH3	1
Terpene	61	39	CYP1051A	39
		2	CYP105CT1	2
		7	CYP105CT1,CYP154M5	14
		6	CYP107AY	6
		4	CYP107AY9,CYP244A10	8
		1	CYP107E37	1
		1	CYP154AJ2	1
		1	CYP154M5	1
transAT- PKS	1	1	CYP113X1	1
transAT- PKS-like	8	8	CYP163B	8

Analysis of P450 smBGCs revealed the presence of 18 types (Tables 4.4 and S2). Among the types, Type I PKS (Polyketide synthase) (T1PKS) was dominant with 223 clusters, followed by nonribosomal peptides (NRPS) (205 clusters) and Type II PKS (T2PKS) (76 clusters) (Tables 4.4 and S1). This suggests that most of the secondary me-tabolites produced by P450 smBGCs are T1PKS. When the P450 smBGCs were further analyzed for the number of P450s and P450 families, the dominant BGC type was not found to be dominant concerning the number of P450s being part of that smBGC type (Tables 4.4 and S1). NRPS had the highest number of P450s (395 P450s), followed by T1PKS (275 P450s), oligosaccharide (121 P450s), and indole (105 P450s) (Tables 4.4 and S1). The difference being not having more P450s despite being dominant smBGCs such as T1PKS is that the other smBGCs have more P450s per se more than one P450 being part of that type (Tables 4.4 and S1). This phenomenon of more than one P450 being part of smBGCs has been reported earlier in other bacterial species (Ngcobo et al., 2020). However, having up to 6 P450s as part of smBGCs is unprecedented (Table 4.4), suggest-ing these clusters produce diverse secondary metabolites. The P450s co-present in different *Salinispora* species were part of the same cluster (Table 4.4). Based on the arrange-ment of P450s concerning their family/subfamily and the number of P450s in smBGCs, it is clear that these smBGCs are orthologs (Table 4.4). These smBGCs

are passed into different *Salinispora* species from a single ancestor before diverging into *S. arenicola*, *S. pacifica*, and *S. tropica*.

#### **4.5. Functional prediction of *Salinispora* species P450s**

Most of the *Salinispora* species P450s are orphans without an assigned biological function. Based on the homolog P450s from other organisms and being part of smBGCs, some P450 functions can be predicted. CYP105 and CYP107 members are involved in the degradation/biotransformation of xenobiotics and biosynthesis of secondary metabolites (Weber et al., 1991, Trefzer et al., 2007, Fujii et al., 2009, Prior et al., 2010, Moody and Loveridge, 2014). CYP107 from *S. arenicola* CNS-205 is involved in secondary metabolite biosynthesis (Wilson et al., 2010). It catalyzes multiple oxidative rearrangement reactions in the biosynthesis of saliniketal and rifampin (Wilson et al., 2010). CYP105 and CYP107 members' enzymatic functions could help *Salinispora* species utilize diverse compounds as carbon sources, detoxify toxic compounds, or kill other bacterial species to thrive in the environment. It is no doubt that due to these beneficial properties, *Salinispora* species enriched these family members in their genomes. CYP125 members conserved in *Salinispora* species are cholesterol and cholest-4-en-3-one hydroxylases (McLean et al., 2009, Ouellet et al., 2010). One can assume that CYP125 members possibly help *Salinispora* species utilize cholesterol or cholesterol-like molecules as carbon sources. Growth of *S. arenicola* CNS-205 on cholesterol where complete degradation of cholesterol was observed (Bergstrand et al., 2016) strongly supports this assumption considering these species do have CYP125 in their genome.

Interestingly, the presence of CYP125 members as part of smBGCs as observed in *Salinispora* species (Table 4.4) is also observed in mycobacterial species (Ngcobo et al., 2020), indicating CYP125 members do have other functions apart from cholesterol oxidation. CYP146 members are involved in  $\beta$ -hydroxytyrosine formation, a precursor for the biosynthesis of vancomycin antibiotics (Cryle and Schlichting, 2008). Interestingly, only a single member was found in *Salinispora* species (Table 4.3) and is not part of smBGCs, complicating predicting its role in these species.

CYP154 members are involved in regio- and stereo-selective hydroxylation of different steroids (Bracco et al., 2013, Subedi et al., 2021). CYP154 from *Nocardia farcinica* IFM10152 is a bifunctional enzyme with O-dealkylation and *ortho*-hydroxylation activities (Choi et al., 2010). This P450 converts formononetin, an isoflavone compound, into ortho-

dihydroxy-isoflavone (Choi et al., 2010). In *Salinispora* species, CYP154 members are dominant, indicating they may attribute the above-said activities to these species. However, the role of CYP154 in the generation of secondary metabolites and these compounds' properties concerning *Salinispora* species is of future interest (Figure 4.3 and Table 4.4).

CYP163A and CYP163B members produce novobiocin, aminocoumarin antibiotic (Chen and Walsh, 2001), and skyllamycin, a potent inhibitor of the platelet-derived growth factor (Uhlmann et al., 2013). CYP162A members are involved in peptidyl nucleoside antibiotic nikkomycin synthesis (Lauer et al., 2000, Xie et al., 2007). CYP161A members are involved in the biosynthesis of antibiotics, pimaricin (Mendes et al., 2005), and amphotericin (Caffrey et al., 2001). CYP113 members are involved in the production a variety of antibiotics erythromycin (Shafiee and Hutchinson, 1987, Stassi et al., 1993), tylosin (Merson-Davies and Cundliffe, 1994, Fouces et al., 1999) and himastatin (Zhang et al., 2013, Ma et al., 2011). The presence of the CYP161-CYP163 and CYP113 members as part of smBGCs in *Salinispora* species (Figure 4.3 and Table 4.4) suggests that these members are certainly involved in the production of secondary metabolites in these species.

CYP244 and CYP245 members are involved in the biosynthesis of antibiotic rapamycin (Aparicio et al., 1996, Molnar et al., 1996). These two P450s together as part of smBGCs clusters in *Salinispora* species (Table 4.4) indicate they are working together in producing secondary metabolite. CYP248A members are involved in the production of antibiotic aureothin (Zocher et al., 2011). *Salinispora* species have 63 CYP248A members (Table 4.3), and 40 of them are part of smBGC (Figure 4.3 and Table 4.4), indicating their prominent role in secondary metabolites production. CYP124 members are known for their terminal hydroxylation of methyl branched-lipids in *M. tuberculosis* (Johnston et al., 2009). None of these members were found as part of smBGCs in *Salinispora* species (Table 4.4), indicating their limited role possibly in the oxidation of different methylated-aliphatic lipids in these species.

It is evident from the data presented in this article that close to half of *Salinispora* species P450s (1236 P450s) are part of smBGCs. Thus, we predict that these P450s play a role in producing different secondary metabolites characteristic of smBGC types (Tables 4.4 and S2). The detailed information on species name, list of P450s part of smBGCs, their cluster information, and BGC type is presented in Table S1.

## CHAPTER 5: CONCLUSIONS AND FUTURE PERSPECTIVES

*Salinispora* species being marine organisms within the phylum *Actinomycetes*, are considered model organisms for studying bacterial diversity and secondary metabolite production. Compared to the genera *Streptomyces* and *Mycobacterium*, the genus *Salinispora* has an unprecedented number of P450s as part of secondary metabolite bio-synthetic gene clusters (smBGCs), indicating a great diversity of secondary metabolites produced by these species. The presence of up to six P450s as part of smBGCs is unusual and not observed in other bacterial species. Future functional characterization of P450s sheds more light on the untapped secondary metabolite biotechnological potentials from *Salinispora* species. Based on the data presented in this article and the literature published on P450s function, we predict that *Salinispora* species enriched or expanded specific P450s in their genome to utilize

diverse compounds as carbon sources to detoxify toxic compounds or kill other bacterial species to thrive in the environment.

## REFERENCES

- ABEGAZ, B. M. & KINFE, H. H. 2019. Secondary metabolites, their structural diversity, bioactivity, and ecological functions: An overview. *Physical Sciences Reviews*, 4.
- AHMED, L., JENSEN, P. R., FREEL, K. C., BROWN, R., JONES, A. L., KIM, B.-Y. & GOODFELLOW, M. 2013. *Salinispora pacifica* sp. nov., an actinomycete from marine sediments. *Antonie Van Leeuwenhoek*, 103, 1069-1078.
- ANDERSEN, J. F., TATSUTA, K., GUNJI, H., ISHIYAMA, T. & HUTCHINSON, C. R. 1993. Substrate specificity of 6-deoxyerythronolide B hydroxylase, a bacterial cytochrome P450 of erythromycin A biosynthesis. *Biochemistry*, 32, 1905-1913.
- ASOLKAR, R. N., KIRKLAND, T. N., JENSEN, P. R. & FENICAL, W. 2010. Arenimycin, an antibiotic effective against rifampin- and methicillin-resistant *Staphylococcus aureus* from the marine actinomycete *Salinispora arenicola*. *J Antibiot (Tokyo)*, 63, 37-9.
- BAK, S., BEISSON, F., BISHOP, G., HAMBERGER, B., HOFER, R., PAQUETTE, S. & WERCK-REICHHART, D. 2011. Cytochromes p450. *Arabidopsis Book*, 9, e0144.

- BARKA, E. A., VATSA, P., SANCHEZ, L., GAVEAU-VAILLANT, N., JACQUARD, C., KLENK, H.-P., CLÉMENT, C., OUHDOUCH, Y. & VAN WEZEL, G. P. 2016. Taxonomy, physiology, and natural products of Actinobacteria. *Microbiology and Molecular Biology Reviews*, 80, 1-43.
- BERDY, J. 2005. Bioactive microbial metabolites. *The Journal of antibiotics*, 58, 1-26.
- BERNHARDT, R. 2006. Cytochromes P450 as versatile biocatalysts. *J Biotechnol*, 124, 128-45.
- BERNHARDT, R. & URLACHER, V. B. 2014. Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Applied microbiology and biotechnology*, 98, 6185-6203.
- BISCHOFF, D., BISTER, B., BERTAZZO, M., PFEIFER, V., STEGMANN, E., NICHOLSON, G. J., KELLER, S., PELZER, S., WOHLLEBEN, W. & SÜSSMUTH, R. D. 2005. The biosynthesis of vancomycin-type glycopeptide antibiotics--a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases. *ChemBiochem*, 6, 267-72.
- BONET, B., TEUFEL, R., CRUSEMANN, M., ZIEMERT, N. & MOORE, B. S. 2015. Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin. *J Nat Prod*, 78, 539-42.
- BOSE, U., HEWAVITHARANA, A. K., NG, Y. K., SHAW, P. N., FUERST, J. A. & HODSON, M. P. 2015. LC-MS-based metabolomics study of marine bacterial secondary metabolite and antibiotic production in *Salinispora arenicola*. *Mar Drugs*, 13, 249-66.
- BOSE, U., HEWAVITHARANA, A. K., VIDGEN, M. E., NG, Y. K., SHAW, P. N., FUERST, J. A. & HODSON, M. P. 2014. Discovering the recondite secondary metabolome spectrum of *Salinispora* species: a study of inter-species diversity. *PLoS One*, 9, e91488.
- BOSE, U., ORTORI, C. A., SARMA, S., BARRETT, D. A., HEWAVITHARANA, A. K., HODSON, M. P., FUERST, J. A., SHAW, P. N. & BODEN, R. 2017. Production of N-acyl homoserine lactones by the sponge-associated marine actinobacteria *Salinispora arenicola* and *Salinispora pacifica*. *FEMS Microbiol Lett*, 364.
- CHIGORIMBO-MUREFU, N., NJOROGE, M., LOUW, S., MUGUMBATE, G. & CHIBALE, K. 2013. Drug Metabolite Generation Using a Laboratory Evolved NADPH Independent Cytochrome P450: Application of in vitro and in silico Approaches. *Drug Metabolism Letters*, 7, 68-77.
- CHOI, S.-S., KIM, H.-J., LEE, H.-S., KIM, P. & KIM, E.-S. 2015. Genome mining of rare actinomycetes and cryptic pathway awakening. *Process Biochemistry*, 50, 1184-1193.
- CIMERMANCIC, P., MEDEMA, M. H., CLAESEN, J., KURITA, K., BROWN, L. C. W., MAVROMMATIS, K., PATI, A., GODFREY, P. A., KOEHRSEN, M. & CLARDY, J. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 158, 412-421.
- CONTADOR, C. A., RODRÍGUEZ, V., ANDREWS, B. A. & ASENJO, J. A. 2019. Use of genome-scale models to get new insights into the marine actinomycete genus *Salinispora*. *BMC systems biology*, 13, 1-14.
- DE LIMA PROCÓPIO, R. E., DA SILVA, I. R., MARTINS, M. K., DE AZEVEDO, J. L. & DE ARAÚJO, J. M. 2012. Antibiotics produced by *Streptomyces*. *The Brazilian Journal of infectious diseases*, 16, 466-471.
- DEBNATH, A., CALVET, C. M., JENNINGS, G., ZHOU, W., AKSENOV, A., LUTH, M. R., ABAGYAN, R., NES, W. D., MCKERROW, J. H. & PODUST, L. M. 2017. CYP51 is an essential drug target for the treatment of primary amoebic meningoencephalitis (PAM). *PLoS Negl Trop Dis*, 11, e0006104.

- DEMAIN, A. L. & FANG, A. 2000. The natural functions of secondary metabolites. *History of modern biotechnology I*. Springer.
- DODANI, S. C., CAHN, J. K., HEINISCH, T., BRINKMANN-CHEN, S., MCINTOSH, J. A. & ARNOLD, F. H. 2014. Structural, functional, and spectroscopic characterization of the substrate scope of the novel nitrating cytochrome P450 TxtE. *Chembiochem*, 15, 2259-67.
- DUNCAN, K. R., CRUSEMANN, M., LECHNER, A., SARKAR, A., LI, J., ZIEMERT, N., WANG, M., BANDEIRA, N., MOORE, B. S., DORRESTEIN, P. C. & JENSEN, P. R. 2015. Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chem Biol*, 22, 460-471.
- ESTEVEZ, F., RUEFF, J. & KRANENDONK, M. 2021. The central role of cytochrome P450 in xenobiotic metabolism—A brief review on a fascinating enzyme family. *Journal of Xenobiotics*, 11, 94-114.
- EUSTAQUIO, A. S., NAM, S. J., PENN, K., LECHNER, A., WILSON, M. C., FENICAL, W., JENSEN, P. R. & MOORE, B. S. 2011. The discovery of salinosporamide K from the marine bacterium "*Salinispora pacifica*" by genome mining gives insight into pathway evolution. *Chembiochem*, 12, 61-4.
- FENICAL, W., JENSEN, P. R., PALLADINO, M. A., LAM, K. S., LLOYD, G. K. & POTTS, B. C. 2009. Discovery and development of the anticancer agent salinosporamide A (NPI-0052). *Bioorganic & medicinal chemistry*, 17, 2175-2180.
- FIEDLER, H. P., BRUNTNER, C., BULL, A. T., WARD, A. C., GOODFELLOW, M., POTTERAT, O., PUDER, C. & MIHM, G. 2005. Marine actinomycetes as a source of novel secondary metabolites. *Antonie Van Leeuwenhoek*, 87, 37-42.
- FRASCH, H. J., MEDEMA, M. H., TAKANO, E. & BREITLING, R. 2013. Design-based re-engineering of biosynthetic gene clusters: plug-and-play in practice. *Curr Opin Biotechnol*, 24, 1144-50.
- GIRVAN, H. M. & MUNRO, A. W. 2016. Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Current opinion in chemical biology*, 31, 136-145.
- GREULE, A., STOK, J. E., DE VOSS, J. J. & CRYLE, M. J. 2018. Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat Prod Rep*, 35, 757-791.
- GRICMAN, Ł., WEISSENBORN, M. J., HOFFMANN, S. M., BORLINGHAUS, N., HAUER, B. & PLEISS, J. 2016. Redox Partner Interaction Sites in Cytochrome P450 Monooxygenases: In Silico Analysis and Experimental Validation. *ChemistrySelect*, 1, 1243-1251.
- GUENGERICH, F. P. 2020. A history of the roles of cytochrome P450 enzymes in the toxicity of drugs. *Toxicological research*, 1-23.
- GUENGERICH, F. P. & MUNRO, A. W. 2013. Unusual cytochrome p450 enzymes and reactions. *J Biol Chem*, 288, 17065-73.
- HASSAN, S. S., ANJUM, K., ABBAS, S. Q., AKHTER, N., SHAGUFTA, B. I., SHAH, S. A. & TASNEEM, U. 2017. Emerging biopharmaceuticals from marine actinobacteria. *Environ Toxicol Pharmacol*, 49, 34-47.
- HWANG, K. S., KIM, H. U., CHARUSANTI, P., PALSSON, B. O. & LEE, S. Y. 2014. Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. *Biotechnol Adv*, 32, 255-68.
- JAWALLAPERSAND, P., MASHELE, S. S., KOVACIC, L., STOJAN, J., KOMEL, R., PAKALA, S. B., KRASEVEC, N. & SYED, K. 2014. Cytochrome P450 monooxygenase CYP53 family in fungi: comparative structural and evolutionary

- analysis and its role as a common alternative anti-fungal drug target. *PLoS One*, 9, e107209.
- JENNEWEIN, S., PARK, H., DEJONG, J. M., LONG, R. M., BOLLON, A. P. & CROTEAU, R. B. 2005. Coexpression in yeast of *Taxus* cytochrome P450 reductase with cytochrome P450 oxygenases involved in Taxol biosynthesis. *Biotechnology and bioengineering*, 89, 588-598.
- JENSEN, P. R., CHAVARRIA, K. L., FENICAL, W., MOORE, B. S. & ZIEMERT, N. 2014. Challenges and triumphs to genomics-based natural product discovery. *J Ind Microbiol Biotechnol*, 41, 203-9.
- JENSEN, P. R. & MAFNAS, C. 2006. Biogeography of the marine actinomycete *Salinispora*. *Environmental Microbiology*, 8, 1881-1888.
- JENSEN, P. R., MOORE, B. S. & FENICAL, W. 2015. The marine actinomycete genus *Salinispora*: a model organism for secondary metabolite discovery. *Nat Prod Rep*, 32, 738-51.
- JENSEN, P. R., WILLIAMS, P. G., OH, D. C., ZEIGLER, L. & FENICAL, W. 2007. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl Environ Microbiol*, 73, 1146-52.
- JOSE, P. A., MAHARSHI, A. & JHA, B. 2021. Actinobacteria in natural products research: Progress and prospects. *Microbiological Research*, 246, 126708.
- KATZ, L. & BALTZ, R. H. 2016. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology and Biotechnology*, 43, 155-176.
- KELLY, S. L. & KELLY, D. E. 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos Trans R Soc Lond B Biol Sci*, 368, 20120476.
- KERSTEN, R. D., LANE, A. L., NETT, M., RICHTER, T. K., DUGGAN, B. M., DORRESTEIN, P. C. & MOORE, B. S. 2013. Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. *Chembiochem*, 14, 955-62.
- KHATER, S., ANAND, S. & MOHANTY, D. 2016. In silico methods for linking genes and secondary metabolites: The way forward. *Synth Syst Biotechnol*, 1, 80-88.
- KHUMALO, M. J. N., N.; PADAYACHEE, T.; CHEN, W.; YU, J.-H.; NELSON, D.; SYED, K. 2020. Comprehensive analyses of cytochrome P450 monooxygenases and secondary metabolite biosynthetic gene clusters in *Cyanobacteria*. *International journal of molecular sciences*, 21.
- KLINGENBERG, M. 1958. Pigments of rat liver microsomes. *Arch Biochem Biophys*, 75, 376-86.
- LAM, K. S. 2006. Discovery of novel metabolites from marine actinomycetes. *Curr Opin Microbiol*, 9, 245-51.
- LAMB, D. C., FOLLMER, A. H., GOLDSTONE, J. V., NELSON, D. R., WARRILOW, A. G., PRICE, C. L., TRUE, M. Y., KELLY, S. L., POULOS, T. L. & STEGEMAN, J. J. 2019. On the occurrence of cytochrome P450 in viruses. *Proceedings of the National Academy of Sciences*, 116, 12343-12352.
- LAMB, D. C. & WATERMAN, M. R. 2013. Unusual properties of the cytochrome P450 superfamily. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 20120434.
- LAMB, D. C., WATERMAN, M. R., KELLY, S. L. & GUENGERICH, F. P. 2007. Cytochromes P450 and drug discovery. *Curr Opin Biotechnol*, 18, 504-12.
- LEPESHEVA, G. I., FRIGGERI, L. & WATERMAN, M. R. 2018. CYP51 as drug targets for fungi and protozoan parasites: past, present and future. *Parasitology*, 145, 1820-1836.

- LETZEL, A. C., LI, J., AMOS, G. C. A., MILLAN-AGUINAGA, N., GINIGINI, J., ABDELMOHSEN, U. R., GAUDENCIO, S. P., ZIEMERT, N., MOORE, B. S. & JENSEN, P. R. 2017. Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. *Environ Microbiol*, 19, 3660-3673.
- LI, L., JIANG, W. & LU, Y. 2017a. New strategies and approaches for engineering biosynthetic gene clusters of microbial natural products. *Biotechnol Adv*, 35, 936-949.
- LI, L., ZHENG, G., CHEN, J., GE, M., JIANG, W. & LU, Y. 2017b. Multiplexed site-specific genome engineering for overproducing bioactive secondary metabolites in actinomycetes. *Metab Eng*, 40, 80-92.
- LI, Z., JIANG, Y., GUENGERICH, F. P., MA, L., LI, S. & ZHANG, W. 2020. Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. *Journal of Biological Chemistry*, 295, 833-849.
- LIU, X. 2016. Generate a bioactive natural product library by mining bacterial cytochrome P450 patterns. *Synth Syst Biotechnol*, 1, 95-108.
- MALDONADO, L. A., FENICAL, W., JENSEN, P. R., KAUFFMAN, C. A., MINCER, T. J., WARD, A. C., BULL, A. T. & GOODFELLOW, M. 2005. *Salinispora arenicola* gen. nov., sp. nov. and *Salinispora tropica* sp. nov., obligate marine actinomycetes belonging to the family Micromonosporaceae. *International Journal of Systematic and Evolutionary Microbiology*, 55, 1759-1766.
- MANIVASAGAN, P., VENKATESAN, J., SIVAKUMAR, K. & KIM, S. K. 2014. Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiol Res*, 169, 262-78.
- MATSUDA, S., ADACHI, K., MATSUO, Y., NUKINA, M. & SHIZURI, Y. 2009. Salinisporamycin, a novel metabolite from *Salinispora arenicola*. [corrected]. *J Antibiot (Tokyo)*, 62, 519-26.
- MEDEMA, M. H., KOTTMANN, R., YILMAZ, P., CUMMINGS, M., BIGGINS, J. B., BLIN, K., DE BRUIJN, I., CHOOI, Y. H., CLAESEN, J., COATES, R. C., CRUZ-MORALES, P., DUDELA, S., DUSTERHUS, S., EDWARDS, D. J., FEWER, D. P., GARG, N., GEIGER, C., GOMEZ-ESCRIBANO, J. P., GREULE, A., HADJITHOMAS, M., HAINES, A. S., HELFRICH, E. J., HILLWIG, M. L., ISHIDA, K., JONES, A. C., JONES, C. S., JUNGSMANN, K., KEGLER, C., KIM, H. U., KOTTER, P., KRUG, D., MASSCHELEIN, J., MELNIK, A. V., MANTOVANI, S. M., MONROE, E. A., MOORE, M., MOSS, N., NUTZMANN, H. W., PAN, G., PATI, A., PETRAS, D., REEN, F. J., ROSCONI, F., RUI, Z., TIAN, Z., TOBIAS, N. J., TSUNEMATSU, Y., WIEMANN, P., WYCKOFF, E., YAN, X., YIM, G., YU, F., XIE, Y., AIGLE, B., APEL, A. K., BALIBAR, C. J., BALSUS, E. P., BARONA-GOMEZ, F., BECHTHOLD, A., BODE, H. B., BORRIS, R., BRADY, S. F., BRAKHAGE, A. A., CAFFREY, P., CHENG, Y. Q., CLARDY, J., COX, R. J., DE MOT, R., DONADIO, S., DONIA, M. S., VAN DER DONK, W. A., DORRESTEIN, P. C., DOYLE, S., DRIESSEN, A. J., EHLING-SCHULZ, M., ENTIAN, K. D., FISCHBACH, M. A., GERWICK, L., GERWICK, W. H., GROSS, H., GUST, B., HERTWECK, C., HOFTE, M., JENSEN, S. E., JU, J., KATZ, L., KAYSSER, L., KLASSEN, J. L., KELLER, N. P., KORMANEC, J., KUIPERS, O. P., KUZUYAMA, T., KYRPIDES, N. C., KWON, H. J., LAUTRU, S., LAVIGNE, R., LEE, C. Y., LINQUAN, B., LIU, X., LIU, W., et al. 2015. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol*, 11, 625-31.
- MILLAN-AGUINAGA, N., CHAVARRIA, K. L., UGALDE, J. A., LETZEL, A. C., ROUSE, G. W. & JENSEN, P. R. 2017. Phylogenomic Insight into *Salinispora* (Bacteria, Actinobacteria) Species Designations. *Sci Rep*, 7, 3564.

- MNGUNI, F. C., PADAYACHEE, T., CHEN, W., GRONT, D., YU, J.-H., NELSON, D. R. & SYED, K. 2020. More P450s are involved in secondary metabolite biosynthesis in *Streptomyces* compared to *Bacillus*, *Cyanobacteria* and *Mycobacterium*. *International Journal of Molecular Sciences*, 21, 4814.
- MOODY, S. C. & LOVERIDGE, E. J. 2014. CYP105-diverse structures, functions and roles in an intriguing family of enzymes in *Streptomyces*. *J Appl Microbiol*, 117, 1549-63.
- MSOMI, N. N., PADAYACHEE, T., NZUZA, N., SYED, P. R., KRYŚ, J. D., CHEN, W., GRONT, D., NELSON, D. R. & SYED, K. 2021. In silico analysis of P450s and their role in secondary metabolism in the bacterial class Gammaproteobacteria. *Molecules*, 26, 1538.
- NAIR, S. & ABRAHAM, J. 2020. Natural products from actinobacteria for drug discovery. *Advances in Pharmaceutical Biotechnology*. Springer.
- NAVEED, A., LI, H. & LIU, X. 2018. Cytochrome P450s: Blueprints for Potential Applications in Plants. *Journal of Plant Biochemistry & Physiology*, 06.
- NELSON, D. R. 1998. Cytochrome P450 nomenclature. *Methods Mol Biol*, 107, 15-24.
- NELSON, D. R. 2006. Cytochrome P450 nomenclature, 2004. *Methods Mol Biol*, 320, 1-10.
- NELSON, D. R. 2013. A world of cytochrome P450s. *Philos Trans R Soc Lond B Biol Sci*, 368, 20120430.
- NELSON, D. R. 2018. Cytochrome P450 diversity in the tree of life. *Biochim Biophys Acta Proteins Proteom*, 1866, 141-154.
- NEWMAN, D. J. & HILL, R. T. 2006. New drugs from marine microbes: the tide is turning. *J Ind Microbiol Biotechnol*, 33, 539-44.
- NZUZA, N., PADAYACHEE, T., SYED, P. R., KRYŚ, J. D., CHEN, W., GRONT, D., NELSON, D. R. & SYED, K. 2021. Ancient Bacterial Class Alphaproteobacteria Cytochrome P450 Monooxygenases Can Be Found in Other Bacterial Species. *International journal of molecular sciences*, 22, 5542.
- OMURA, T. & SATO, R. 1962. A new cytochrome in liver microsomes. *J Biol Chem*, 237, 1375-6.
- PADAYACHEE, T., NZUZA, N., CHEN, W., NELSON, D. R. & SYED, K. 2020. impact of lifestyle on cytochrome P450 monooxygenase repertoire is clearly evident in the bacterial phylum Firmicutes. *Scientific reports*, 10, 1-12.
- PARVEZ, M., QHANYA, L. B., MTHAKATHI, N. T., KGOSIEMANG, I. K., BAMAL, H. D., PAGADALA, N. S., XIE, T., YANG, H., CHEN, H., THERON, C. W., MONYAKI, R., RASELEMANE, S. C., SALEWE, V., MONGALE, B. L., MATOWANE, R. G., ABDALLA, S. M., BOOI, W. I., VAN WYK, M., OLIVIER, D., BOUCHER, C. E., NELSON, D. R., TUSZYNSKI, J. A., BLACKBURN, J. M., YU, J. H., MASHELE, S. S., CHEN, W. & SYED, K. 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Sci Rep*, 6, 33099.
- PENN, K., JENKINS, C., NETT, M., UDWARY, D. W., GONTANG, E. A., MCGLINCHEY, R. P., FOSTER, B., LAPIDUS, A., PODELL, S., ALLEN, E. E., MOORE, B. S. & JENSEN, P. R. 2009. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J*, 3, 1193-203.
- PETERS, L., KONIG, G. M., WRIGHT, A. D., PUKALL, R., STACKEBRANDT, E., EBERL, L. & RIEDEL, K. 2003. Secondary Metabolites of *Flustra foliacea* and Their Influence on Bacteria. *Applied and Environmental Microbiology*, 69, 3469-3475.
- PODUST, L. M. & SHERMAN, D. H. 2012. Diversity of P450 enzymes in the biosynthesis of natural products. *Nat Prod Rep*, 29, 1251-66.
- QHANYA, L. B., MATOWANE, G., CHEN, W., SUN, Y., LETSIMO, E. M., PARVEZ, M., YU, J. H., MASHELE, S. S. & SYED, K. 2015. Genome-wide annotation and

- comparative analysis of cytochrome P450 monooxygenases in Basidiomycete biotrophic plant pathogens. *PLoS One*, 10, e0142100.
- RUDOLF, J. D., CHANG, C. Y., MA, M. & SHEN, B. 2017. Cytochromes P450 for natural product biosynthesis in *Streptomyces*: sequence, structure, and function. *Nat Prod Rep*, 34, 1141-1172.
- SENATE, L. M., TJATJI, M. P., PILLAY, K., CHEN, W., ZONDO, N. M., SYED, P. R., MNGUNI, F. C., CHILIZA, Z. E., BAMAL, H. D., KARPOORMATH, R., KHOZA, T., MASHELE, S. S., BLACKBURN, J. M., YU, J. H., NELSON, D. R. & SYED, K. 2019. Similarities, variations, and evolution of cytochrome P450s in *Streptomyces* versus *Mycobacterium*. *Sci Rep*, 9, 3962.
- SHARMA, A., KUMARI, N. & MENGHANI, E. 2014. Bioactive secondary metabolites: An overview. *International journal of Scientific and Engineering Research*, 5, 1395.
- SHUMYANTSEVA, V. V., KUZIKOV, A. V., MASAMREKH, R. A., BULKO, T. V. & ARCHAKOV, A. I. 2018. From electrochemistry to enzyme kinetics of cytochrome P450. *Biosens Bioelectron*, 121, 192-204.
- SIRIM, D., WIDMANN, M., WAGNER, F. & PLEISS, J. 2010. Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Struct Biol*, 10, 34.
- SONO, M., ROACH, M. P., COULTER, E. D. & DAWSON, J. H. 1996. Heme-containing oxygenases. *Chemical reviews*, 96, 2841-2888.
- SUBRAMANI, R. & AALBERSBERG, W. 2012. Marine actinomycetes: an ongoing source of novel bioactive metabolites. *Microbiol Res*, 167, 571-80.
- SYED, K., NELSON, D. R., RILEY, R. & YADAV, J. S. 2013. Genomewide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the polypore species *Bjerkandera adusta*, *Ganoderma* sp. and *Phlebia brevispora*. *Mycologia*, 105, 1445-55.
- THIRUMURUGAN, D., CHOLARAJAN, A., RAJA, S. S. & VIJAYAKUMAR, R. 2018. An Introductory Chapter: Secondary Metabolites. *Secondary Metabolites-Sources and Applications*. IntechOpen.
- TYC, O., SONG, C., DICKSCHAT, J. S., VOS, M. & GARBEVA, P. 2017. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends Microbiol*, 25, 280-292.
- UDWARY, D. W., ZEIGLER, L., ASOLKAR, R. N., SINGAN, V., LAPIDUS, A., FENICAL, W., JENSEN, P. R. & MOORE, B. S. 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci U S A*, 104, 10376-81.
- URLACHER, V. B. & EIBEN, S. 2006. Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol*, 24, 324-30.
- URLACHER, V. B. & GIRHARD, M. 2012. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol*, 30, 26-36.
- VAISHNAV, P. & DEMAIN, A. L. 2011. Unexpected applications of secondary metabolites. *Biotechnology Advances*, 29, 223-229.
- VAN BEILEN, J. B., HOLTACKERS, R., LÜSCHER, D., BAUER, U., WITHOLT, B. & DUETZ, W. A. 2005. Biocatalytic production of perillyl alcohol from limonene by using a novel *Mycobacterium* sp. cytochrome P450 alkane hydroxylase expressed in *Pseudomonas putida*. *Applied and environmental microbiology*, 71, 1737-1744.
- WEBER, T. & KIM, H. U. 2016. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synthetic and Systems Biotechnology*, 1, 69-79.
- WHITE, R. E. & COON, M. J. 1980. Oxygen activation by cytochrome P-450. *Annu Rev Biochem*, 49, 315-56.

- WILSON, M. C., GULDER, T. A., MAHMUD, T. & MOORE, B. S. 2010. Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the sare1259-encoded cytochrome P450. *Journal of the American Chemical Society*, 132, 12757-12765.
- XU, L. H. & DU, Y. L. 2018. Rational and semi-rational engineering of cytochrome P450s for biotechnological applications. *Synth Syst Biotechnol*, 3, 283-290.
- ZIEMERT, N., LECHNER, A., WIETZ, M., MILLÁN-AGUIÑAGA, N., CHAVARRIA, K. L. & JENSEN, P. R. 2014. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proceedings of the National Academy of Sciences*, 111, E1130-E1139.



## ANNEXURE

The following supporting information can be downloaded at:

<https://www.mdpi.com/article/10.3390/microorganisms10050871/s1>.

Table S1: Information on species used in the study. Identification of P450s that are part of secondary metabolite biosynthetic gene clusters (smBGCs) in *Salinispora* species. Cluster-ID and BGC type is retrieved from Integrated Microbial Genomes & Microbiomes (IMG/M) database. smBGC Type was indicated for consistency with the standard BGC Type name terminology available in the anti-SMASH database.

Table S2: P450 sequences identified and annotated in *Salinispora* species. Each P450 is presented with its assigned name followed by gene ID (in parenthesis) and species name.

Figure S1. Phylogenetic analysis of *Salinispora* species P450s. 2643 P450s were used to construct the tree, and the members of the eight most abundant P450 families are highlighted in different colors and indicated in the figure. P450 protein sequences used to build the tree are listed in Table S2.