UNIVERSITY OF ZULULAND

Annotation and comparative analysis of P450s, their redox partners and secondary metabolite gene clusters in the bacterial phylum *Firmicutes*

Tiara Padayachee
201509566

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

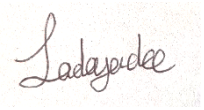
Supervisor: Prof Khajamohiddin Syed

KwaDlangezwa

November 2021

DECLARATION

I, Tiara Padayachee, declare that this dissertation is entirely my own work and has not been taken from the work of others, except where I have appropriately acknowledged and referenced the original source. This dissertation has never been submitted for any degree for examination at any university. Considering the quality is more important than the quantity (Gould, 2016), care has been taken to present the dissertation in a publication format to enable presentation of data in a concise manner and for easy understanding of the work. I also state that the ferredoxin work was carried out in collaboration with my colleague, Ms Nomfundo Nzuza, and the work presented in this dissertation has been published as indicated in the research outputs section.



_____

Signed on the 15th of September 2021

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

ABSTRACT

Cytochrome P450 monooxygenases (CYPs/P450s), heme thiolate proteins, are well known for their role in organisms' primary and secondary metabolism. P450s are also involved in the generation of secondary metabolites with biotechnological value. P450s require electrons to perform their function, and these electrons are provided by iron-sulfur (Fe-S) cluster proteins known as ferredoxins. However, to date, comparative analysis of P450s, redox proteins and P450s involved in secondary metabolite production is not reported in the bacterial phylum *Firmicutes*. This study aimed to address this research gap. Genome-wide analysis of P450s in 972 *Firmicutes* species belonging to 158 genera revealed that only 229 species belonging to 37 genera have P450s; 38% of *Bacilli* species, followed by 14% of *Clostridia* and 2.7% of other *Firmicutes* species, have P450s. The pathogenic or commensal lifestyle influences P450 content to such an extent that species belonging to the genera *Streptococcus*, *Listeria*, *Staphylococcus*, *Lactobacillus*, *Lactococcus* and *Leuconostoc* do not have P450s, except for a handful of *Staphylococcus* species that have a single P450. Only 18% of P450s are found to be involved in secondary metabolism and 89 P450s that function in the synthesis of specific secondary metabolites are predicted.

A total of 281 ferredoxins were found amongst 227 *Firmicutes* species. Four types of ferredoxins were found in this phylum, namely, 2Fe-2S, 4Fe-4S, 7Fe-8S and 2[4Fe-4S]. The dominant type found amongst this phylum was 4Fe-4S with 140 ferredoxins followed by 2Fe-2S with 97 ferredoxins, 7Fe-8S with 32 ferredoxins and 2[4Fe-4S] with 12 ferredoxins. To date, evolutionary analysis of these proteins across the domains of life is confined to observing the abundance of Fe-S cluster types and the diversity of ferredoxins within a type is not reported. To address this research gap, here a subtype classification and nomenclature for ferredoxins based on the characteristic spacing between the cysteine amino acids of the Fe-S binding motif as a subtype signature was proposed. To test this hypothesis, comparative analysis of ferredoxins between *Firmicutes* and ferredoxins collected from species of different domains of life that are reported in the literature has been carried out. Various ferredoxin subtypes were found amongst *Firmicutes* species. Eleven different subtypes were found amongst 2Fe-2S ferredoxins with subtype 20 being the dominant subtype. 4Fe-4S had five subtypes with subtype 2 being dominant, and 2[4Fe-4S] had three subtypes with subtype 9 being dominant. It is interesting to note that 7Fe-8S only had one subtype showing a preference to that specific ferredoxin subtype. Three subtypes of 2Fe-2S and two subtypes of

2[4Fe-4S] were found to be common between the Archaea and *Firmicutes* species, indicating their shared common ancestral origin.

Overall, the study results supported the hypothesis proposed by our laboratory that organisms' lifestyles impact P450 repertoire in their genomes. This is clearly evident in the bacterial phylum *Firmicutes* where a pathogenic or commensal lifestyle resulted in complete or nearly loss of P450s in the species. This study also is the first to propose ferredoxin subtype classification and identification of ferredoxins that have common ancestor origin (between Archean and Bacteria) and those are subjected to lateral gene transfer from prokaryotes (Archean/Bacteria) to eukaryotes.

DEDICATION

This work is dedicated to my family, my father Raj Padayachee, my mother Sarah Padayachee, my brother Tevin Padayachee and my boyfriend Thaveshan Govender.

ACKNOWLEDGEMENTS

Firstly, I would like to thank God for giving me the strength and support I needed. Secondly, I would like to thank my supervisor, Professor Khajamohiddin Syed, for believing in my abilities and pushing me to become a better researcher/scientist. I would also like to extend my appreciation to the following people: Prof David R Nelson (University of Tennessee, Memphis USA), Dr Wanping Chen (University of Göttingen, Germany) and Dr Dominik Gront (University of Warsaw, Poland) for their collaboration.

Lastly, I would like to thank my family for being my support system and always being there for me when I needed them. Without your constant strength and motivation all this would not have been possible. God bless you all.

RESEARCH OUTPUTS

- **Articles**

1. **Tiara Padayachee**, Nomfundo Nzuza, Wanping Chen, David R Nelson, Khajamohiddin Syed. Impact of lifestyle on cytochrome P450 monooxygenase repertoire is clearly evident in the bacterial phylum *Firmicutes* (2020) *Scientific reports*, *10*(1), pp.1-12.

2. Makhosazana Jabulile Khumalo, Nomfundo Nzuza, **Tiara Padayachee**, Wanping Chen, Jae-Hyuk Yu, David R Nelson, Khajamohiddin Syed. Comprehensive analyses of cytochrome P450 monooxygenases and secondary metabolite biosynthetic gene clusters in *Cyanobacteria* (2020) *International journal of molecular sciences*, *21*(2), p.656.

3. Fanele Cabangile Mnguni, **Tiara Padayachee**, Wanping Chen, Dominik Gront, Jae-Hyuk Yu, David R Nelson, Khajamohiddin Syed. More P450s are involved in secondary metabolite biosynthesis in *Streptomyces* compared to *Bacillus*, *Cyanobacteria*, and *Mycobacterium* (2020) *International journal of molecular sciences*, *21*(13), p.4814.

4. Ntombizethu Nokuphiwa Msomi, **Tiara Padayachee**, Nomfundo Nzuza, Puleng Rosinah Syed, Justyna Dorota Kryś, Wanping Chen, Dominik Gront, David R Nelson, Khajamohiddin Syed. In silico analysis of P450s and their role in secondary metabolism in the bacterial class *Gammaproteobacteria* (2021) *Molecules*, *26*(6), p.1538.

5. Nomfundo Nzuza, **Tiara Padayachee**, Puleng Rosinah Syed, Justyna Dorota Kryś, Wanping Chen, Dominik Gront, David R. Nelson and Khajamohiddin Syed. Ancient bacterial class alphaproteobacteria cytochrome P450 monooxygenases can be found in other bacterial species (2021) *International journal of molecular sciences*, *22*(11), p.5542.

6. Nomfundo Nzuza[@], **Tiara Padayachee**[@], Wanping Chen, Dominik Gront, David R Nelson, Khajamohiddin Syed. Diversification of ferredoxins across living organisms (2021) *Current Issues in Molecular Biology* 43(3), 1374-1390.

@ equally contributed to the study

- **Conference Attendance**

Online conference: Fireside Chat: Current IVD Status and Innovation for the Future, June 23, 2021 10:00 AM Harare, Pretoria.

- **Conference Presentation**

Tiara Padayachee and Khajamohiddin Syed. Impact of lifestyle on cytochrome P450 monooxygenase repertoire is clear in the bacterial phylum *Firmicutes*. 11th Annual Biomedical Research & Innovation Platform (BRIP) Symposium 18-19 October 2021, Cape Town, South Africa.

IN NEWS

- National Research Foundation facebook posts on 29 October, 2021 : https://www.facebook.com/pg/NRFSouthAfrica/posts/
- YIBA news, 11 October 2021 : https://yiba.co.za/unizulu-researchers-solved-the-six-decades-old-mystery-of-the-evolution-of-electron-transfer-proteins/
- Zululand observer on 15 October, 2021: https://www.pressreader.com/south-africa/zululand-observer/textview
- UNIZULU website on 6 October 2021: http://www.unizulu.ac.za/unizulu-researchers-solved-the-six-decades-old-mystery-of-the-evolution-of-electron-transfer-proteins/

ABBREVIATIONS

Anti-SMASH	Antibiotics & Secondary Metabolite Analysis Shell
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
CYP/P450	Cytochrome P450 monooxygeanse
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
iTOL	Interactive Tree of Life
KEGG	Kyoto Encyclopedia of Genes and Genomes
LGT	Lateral gene transfer
MAFFT	Multiple alignment using fast Fourier transform
MSA	Multiple sequence alignment
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NRPS	Non-ribosomal peptide synthases
PDB	Protein data bank
PKS	Polyketide synthase
Trex	Tree and Reticulogram Reconstruction
SCFAs	Short chain fatty acids

TABLE OF CONTENTS

DECLARATION 1
ABSTRACT 2
DEDICATION 4
ACKNOWLEDGEMENTS 5
RESEARCH OUTPUTS 6
ABBREVIATIONS 8
LIST OF FIGURES 11
LIST OF TABLES 12
CHAPTER 1: INTRODUCTION 13
1.1 Introduction 13
1.2 Problem statement 14
1.3 Value to the body of knowledge 14
1.4 Aim and objectives 15
1.4.1 Study aim 15
1.4.2 Study objectives 15
1.5 Dissertation overview 15
CHAPTER 2: LITERATURE REVIEW 16
2.1 *Firmicutes* 16
2.2 Secondary metabolites produced by *Firmicutes* 17
2.3 P450s role in production of secondary metabolites 19
2.4 Redox proteins 21
CHAPTER 3: IMPACT OF LIFESTYLE ON CYTOCHROME P450 MONOOXYGENASE REPERTOIRE IS CLEARLY EVIDENT IN THE BACTERIAL PHYLUM *FIRMICUTES* 23
3.1 Introduction 23
3.2 Results and discussion 25
3.2.1 Only a few *Firmicutes* species have P450s. 25
3.2.2 *Firmicutes* species have the lowest P450 diversity. 28
3.2.3 A small proportion of P450s are involved in secondary metabolism in *Firmicutes* species 33
3.2.4 Most *Firmicutes* species P450s are orphans of unknown function. 34
3.3 Methods 38
3.3.1 Species and database. 38
3.3.2 Genome data mining and annotation of P450s. 38

3.3.3 Phylogenetic analysis of P450s. 39
3.3.4 Generation of P450 profile heat-maps. 39
3.3.5 Secondary metabolite BGCs analysis and P450s identification. 39
3.3.6 Bacterial P450s and gene cluster data. 39
3.3.7 Functional prediction of P450s. 40
CHAPTER 4: COMPARATIVE ANALYSIS OF FERREDOXINS IN THE BACTERIAL PHYLUM *FIRMICUTES* WITH OTHER DOMAINS OF LIFE 41
4.1 Introduction 41
4.2 Materials and methods 45
4.2.1 Species and database 45
4.2.2 Genome data mining and annotation of ferredoxins 45
4.2.3 Ferredoxins subtype classification and nomenclature 50
4.2.4 Assigning Fe-S cluster subtypes to the ferredoxins retrieved from literature 51
4.2.5 Phylogenetic analysis of ferredoxins 51
4.2.6 Generation of ferredoxin subtype profile heat maps 51
4.3 Results and discussion 51
4.3.1 *Firmicutes* species have different Fe-S cluster type ferredoxins in their genomes 51
4.3.2.1 2Fe-2S 53
4.3.2.2 4Fe-4S 57
4.3.2.3 7Fe-8S 57
4.3.2.4 2[4Fe-4S] 57
4.4 Evolutionary linkage of ferredoxins subtype classification 57
4.5 Common ancestral linkage and lateral gene transfer of ferredoxins between *Firmicutes* and other domains of life 58
CHAPTER 5: Conclusions and future perspectives 60
REFERENCES 62
ANNEXURE 79

LIST OF FIGURES

Figure 1.1. Reactions that P450s catalyse, showing their versatility (Bernhardt, 2006) 13
Figure 2.1.Dominance of *Firmicutes* in the human gut relative to other microbes (Kong et al., 2014). 16
Figure 2.2. Intramolecular cyclization reaction (Podust and Sherman, 2012).....20
Figure 2.3. Substrate mediated proton delivery (Podust and Sherman, 2012) 20
Figure 2.4. The addition of OH groups to filipin (Podust and Sherman, 2012) 21
Figure 2.5. Cytochrome P450 catalytic cycle, showing the function of redox partners (McLean et al., 2015).....22
Figure 2.6. Schematic representation of different P450-redox protein combinations (Hannemann et al., 2007).....22
Figure 3.1. Analysis of P450s in *Firmicutes* species.....26
Figure 3.2. Phylogenetic tree of *Firmicutes* species P450s 29
Figure 3.3. Analysis of presence of P450 family (red) or its absence (green) in 229 *Firmicutes* species.....32
Figure 3.4. Comparative analysis of P450s associated with secondary metabolism in *Firmicutes* species.....32
Figure 4.1. Structures of the two-, three- and four-iron cluster types found in bacterial ferredoxins (Bruschi and Guerlesquin, 1988).....42
Figure 4.2. Unusual cluster binding motifs found in *Mycobacterium marinum* (Child et al., 2018) 44
Figure 4.3. Ferredoxin nomenclature based on the spacing between the cysteine amino acids of Fe-S cluster binding motif.....50
Figure 4.4. Comparative analysis of ferredoxins iron-sulfur (Fe-S) cluster in *Firmicutes* species.....52
Figure 4.5. Phylogenetic analysis of ferredoxins.....58
Figure 4.6. Heat map figure representing the presence (red) and absence (green) of ferredoxin subtypes in *Firmicutes,* Archaea, Bacteria and Eukarya 59

LIST OF TABLES

Table 2.1. Various secondary metabolites produced by *Firmicutes* species, their properties and respective references 18
Table 3.1. Comparative analysis of key features of P450s and their association with secondary metabolism between *Firmicutes* species and different bacterial species..... 27
Table 3.2. Comparative analysis of P450 families and subfamilies in *Firmicutes* species 30
Table 3.3. Functional prediction of *Firmicutes* species P450s' involvement in synthesis of secondary metabolites. 35
Table 4.1. Information on ferredoxins that are used as reference proteins in the study 47
Table 4.2. Subtype classification of ferredoxins across the domains of life..... 53

CHAPTER 1: INTRODUCTION

1.1 Introduction

Cytochrome P450 monooxygenases - also abbreviated as CYPs/P450s - are believed to have evolved from a common ancestor about three billion years ago (Nelson and Strobel, 1987). Initially discovered in rat liver microsomes, P450s were found to be hemeproteins and ubiquitously distributed in species belonging to all kingdoms (Nelson, 2018). Many P450s have been discovered, with P450BM3 isolated from *Bacillus megaterium* being extensively researched and modified to perform a range of functions (Notonier et al., 2016). P450s are known to be versatile biocatalysts having the profound ability to catalyze the oxidation of non-activated hydrocarbons, thereby overcoming the challenge many chemical catalysts face (Urlacher and Girhard, 2012) (Figure 1.1). Due to such versatility, they are regarded as promising subjects in the biotechnological field (Notonier et al., 2016). Catalytic versatility of P450s with respect to types of enzymatic reactions is shown in Figure 1.1.

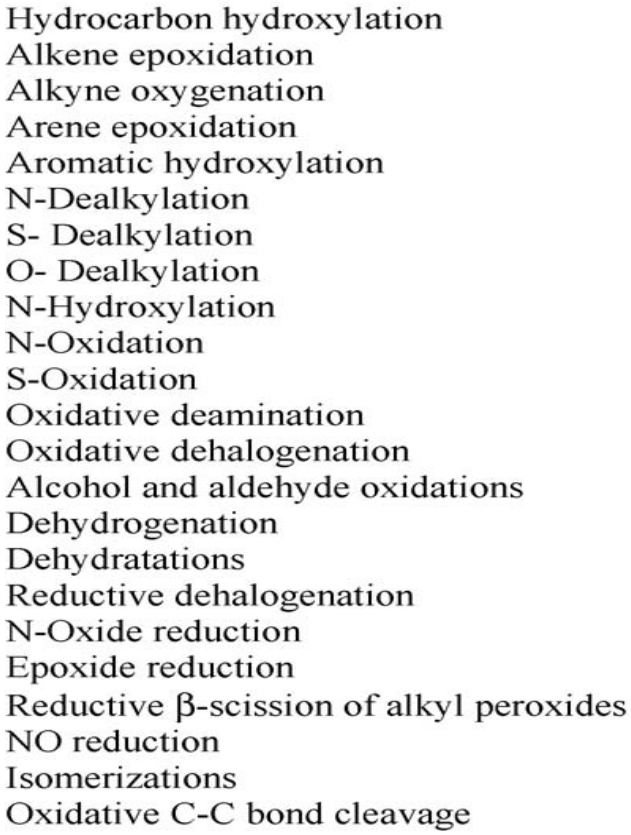


Figure 1.1. Reactions that P450s catalyse, showing their versatility (Bernhardt, 2006)

In bacteria, P450s are involved in primary and secondary metabolism including the biosynthesis of secondary metabolites and degradation of xenobiotic compounds (Paul R. Ortiz de Montellano, 2015). Understanding the P450s' role in secondary metabolites is now gaining momentum (Greule et al., 2018). Secondary metabolites are natural products found to have beneficial values in many scientific areas such as human and veterinary medicine, agriculture, and manufacturing (Cimermancic et al., 2014) and are known to mediate a variety of microbe-host and microbe-microbe interactions (Wilson et al., 2017).

P450s are involved in the synthesis of secondary metabolites ranging from terpenes, shikimates, alkaloids, polyketides, and peptides (Greule et al., 2018). Genes involved in the synthesis of secondary metabolites are found to be clustered together and form a gene cluster known as biosynthetic gene clusters (BGCs) and bacterial species have been found to have more than 1000 different types of BGCs involved in the synthesis of known and unknown secondary metabolites (Cimermancic et al., 2014). In this direction our laboratory unraveled P450s and their role in generation of secondary metabolites in bacterial species belonging to the genera *Bacillus* (Mthethwa et al., 2018), *Mycobacterium* and *Streptomyces* (Senate et al., 2019) and the bacterial phylum *Cyanobacteria* (Khumalo et al., 2020).

1.2 Problem statement

It is a well-known fact that P450 enzymes not only play a role in the generation of secondary metabolites but also contribute to secondary metabolite diversity (Podust and Sherman, 2012) and these P450s need electrons to perform their enzymatic reaction. These electrons were supplied by redox proteins. *Firmicutes* are well-known to generate secondary metabolites that play a role in human health and have biotechnological applications. However, to date, comparative analysis of P450s and their role in secondary metabolite production and redox proteins (ferredoxins) has not been reported. Thus, this study aimed to address this research gap. This study has the potential to identify P450s involved in secondary metabolites and biosynthetic gene clusters that can be used to produce human valuable compounds at a commercial level.

1.3 Value to the body of knowledge

This study generated fundamental knowledge on P450s, their redox proteins and secondary metabolite clusters in *Firmicutes*. This is a great contribution to the scientific community

from the University of Zululand and from South Africa. The study results serve as a reference for P450s, their redox proteins and secondary metabolites for the entire research community.

1.4 Aim and objectives

1.4.1 Study aim

To perform the annotation and comparative analysis of cytochrome P450s, their redox partners and secondary metabolite gene clusters in the human inhabiting bacterial species of the phylum *Firmicutes*.

1.4.2 Study objectives

- To perform genome data mining, annotation of P450s and their redox proteins in the phylum *Firmicutes*.
- To perform phylogenetic analysis of P450s and their redox proteins in the phylum *Firmicutes*.
- To perform genome data mining and identification of secondary metabolite gene clusters in the phylum *Firmicutes*.
- To identify P450s involved in secondary metabolites in the phylum *Firmicutes*.

1.5 Dissertation overview

This dissertation is divided into five chapters. Chapter 1 provides a brief overview of the work, with introduction, problem statement, and value to the body of knowledge, study aim and objectives. Chapter 2 consists of a literature review. Chapter 3 is presented in the format of a research article that was published in Scientific Reports detailing the P450s and their role in secondary metabolism in *Firmicutes* species. Chapter 4 consists of study results on ferredoxins that are under review in Current Issues in Molecular Biology. Overall conclusions and future perspectives are presented as part of Chapter 5. Supplementary data is presented in the Annexure.

CHAPTER 2: LITERATURE REVIEW

2.1 *Firmicutes*

There are countless bacteria that inhabit the human gut, and *Firmicutes* is one of the beneficial bacterial groups that dominates the gut (Ley et al., 2006b) (Figure 2.1). *Firmicutes* are a diverse phylum of bacteria that consists of many bacterial species classified into seven subphyla, namely *Bacilli*, *Clostridia*, *Erysipelotrichia*, *Limnochordia*, *Negativicutes*, *Termolithobacteria*, and *Tissierellia*. *Firmicutes* is known to influence human nutrition and metabolism. Amongst the genera, *Clostridia* contain eminent species such as *Eubacterium*, *Faecalibacterium*, *Roseburia* and *Ruminococcus* (Ramakrishna, 2013). *Bacillus* species are known for their secondary metabolite production and for producing dormant endospores during unfavorable growth conditions (Mthethwa et al., 2018). The increased ratio between *Bacteroidetes* and *Firmicutes* have been shown to influence conditions such as Crohn's disease and obesity (Marciano and Vajro, 2017).

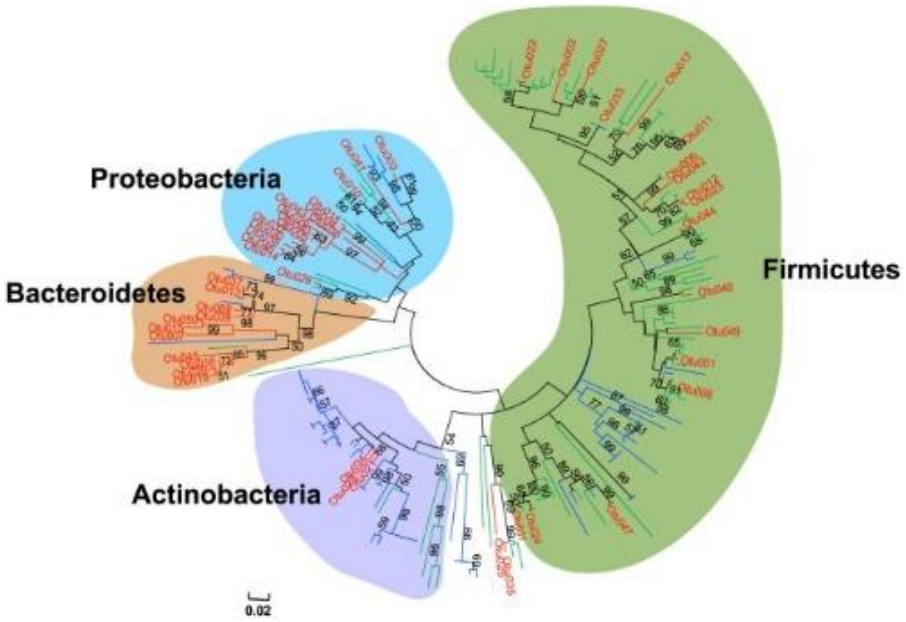


Figure 2.1. Dominance of *Firmicutes* in the human gut relative to other microbes (Kong et al., 2014).

Firmicutes are found to play a role in the energy extracted from the diet and consequently have an impact on obesity in humans (Gosalbes et al., 2011). Studies have shown that subjects who are obese or overweight tend to have a higher *Firmicutes* content in their gut (Chakraborti, 2015). Researchers proposed that manipulation of these gut microbes

might provide advancements in the treatment of obesity (Ley et al., 2006b). Many species belonging to this phylum are vital carbohydrate fermenters and may be associated with the salvage of energy from unabsorbed dietary carbohydrates, thus explaining the link between the gut microbiota and nutrition (Ramakrishna, 2013). Another study showed the importance of *Firmicutes* in carbohydrate metabolism on a transcriptional level (Ottman et al., 2012). Due to its major composition of the human gut (Castaner et al., 2018), *Firmicutes* are bound to have some contribution or effect on intestinal diseases or conditions such as Crohn's disease. Several studies have been done on this topic and the overall result showed a general decrease in *Firmicutes* species in Crohn's disease and ulcerative colitis (Manichanh et al., 2006). The decrease of these microbes further provides evidence that *Firmicutes* play a major role in the gut and, if absent, conditions may develop.

Firmicutes are composed of a large number of species including *Lactobacillus*, *Streptococcus*, *Mycoplasma* and *Clostridium* that are capable of producing several short chain fatty acids (SCFAs) like butyrate (Marciano and Vajro, 2017). This is an important characteristic as SCFAs such as butyrate have essential functions. Butyrate assists in trans-epithelial ion transport, cell growth and differentiation of the gut lumen, inflammation and oxidative stress and defensive properties in the human gut, and assists on an extra-intestinal level in terms of genetic metabolic diseases, obesity and insulin resistance (Canani et al., 2011). *Firmicutes* may also be used to explain certain age dependent events as the level of *Firmicutes* tends to increase with age (Mariat et al., 2009). Despite all the beneficial bacterial species in this phylum, pathogens are present such as species belonging to *Streptococcus* and *Staphylococcus*. An interesting fact about pathogens from this phylum is that they are not capable of growing outside their hosts, unlike pathogens such as *Vibrio cholera*. This suggests that transmission of this pathogen is possibly dependent on close contact of persons (Ley et al., 2006a).

2.2 Secondary metabolites produced by *Firmicutes*

Firmicutes produce various useful secondary metabolites. *Bacillus* species were responsible for most of the secondary metabolites produced as shown in Table 2.1. Polyketide synthase (PKS) is known to be a potential source of vital compounds in bacteria that have vigorous antibacterial, antiviral and cytotoxic properties such as macrolactin, bacillaene and difficidin (Yang et al., 2016). Difficidin is known to have actions against the rice pathogen (Xanthomonas oryzae) (Wu et al., 2015), All these metabolites are produced by various species in this phylum. Aurantinins B-D is synthesized by the PKS1 gene. This metabolite

has potent antimicrobial characteristics, especially towards *Staphylococcus aureus* (Yang et al., 2016). Previous studies on this metabolite show that it could be a possible source of environmentally friendly anti-microbial agents (Yang et al., 2016). Most of the secondary metabolites produced have anti-bacterial properties such as tauramamide that inhibits *Enterococcus* sp. (Kang et al., 2015), bacillaene that is known to be an unstable antibiotic (Moldenhauer et al., 2007) and fengycin that is an antimicrobial lipopeptide (Sur et al., 2018). Surfactin was originally produced by *Bacillus amyloliquefaciens* and has inhibitory actions towards *Candida albicans.* Apart from being an anti-fungal agent it also has anti-viral and anti-microbial properties (Kang et al., 2015). Lichenysin is a molecule that closely resembles surfactin in terms of its structure with some differences. Due to these structural differences this metabolite is known to be a more efficient cation chelator (Grangemard et al., 2001). Bacillibactin is classified as a catechol siderophore that binds and transports iron (Hertlein et al., 2014). This metabolite is produced by the honeybee pathogen (*Paenibacillus larvae*) that is also a species in this phylum (Hertlein et al., 2014). Enzymes in bacterial species are responsible for delivering amino acids to a growing protein chain, an example of such a mega-enzyme is a non-ribosomal peptide synthetase which is also produced by *Firmicutes* (Hertlein et al., 2014). Products of such enzymes tend to have various vital properties including anti-tumour and cytotoxic properties (Hertlein et al., 2014). *Fusarium graminearum* is a fungal pathogen that destroys the production and quality of wheat and barley. Various *Bacillus* species produce a lipopeptide bacillomycin that in a previous study showed evident antifungal properties towards such pathogens (Gu et al., 2017). Other antifungal metabolites include rhizocticin and basiliskamides (Borisova et al., 2010, Barsby et al., 2002). Taking into consideration all the functions that these metabolites produce, it can be stated that *Firmicutes* is a diverse phylum with many potential benefits.

Table 2.1. Various secondary metabolites produced by *Firmicutes* species, their properties and respective references

Secondary metabolites	Species (Source)	Role	Reference
Macrolactin	*Bacillus subtilis*	Antiviral, anticancer and antimicrobial	(Yuan et al., 2016, Romero-Tabarez et al., 2006)
Bacillaene	*Bacillus subtilis* 168	Unstable antibiotic	(Moldenhauer et al.,

	and *Bacillus amyloliquefaciens* FZB 42		2007, Butcher et al., 2007)
Difficidin	*Bacillus amyloliquefaciens* FZB 42	Acts against the rice pathogen	(Wu et al., 2015, Chen et al., 2006)
Aurantinins B-D	*Bacillus subtilis* fmb60	Antimicrobial	(Yang et al., 2016)
Tauramamide	*Brevibacillus laterosporus*	Antibacterial	(Kang et al., 2015)
Fengycin	*Bacillus subtilis* subsp. inaquosorum	Antibacterial	(Sur et al., 2018, Knight et al., 2018)
Surfactin	*Bacillus subtilis* and *Bacillus amyloliquefaciens*	Antiviral, antimicrobial and antifungal	(Peypoux et al., 1999, Grangemard et al., 2001)
Lichenysin	*Bacillus licheniformis*	Antiviral, antimicrobial and antifungal	(Grangemard et al., 2001)
Bacillibactin	*Bacillus sp.* PKU-MA00093 and PKU-MA00092 and *Paenibacillus larvae*	Transports iron	(Zhou et al., 2018, Hertlein et al., 2014)
Bacillomycin	*Bacillus sp.* PKU-MA00093 and PKU-MA00092	Antifungal	(Zhou et al., 2018, Gu et al., 2017)
Basiliskamides	*Bacillus laterosporus*	Antifungal	(Barsby et al., 2002)

2.3 P450s role in production of secondary metabolites

P450s are well known to be involved in the synthesis of a vast collection of secondary metabolites ranging from macrolides and glycopeptides to alkaloids and aromatic polyketides (Podust and Sherman, 2012). Their enzymatic role is crucial in the production of such metabolites. P450s are versatile enzymes and thus perform diverse types of reactions such as

the rearrangement of carbon skeletons, cleavage of carbon-carbon bonds and ring-coupling (Podust and Sherman, 2012). CYP154A1 performs an intramolecular cyclization reaction that produces a product that resembles an antibiotic, namely methylenomycin C, as shown in Figure 2.2. Unlike most P450 enzymatic reactions, this does not require a redox protein. CYP158A1 and CYP158A2 exhibit indirect substrate mediated proton delivery as shown in Figure 2.3 which results in the dimerization of flaviolin. These two P450s are also capable of performing ring-coupling reactions (Podust and Sherman, 2012). CYP3A4 that is found in the human liver is known to be involved in hydroxylation reactions. CYP105P1 and CYP105D6 introduce one OH group each to a polyene macrolide antibiotic, filipin (Figure 2.4). The antibiotic hydroxylases both P450s for defensive purposes such as the degradation of xenobiotics (Podust and Sherman, 2012).

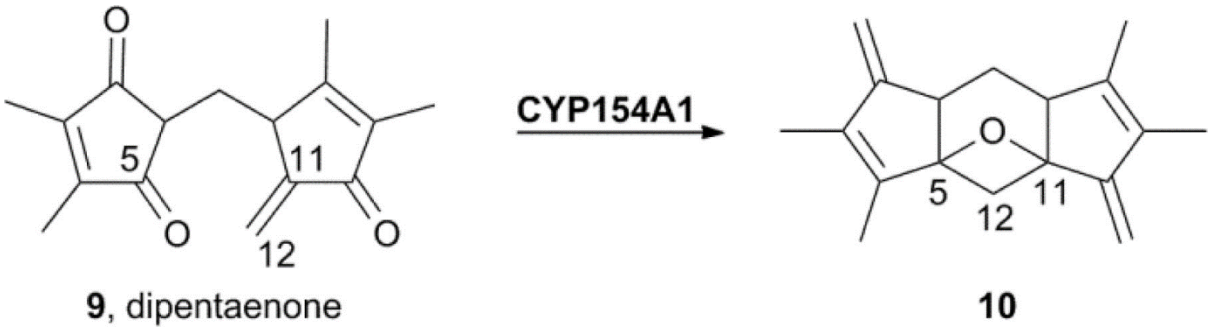


Figure 2.2. Intramolecular cyclization reaction (Podust and Sherman, 2012)

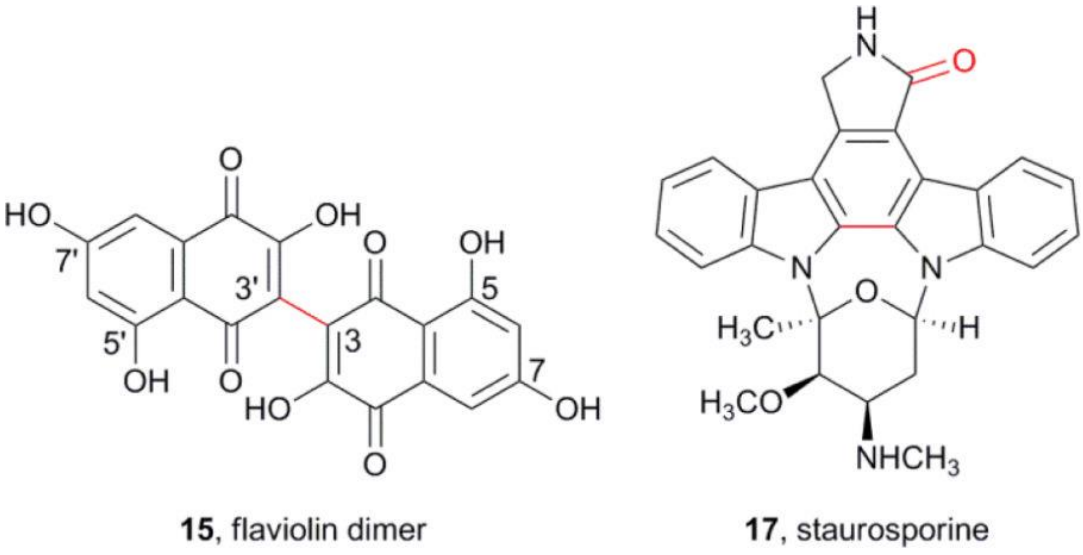


Figure 2.3. Substrate mediated proton delivery (Podust and Sherman, 2012)

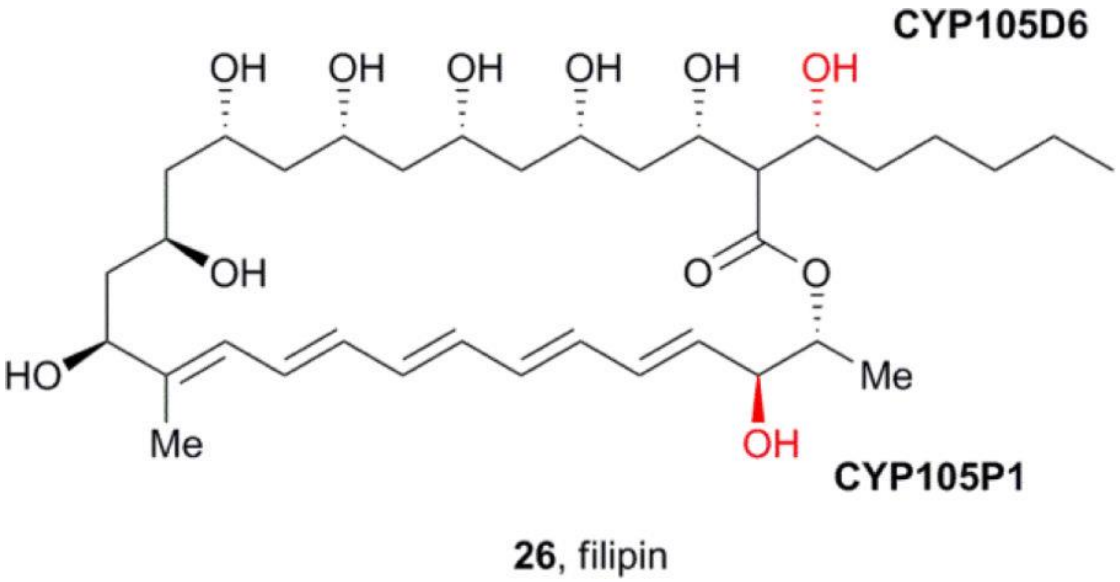


Figure 2.4. The addition of OH groups to filipin (Podust and Sherman, 2012)

2.4 Redox proteins

P450s are known to be involved in oxidation reactions. These enzymes require electrons in order to perform these reactions which are usually derived by NADH and NADPH (Lewis and Hlavica, 2000). Self-sufficent P450s do not require an additional redox partner. However, with the exception of self-sufficient P450s, all P450s interact with redox partners for electrons (Hannemann et al., 2007). Redox partner enzymes are responsible for delivering these electrons as shown in Figure 2.5 (Munro et al., 2007). Due to the versatility of these enzymes, there are numerous P450-redox combinations that may occur as depicted in Figure 2.6. There are two classes of redox protein systems, adrenal mitochondrial P450 systems and the liver microsomal P450 system. Both systems gain electrons from NADPH via different pathways, namely adrenodoxin reductase and adrenodoxin and FAD and FMN-containing P450 reductase respectively (Hannemann et al., 2007). Most bacterial P450 systems generally have a FAD-containing reductase which transfers electrons from NADPH to ferredoxin which reduces the P450 (Hannemann et al., 2007). Eukaryotic P450 systems have two major proteins which is the cytochrome P450 and the NADPH-cytochrome P450 reductase which contains FAD and FMN. Other potential redox partners of P450s include FAD-containing ferredoxin reductase, ferredoxin oxidoreductase, P450-ferredoxin fusion protein and flavodoxin-P450 fusion protein (Hannemann et al., 2007).

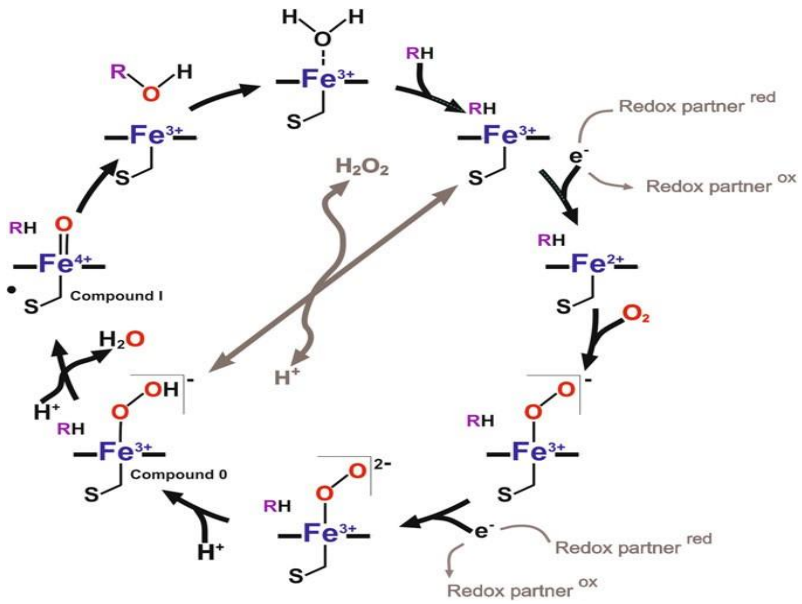


Figure 2.5. Cytochrome P450 catalytic cycle, showing the function of redox partners (McLean et al., 2015)

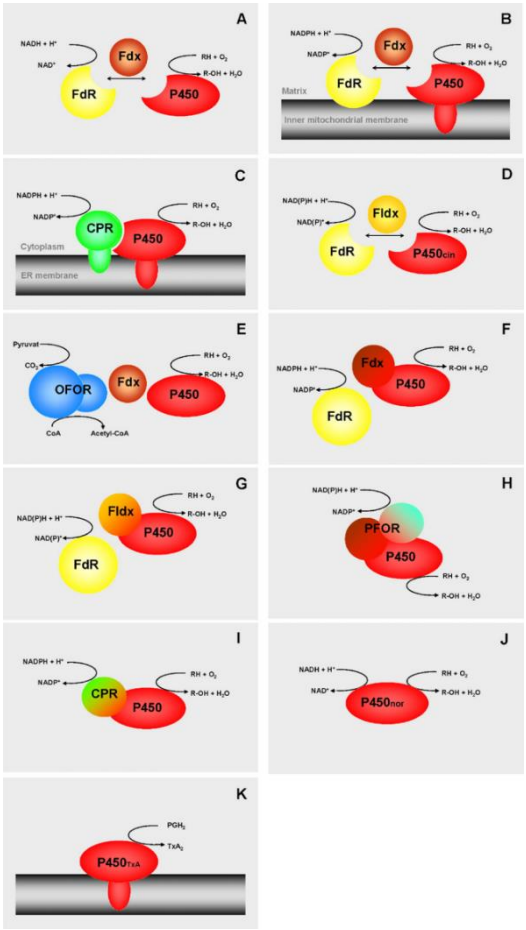


Figure 2.6. Schematic representation of different P450-redox protein combinations (Hannemann et al., 2007)

CHAPTER 3: IMPACT OF LIFESTYLE ON CYTOCHROME P450 MONOOXYGENASE REPERTOIRE IS CLEARLY EVIDENT IN THE BACTERIAL PHYLUM *FIRMICUTES*

3.1 Introduction

Among the bacteria that inhabit the human gut, species belonging to the bacterial phylum *Firmicutes* and *Bacteroides* are dominant (Arumugam et al., 2011, Dieterich et al., 2018, Thursby and Juge, 2017). *Firmicutes* species are gram-positive microorganisms with rod or sphere shapes and reproduce *via* binary fission. This phylum contains bacteria possessing diverse characteristics that are adapted to diverse ecological niches. Some members are saprophytes that live in soil and aquatic environments, performing mainly decomposition and recycling of organic matter, some members are commensals of humans and some members are pathogens of animals, including humans, and plants (Seong et al., 2018). Some members have been subjected to thorough investigation to gain understanding of endospore formation and survival (Filippidou et al., 2016) and the biotechnological potential of these organisms has been explored for the production of dairy products (Bintsis, 2018), enzymes (Verma et al., 2018) and antibiotics (Caulier et al., 2019).

The *Firmicutes* phylum is further divided into seven subphyla, namely *Bacilli*, *Clostridia*, *Erysipelotrichia*, *Limnochordia*, *Negativicutes*, *Termolithobacteria*, and *Tissierellia* (Seong et al., 2018). Species belonging to the subphyla *Bacilli* are well known for the production of secondary metabolites valuable to humans, organic compounds that have medicinal properties (Table 2.1). *Clostridia* consist of species that produce short chain fatty acids in the human gut, such as butyrate, which is essential fuel for colonocytes (enterocytes referred to as colonocytes in the colon) (Lopetuso et al., 2013) and also species causing botulism, tetanus, gas gangrene, food poisoning, pseudomembranous colitis, and antibiotic associated diarrhea in humans (Rood, 2016, Wells and Wilkins, 1996). Quite a number of studies have explored the relation between the changes in the percentage of *Firmicutes* species in the human gut and human conditions such as obesity, inflammatory bowel disease, systemic lupus erythematosus and psoriasis but results are not conclusive (Dieterich et al., 2018, Kho and Lal, 2018, Koliada et al., 2017).

Because of the potential biotechnological applications of secondary metabolites produced by *Bacillus* species (Table 2.1), comprehensive analysis of cytochrome P450 monooxygenases (CYPs/P450s) in the species belonging to the genus *Bacillus* has recently

been carried out and their association with secondary metabolism has been unraveled (Mthethwa et al., 2018).

P450s are heme thiolate proteins that play an important role in organisms' primary and secondary metabolism. Because of their diverse enzymatic reactions, P450s are found to play key roles such as conferring diversity on secondary metabolites (Podust and Sherman, 2012, Greule et al., 2018). The study revealed the presence of 507 P450s belonging to 13 P450 families and 28 P450 subfamilies in 128 *Bacillus* species and 112 P450s were found to be part of secondary metabolite biosynthetic gene clusters (BGCs) (Mthethwa et al., 2018). BGCs are groups of genes clustered together that are responsible for producing secondary metabolites in organisms (Medema et al., 2015). However, this study was limited to the genus *Bacillus*. None of the species belonging to other genera or a comprehensive analysis of P450s in *Firmicutes* species has been reported.

Studies on the role of P450s in organism's adaptations or changes in P450 profiles according to an organism's life have been conducted on eukaryotic organisms such as fungi (Ngwenya et al., 2018, Matowane et al., 2018, Qhanya et al., 2015, Kgosiemang et al., 2014, Jawallapersand et al., 2014, Syed K, 2014, Suzuki et al., 2012, Akapo et al., 2019), oomycetes (Sello et al., 2015), plants (Hamberger and Bak, 2013), and animals (Feyereisen, 2012). However, this phenomenon is scarcely reported in prokaryotes; currently data are available only for bacterial species belonging to the genera *Streptomyces* (Senate et al., 2019, Mnguni et al., 2020), *Mycobacterium* (Parvez et al., 2016, Syed et al., 2019) and the phylum *Cyanobacteria* (Khumalo et al., 2020), suggesting that more research is needed to unravel P450s' role in other bacterial species. The presence of species with diverse characteristics that are adapted to diverse ecological niches, such as in *Firmicutes*, will enhance the understanding of the role of P450s in bacterial species, especially with respect to their role in those organism's adaptations. Since *Firmicutes* species have diverse lifestyles, it would be interesting to see if there are any changes in P450 profiles with respect to their lifestyle, as studies have shown that organisms lose a considerable number of P450s owing to their lifestyle. Examples are adapting to utilize simple carbon sources, as observed in *Saccharomyces* species (Kgosiemang et al., 2014), or readily available abundant carbon sources in the host, as observed in mycobacterial species (Parvez et al., 2016), or making more copies of specific P450s both at P450 family and subfamily level in their genomes (P450 blooms) owing to adaptation to diferent ecological niches (Ngwenya et al., 2018,

Qhanya et al., 2015, Jawallapersand et al., 2014, Syed K, 2014, Suzuki et al., 2012, Akapo et al., 2019, Sello et al., 2015, Feyereisen, 2012).

In order to address the above research gaps on the impact of lifestyle on *Firmicutes* species P450s, if any, in this study comprehensive analysis of P450s and their association with secondary metabolism in 972 *Firmicutes* species belonging to 158 genera has been carried out.

3.2 Results and discussion

3.2.1 Only a few *Firmicutes* species have P450s.

Comprehensive genome-wide data mining and annotation of P450s in 972 *Firmicutes* species revealed that only 229 species have P450s in their genomes (Fig. 3.1A). P450 sequences found in this phylum is represented in Supplementary dataset 1. This indicates that most of the *Firmicutes* species do not have P450s in their genomes (Fig. 3.1A). Subphylum level analysis revealed that 38% of *Bacilli* species have P450s (Fig. 3.1B). In contrast to *Bacilli*, only 14% and 2.7% of *Clostridia* and other *Firmicutes* species have P450s (Fig. 3.1B). Owing to the availability of few species genomes that belong to the other five *Firmicutes* subphyla (*Erysipelotrichia*, *Limnochordia*, *Negativicutes*, *Termolithobacteria*, and *Tissierellia*), these species were kept under "Others" at KEGG (Kanehisa et al., 2015). Thus, we indicated the P450 profiles in these species under the name "other species". Only one species (*Limnochorda pilosa*) among 37 species in this category had P450s. Analysis of the *Firmicutes* genera disclosed that of the 158 genera, species belonging to 37 genera have P450s in their genomes (Fig. 3.1C). Most of the species belonging to the genus *Bacillus* have P450s, followed by *Paenibacillus* and *Clostridium* (Fig. 3.1C). Based on the number of species used in the study, we can safely conclude that species belonging to the genera *Streptococcus*, *Lactobacillus*, *Listeria*, *Geobacillus*, *Lactococcus* and *Leuconostoc* do not have P450s (Fig. 3.1C). Furthermore, among 86 *Staphylococcus* species only nine species have P450s (Fig. 3.1C). On average three P450s were found in 229 species, whereas 92 species had a single P450 in their genome. Among the *Firmicutes*, species belonging to the genus *Paenibacillus* had the highest number of P450s in their genomes. *P. mucilaginous* 3016 had the highest number of P450s (11 P450s) in its genome, followed by *P. mucilaginous* K02 and *P. mucilaginous* KNP414 (10 P450s each). Comparative analysis of different bacterial species revealed that *Firmicutes* species had the lowest average number of P450s in their genomes compared to the bacterial species belonging to the genera *Streptomyces* and *Mycobacterium* and the phylum *Cyanobacteria* (Table 3.1). The absence of

P450s in species belonging to the genera *Streptococcus*, *Lactobacillus*, *Listeria*, *Lactococcus*, *Leuconostoc* or the presence of a single P450 in a handful of *Staphylococcus* species strongly suggests that the impact of lifestyle profoundly influenced the P450 repertoire in the species, as also observed in mycobacterial species (Parvez et al., 2016) and in *Saccharomyces* species (Kgosiemang et al., 2014). The pathogenic lifestyle of species belonging to the genera *Streptococcus*, *Listeria* and *Staphylococcus* led them to adapt readily available carbon sources in the host and thus possibly led to the loss of P450s, similarly to mycobacterial species (Parvez et al., 2016). Species belonging to the genera *Lactobacillus*, *Lactococcus* and *Leuconostoc* are involved in fermentation processing (industrial scale or inside the human gut), indicating that adaption to thrive on simple carbon sources led to the loss of P450s, just as observed in *Saccharomyces* species (Kgosiemang et al., 2014).

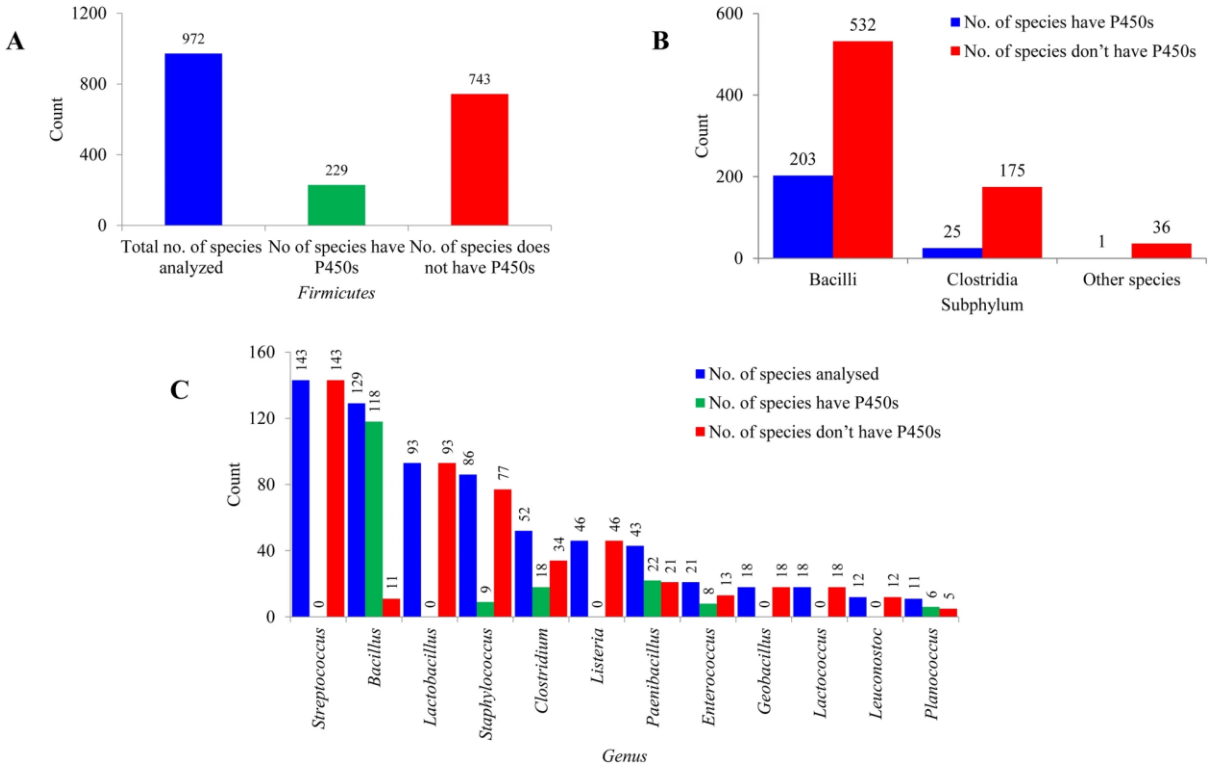


Figure 3.1. Analysis of P450s in *Firmicutes* species. Comparative analysis of P450s at species level (A), subphylum level (B) and genera level (C) is presented in the figure. Owing to the availability of few species genomes belonging to the other five subphyla of *Firmicutes* (*Erysipelotrichia*, *Limnochordia*, *Negativicutes*, *Termolithobacteria*, and *Tissierellia*), these species were kept under "Others" at KEGG (Kanehisa et al., 2015). Thus, we indicated the P450 profiles in these species under the name Other species. Numbers next to the bars indicate the number of species

Table 3.1. Comparative analysis of key features of P450s and their association with secondary metabolism between *Firmicutes* species and different bacterial species

	Firmicutes species	*Streptomyces* species	Mycobacterial species	*Bacillus* species	Cyanobacterial species
Total No. of species analyzed	972	203	60	128	114
No. of P450s	712	5460	1784	507	341
No. of families	15	253	77	13	36
No. of subfamilies	53	698	132	28	79
Dominant P450 family	CYP107	CYP107	CYP125	CYP107	CYP110
Average no. of P450s	1	27	30	4	3
No of BGCs*	2186	4457	898	1098	770
Average no. of BGCs	2	31	15	9	7
No. of P450s part of BGCs	126	1231	204	112	27
Percentage of P450s part of BGCs	18	32	11	22	8
Reference	(Padayachee et al., 2020)	(Mnguni et al., 2020, Senate et al., 2019)	(Senate et al., 2019, Parvez et al., 2016)	(Mthethwa et al., 2018)	(Khumalo et al., 2020)

3.2.2 *Firmicutes* species have the lowest P450 diversity.

Firmicutes species P450s were grouped into different P450 families and P450 subfamilies following the International P450 Nomenclature Committee rules that include phylogenetic analysis of P450s (Fig. 3.2) (Nelson et al., 1993, Nelson, 1998, Nelson, 2006). Based on the percentage identity of >40% for a family and >55% for a subfamily and following the evolutionary analysis where P450s belonging to the same family were grouped together (Fig. 3.2), all 712 P450s found in 229 *Firmicutes* species were grouped into 14 P450 families and 53 P450 subfamilies (Table 3.2). The number of P450 families found in *Firmicutes* species is exceptionally low compared to other bacterial species (Table 3.1), indicating the lowest P450 diversity. As predicted, P450 families such as CYP107, CYP102, CYP152, CYP109 and CYP106 were expanded in *Firmicutes* species, as the percentage contribution of these families to the total number of P450s was higher compared to the rest of the P450 families (Table 3.2). Among P450 families, CYP107 had the highest number of P450s (199 P450s), contributing 28% of 712 P450s, followed by CYP102 (179 P450s), CYP152 (110 P450s), CYP109 (95 P450s) and CYP106 (57 P450s) (Table 3.2). Blooming of certain P450 families in species is a common phenomenon and is observed in species belonging to different biological kingdoms as a potential indication of adaptation to an ecological niche (Ngwenya et al., 2018, Qhanya et al., 2015, Jawallapersand et al., 2014, Syed K, 2014, Suzuki et al., 2012, Akapo et al., 2019, Sello et al., 2015, Feyereisen, 2012). Comparative analysis of dominant P450 families across different bacterial species revealed that the CYP107 family is dominantly present in *Firmicutes* species and *Streptomyces* species and CYP125 and CYP110 are dominantly present in mycobacterial and cyanobacterial species, respectively (Table 3.1).

Analysis of P450 subfamilies revealed subfamily-level blooming in *Firmicutes* species as some subfamilies were expanded in a family (Table 3.2). CYP107 had the highest number of P450 subfamilies (12 subfamilies), followed by CYP152 (11 subfamilies), CYP109 (10 subfamilies), CYP197 (6 subfamilies), CYP106 (3 subfamilies) and CYP134 and CYP1731 and (Table 3.2). Seven of 14 P450 families had a single subfamily (Table 3.2). The dominant subfamilies in expanded P450 families were the following: CYP107 family had the subfamily “J” as the dominant subfamily; subfamily “B” was dominant in P450 families CYP106 and CYP109 and subfamily “A” was dominant in the CYP152 family. It is interesting to note that the CYP102 P450 family, despite contributing the second largest number of P450s, had a single subfamily “A” (Table 3.2). Blooming of certain P450

subfamilies in a family was also observed in species of different biological kingdoms and it is assumed that the P450 blooms possibly bestow certain advantages on organisms in adapting to ecological niches (Ngwenya et al., 2018, Qhanya et al., 2015, Jawallapersand et al., 2014, Syed K, 2014, Suzuki et al., 2012, Akapo et al., 2019, Sello et al., 2015, Feyereisen, 2012). Analysis of conservation of P450 families in 229 *Firmicutes* species revealed that none of the 14 P450 families was conserved in these species (Fig. 3.3). However, based on the heat-map profile of P450 families, the P450 families CYP152, CYP107, CYP120 and CYP109 were found to be a co-presence in most *Firmicutes* species (Fig. 3.3). Cyanobacterial species also had no P450 family conserved, but P450 families CYP110 and CYP120 were found to be a co-presence in most of these species (Khumalo et al., 2020). the large CYP107 family was found to be conserved in *Streptomyces* species (Mnguni et al., 2020) and several P450 families were found to be conserved in mycobacterial species (Parvez et al., 2016).

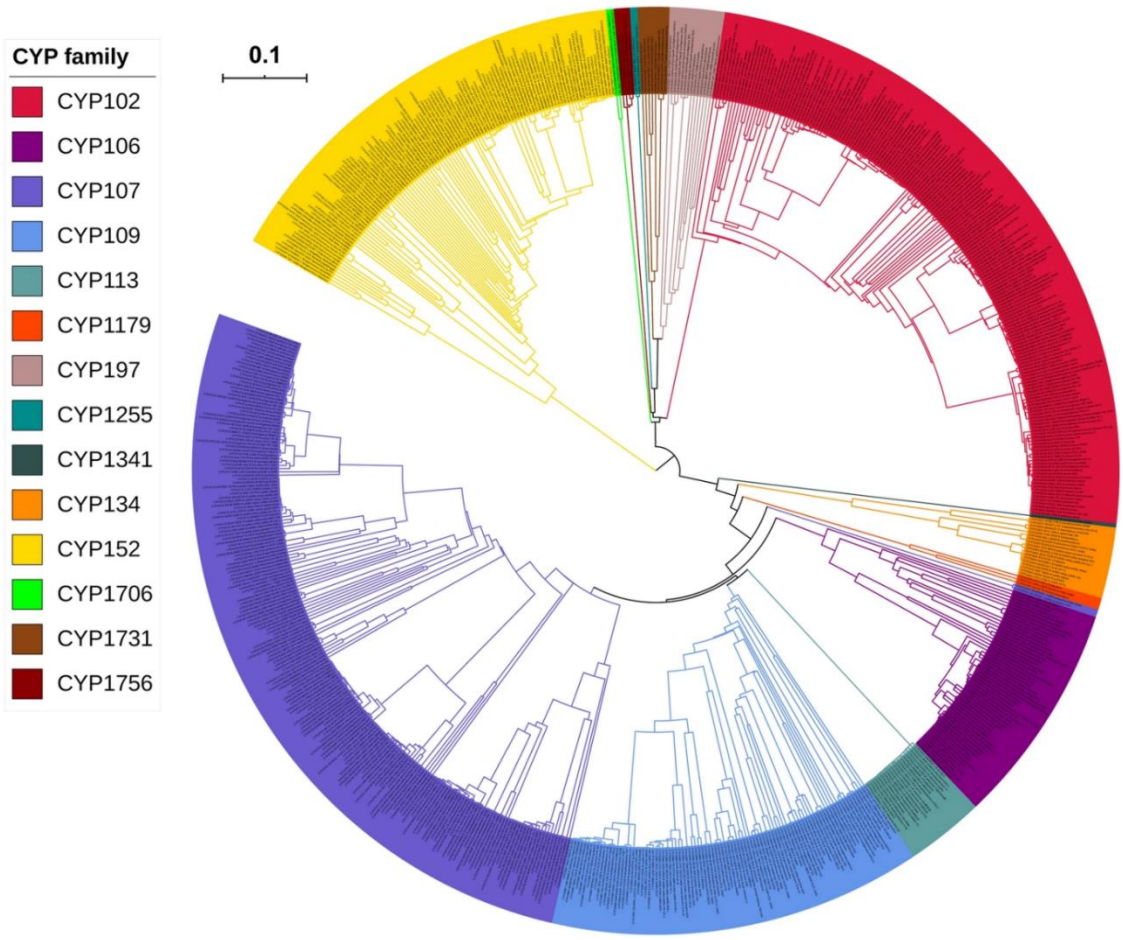


Figure 3.2. Phylogenetic tree of *Firmicutes* species P450s. Different P450 families are indicated with different colours

Table 3.2. Comparative analysis of P450 families and subfamilies in *Firmicutes* species. The percentage contribution of a particular family and its subfamilies to the total number of P450s is also presented in the table

Family	P450 count	Percentage count	Subfamily	P450 count	Percentage contribution
CYP102	179	25.14	A	179	25.14
CYP106	57	8.01	A	15	2.11
			B	41	5.76
			C	1	0.14
CYP107	199	27.95	CB	3	0.42
			DE	3	0.42
			DF	12	1.69
			DY	8	1.12
			H	47	6.60
			J	66	9.27
			JF	7	0.98
			JG	2	0.28
			JH	3	0.42
			JF	1	0.14
			K	45	6.32
			NJ	2	0.28
CYP109	95	13.34	A	23	3.23
			AJ	1	0.14
			B	44	6.18
			E	4	0.56
			J	3	0.42
			T	12	1.69
			U	1	0.14
			V	3	0.42
			W	3	0.42
			X	1	0.14

CYP113	20	2.81	L	20	2.81
CYP1179	3	0.42	A	3	0.42
CYP1255	2	0.28	A	2	0.28
CYP1341	1	0.14	C	1	0.14
CYP134	18	2.53	A	14	1.97
			C	4	0.56
CYP152	110	15.45	A	61	8.57
			AC	1	0.14
			AK	1	0.14
			AL	1	0.14
			AM	1	0.14
			AN	5	0.70
			J	1	0.14
			K	12	1.69
			L	11	1.54
			M	14	1.97
			N	2	0.28
CYP1706	2	0.28	B	2	0.28
CYP1731	8	1.12	A	5	0.70
			B	3	0.42
CYP1756	4	0.56	A	4	0.56
CYP197	14	1.97	A	1	0.14
			AD	3	0.42
			AE	1	0.14
			AF	3	0.42
			AH	3	0.42
			S	3	0.42

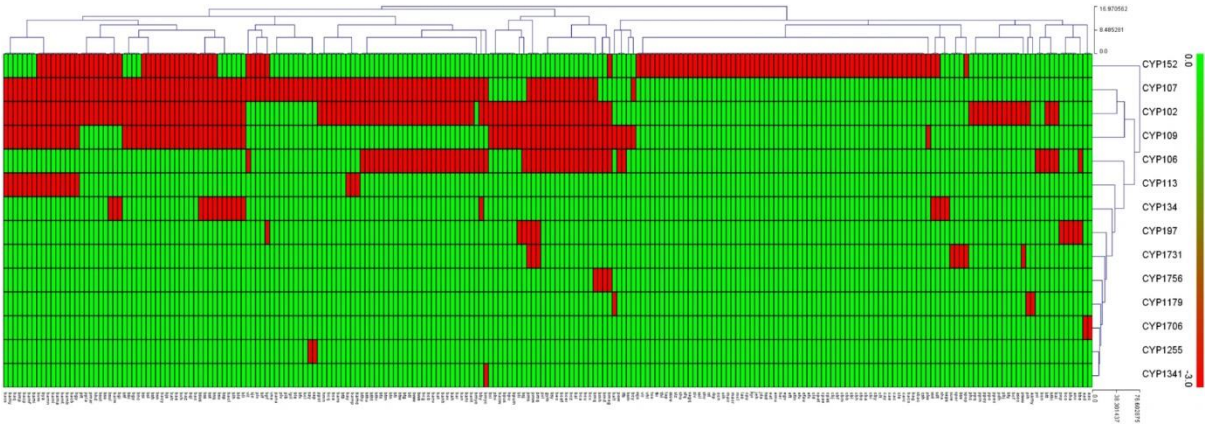


Figure 3.3. Analysis of presence of P450 family (red) or its absence (green) in 229 *Firmicutes* species

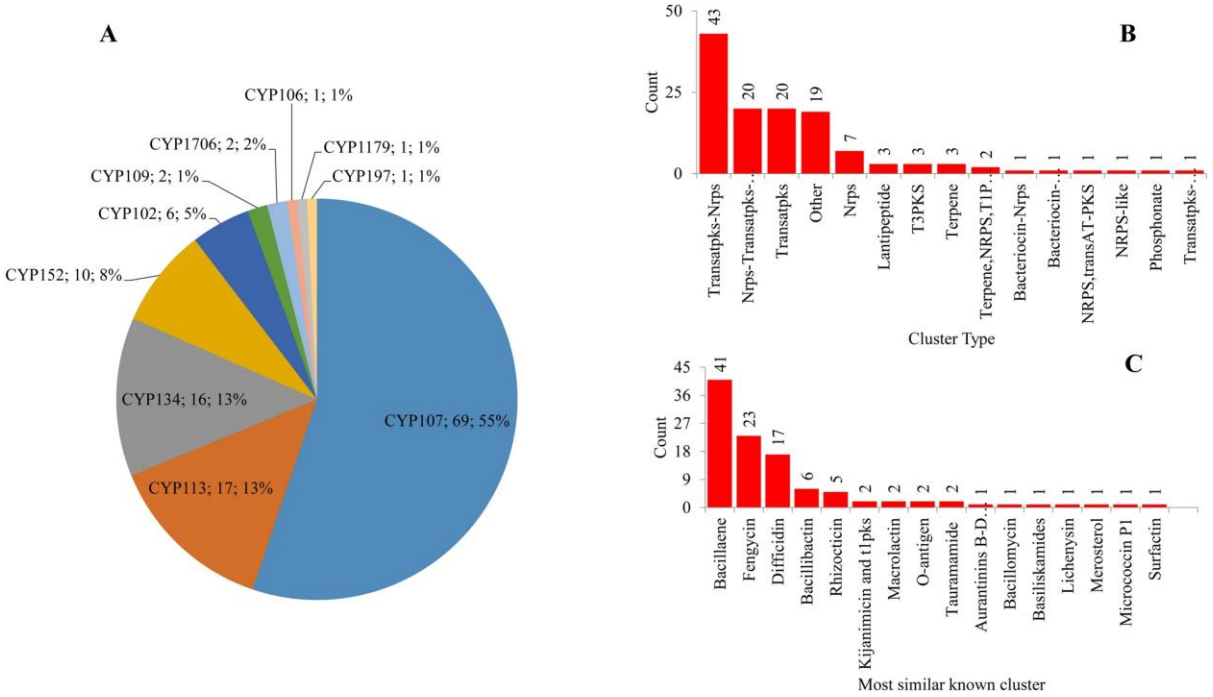


Figure 3.4. Comparative analysis of P450s associated with secondary metabolism in *Firmicutes* species. (A) Comparative analysis of P450 families that are part of secondary metabolite biosynthetic gene clusters (BGCs). The P450 family name, number of P450s and their percentage of the total number of P450s part of BGCs are presented in the figure. (B) Comparative analysis of types of BGCs. The number at the top of each bar represents the number of P450s in the type of BGC. (C) Comparative analysis of most similar known clusters that have P450s. The number at the top of each bar represents the total number of similar clusters

3.2.3 A small proportion of P450s are involved in secondary metabolism in *Firmicutes* species

Among 712 P450s identified in 229 *Firmicutes* species, 125 P450s (18%) of 62 *Firmicutes* species were found to be part of secondary metabolite BGCs (Fig. 11). P450s that were part of BGCs were from species belonging to the subphylum *Bacilli* and most of these species belonged to the genera *Bacillus* (49 species). Most of the *Firmicutes* species P450 families were found to be part of BGCs (Fig. 11A). Among 14 P450 families, ten families, namely CYP107, CYP113, CYP134, CYP152, CYP102, CYP109, CYP1706, CYP106, CYP1179 and CYP197, were found to be part of different secondary metabolite BGCs (Fig. 11A). Among these families, P450s belonging to the CYP107 family were dominantly present in BGCs with more than half of the P450s (55%) being apart of BGCs (Fig. 11A). A point to be noted is that P450 families such as CYP107, CYP152, CYP102 and CYP109 are expanded in *Firmicutes* and part of the BGCs also clearly support a previous hypothesis that "species populate specific P450s if they are useful in their adaptation to certain ecological niches or useful in their physiology" (Ngwenya et al., 2018, Qhanya et al., 2015, Jawallapersand et al., 2014, Syed K, 2014, Suzuki et al., 2012, Akapo et al., 2019, Sello et al., 2015, Feyereisen, 2012). Interestingly, some P450 families such as CYP113, CYP134, CYP197, CYP1706 and CYP1179 that are scarcely present in *Firmicutes* species, but they are also part of BGCs, indicating their important role in producing these metabolites.

An interesting phenomenon is observed when comparing at the subfamily level where P450 subfamilies "H" and "K" are part of BGCs and the subfamily "J" P450s are not part of BGCs despite this subfamily's dominance in the CYP107 family (Table 4), indicating subfamily level selectivity by species. P450s were found to be part of 15 BGC types (Fig. 11B). Most of the P450s are part of BGC types such as Transatpks-Nrps (Trans-AT polyketide synthase-Non-ribosomal peptide synthetase cluster) (43 P450s), followed by Nrps-Transatpks-Otherks (Nrps-Transatpks-Other types of PKS cluster) and Transatpks (each 20 P450s) and Other (19 P450s) (Fig. 11B). Analysis of most similar known clusters revealed that the majority of P450 clusters were bacillaene (41 clusters), followed by fengycin (23 clusters) and diffcidin (17 clusters) (Fig. 11C). Analysis of the association between P450 families and BGCs showed that CYP107 family P450s are mostly associated with BGCs Nrps-Transatpks Otherks, Transatpks, Transatpks-Nrps and Other, a putative gene cluster; CYP113 family P450s are associated with Transatpks-Nrps and Transatpks BGC, and CYP134 family P450s are associated with Transatpks-Nrps and Other. The percentage

identity of *Firmicutes* BGCs containing P450s with similar known clusters revealed that most of the clusters have 100% sequence identity, indicating that these clusters indeed produce the expected metabolites.

Comparative analysis of P450s involved in secondary metabolism among different bacterial species revealed that more P450s from *Firmicutes* species (18%) are involved in secondary metabolism compared to the P450s from mycobacterial (11%) and cyanobacterial (8%) species (Table 3). However, P450s from *Firmicutes* species are only second to *Streptomyces* species P450s (23%) in terms of their involvement in secondary metabolism (Table 3). These results strongly support the recent observation that more P450s are involved in secondary metabolism in *Streptomyces* species compared to other bacterial species (Mnguni et al., 2020) and it is no surprise that two thirds of all known antibiotics in the world come from these species (Procópio et al., 2012).

3.2.4 Most *Firmicutes* species P450s are orphans of unknown function.

Among the 712 *Firmicutes* species P450s, only a handful of P450s are characterized for their physiological functions. The well-known and well-studied P450 CYP102A1 (P450-BM3) from *B. megaterium* is found to be a fatty acid hydroxylase (Munro et al., 2002, Li and Poulos, 1997, Noble et al., 1999). CYP152A1 from *B. subtilis* and CYP152K6 from *B. methanolicus* were found to be peroxygenases that use hydrogen peroxide to drive hydroxylation and decarboxylation of fatty acids (Lee et al., 2003, Lee et al., 2002, Girvan et al., 2018). CYP107H1 from *B. subtilis* is a pimelic acid hydroxylase involved in biotin synthesis (Bower et al., 1996, Cryle and Schlichting, 2008). CYP134A1 from *B. subtilis* is involved in synthesis of pulcherriminic acid, a natural product, by three-step oxidative transformation of the diketopiperazine cyclo-l-leucyl-l-leucyl (Cryle et al., 2010). Based on the P450s' location in different BGCs and their percentage identity with known similar clusters (most of them have 100% sequence identity), we predict 89 P450 functions in the synthesis of diferent secondary metabolites such as polyketides (macrolactin, bacillaene, and diffcidin), lipopeptides (surfactin, lichenysin and fengycin), phosphono-oligopeptides (rhizocticin) and siderophores (Bacillibactin) (Table 3.3). These secondary metabolites are well known for their potential biotechnological applications such as antibacterial, antiviral and cytotoxic properties as listed in Table 3.3.

Analysis of association of P450s with BGCs revealed that specific P450 orthologues are involved in the production of the same secondary metabolite (Table 3.3), indicating

horizontal gene transfer of these BGCs among *Bacillus* species, a well-known phenomenon of gene-cluster transfer between bacterial species (Cimermancic et al., 2014, Tran et al., 2019, Medema et al., 2015). Prominent observations include CYP107K P450s' involvement in biosynthesis of bacillaene, CYP113L1 P450s' involvement in the biosynthesis of diffcidin and CYP107H4 P450s' involvement in the biosynthesis of fengycin (Table 3.3). Considering the high conservation of ortholog P450s, it can be assumed that BGCs such as bacillaene, diffcidin and fengycin populated among *Bacillus* species via horizontal gene-cluster transfer.

An interesting comparison can be drawn between the role of secondary metabolites produced by *Firmicutes* where P450s are involved (Table 3.3) and the role of secondary metabolites produced by *Streptomyces* species (Senate et al., 2019, Mnguni et al., 2020). In both cases, these secondary metabolites seem to be helping organisms gain the upper hand in their ecological niches, as the secondary metabolites produced by these organisms have anti-bacterial, anti-fungal and anti-viral properties, suggesting that these compounds help both the *Firmicutes* (Caulier et al., 2019) and the *Streptomyces* (Senate et al., 2019, Mnguni et al., 2020) species to thrive in their ecological niches by eliminating other organisms. The structure and biological functions of bacillaene, diffcidin, fengycin, lichenysin, macrolactin, rhizocticin and surfactin produced by *Firmicutes* were recently reviewed (Caulier et al., 2019).

Table 3.3. Functional prediction of *Firmicutes* species P450s' involvement in synthesis of secondary metabolites

P450	Species name	Predicted biological role
CYP134A1	*Bacillus subtilis* subsp. *subtilis* 168	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *subtilis* RO-NN-1	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *subtilis* 6051-HGW	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *subtilis* BAB-1	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *subtilis* AG1839	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *subtilis* JH642	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *subtilis* OH 131.1	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *spizizenii* W23	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *spizizenii* TU-B-10	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* BSn5	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* QB928	Bacillaene biosynthesis

CYP107K1	*Bacillus subtilis* XF-1	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* PY79	Bacillaene biosynthesis
CYP107K1	*Bacillus* sp. JS	Bacillaene biosynthesis
CYP107K1	*Bacillus* sp. YP1	Bacillaene biosynthesis
CYP107K1	*Bacillus* sp. BS34A	Bacillaene biosynthesis
CYP107K1	*Bacillus* sp. LM 4-2	Bacillaene biosynthesis
CYP107K1	*Bacillus subtilis* subsp. *subtilis* BSP1	Bacillaene biosynthesis
CYP107K1	*Bacillus gibsonii*	Bacillaene biosynthesis
CYP107K2	*Bacillus atrophaeus* NRS 1221A	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* FZB42	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* CAU B946	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* YAU B9601-Y2	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* UCMB5036	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* UCMB5033	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* UCMB5113	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* NAU-B3	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* TrigoCor1448	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis* SQR9	Bacillaene biosynthesis
CYP107K3	*Bacillus velezensis*	Bacillaene biosynthesis
CYP107K3	*Bacillus amyloliquefaciens* DSM 7	Bacillaene biosynthesis
CYP107K3	*Bacillus amyloliquefaciens* TA208	Bacillaene biosynthesis
CYP107K3	*Bacillus amyloliquefaciens* LL3	Bacillaene biosynthesis
CYP107K3	*Bacillus amyloliquefaciens* XH7	Bacillaene biosynthesis
CYP107K3	*Bacillus amyloliquefaciens* Y2	Bacillaene biosynthesis
CYP107K3	*Bacillus amyloliquefaciens* IT-45	Bacillaene biosynthesis
CYP107K3	*Bacillus amyloliquefaciens* LFB112	Bacillaene biosynthesis
CYP107K3	*Bacillus vallismortis*	Bacillaene biosynthesis
CYP107K3	*Bacillus* sp. Pc3	Bacillaene biosynthesis
CYP107K3	*Bacillus* sp. BH072	Bacillaene biosynthesis
CYP107K3	*Bacillus* sp. SDLI1	Bacillaene biosynthesis
CYP152K2	*Solibacillus silvestris* DSM 12223	Bacillibactin biosynthesis
CYP113L1	*Bacillus* sp. Pc3	Bacillibactin biosynthesis

CYP113L1	*Bacillus velezensis* FZB42	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* CAU B946	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* YAU B9601-Y2	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* AS43.3	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* UCMB5036	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* UCMB5033	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* UCMB5113	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* NAU-B3	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis* SQR9	Diffcidin biosynthesis
CYP113L1	*Bacillus velezensis*	Diffcidin biosynthesis
CYP113L1	*Bacillus amyloliquefaciens* Y2	Diffcidin biosynthesis
CYP113L1	*Bacillus amyloliquefaciens* IT-45	Diffcidin biosynthesis
CYP113L1	*Bacillus amyloliquefaciens* CC178	Diffcidin biosynthesis
CYP113L1	*Bacillus amyloliquefaciens* LFB112	Diffcidin biosynthesis
CYP113L1	*Bacillus vallismortis*	Diffcidin biosynthesis
CYP113L1	*Bacillus* sp. SDLI1	Diffcidin biosynthesis
CYP107H4	*Bacillus* sp. BH072	Diffcidin biosynthesis
CYP107H4	*Bacillus velezensis* FZB42	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* CAU B946	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* YAU B9601-Y2	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* UCMB5036	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* UCMB5033	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* UCMB5113	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* NAU-B3	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* TrigoCor1448	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis* SQR9	Fengycin biosynthesis
CYP107H4	*Bacillus velezensis*	Fengycin biosynthesis
CYP107H4	*Bacillus amyloliquefaciens* Y2	Fengycin biosynthesis
CYP107H4	*Bacillus amyloliquefaciens* IT-45	Fengycin biosynthesis
CYP107H4	*Bacillus amyloliquefaciens* CC178	Fengycin biosynthesis
CYP107H4	*Bacillus amyloliquefaciens* LFB112	Fengycin biosynthesis
CYP107H4	*Bacillus vallismortis*	Fengycin biosynthesis

CYP107H4	*Bacillus* sp. Pc3	Fengycin biosynthesis
CYP107H4	*Bacillus* sp. BH072	Fengycin biosynthesis
CYP107H4	*Bacillus* sp. SDLI1	Fengycin biosynthesis
CYP113L1	*Bacillus velezensis* AS43.3	Fengycin biosynthesis
CYP107H2	*Bacillus amyloliquefaciens* DSM 7	Fengycin biosynthesis
CYP107H2	*Bacillus amyloliquefaciens* TA208	Fengycin biosynthesis
CYP107H2	*Bacillus amyloliquefaciens* LL3	Fengycin biosynthesis
CYP107H2	*Bacillus amyloliquefaciens* XH7	Fengycin biosynthesis
CYP1179A4	*Bacillus xiamenensis*	Lichenysin biosynthesis
CYP1179A4	*Bacillus altitudinis*	Lichenysin biosynthesis
CYP107K3	*Bacillus velezensis* YAU B9601-Y2	Macrolactin biosynthesis
CYP107K3	*Bacillus velezensis* AS43.3	Macrolactin biosynthesis
CYP152A1	*Bacillus subtilis* subsp. *spizizenii* W23	Rhizocticin biosynthesis
CYP152A9	*Bacillus atrophaeus* 1942	Surfactin biosynthesis

3.3 Methods

3.3.1 Species and database.

Nine hundred and seventy-two *Firmicutes* species genomes publicly available at Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2015) were used in this study.

3.3.2 Genome data mining and annotation of P450s.

Genome data mining of P450s in *Firmicutes* species was carried out following the standard method described elsewhere (Mthethwa et al., 2018, Senate et al., 2019, Parvez et al., 2016, Khumalo et al., 2020). Briefly, using the information and databases, the whole proteome of each *Firmicutes* species was downloaded and submitted to the NCBI Batch Web CD-Search Tool (Marchler-Bauer et al., 2017). Proteins that were assigned as P450 superfamily were chosen and assessed for the presence of characteristics EXXR and CXG motifs (Syed and Mashele, 2014, Graham and Peterson, 1999). Proteins having both motifs were selected for annotation. Annotation of P450s was accomplished following the International P450 Nomenclature Committee rules (Nelson et al., 1993, Nelson, 1998, Nelson, 2006). Proteins with an identity >40% were classified under the same family and proteins with an identity >55% were classified under the same subfamily. Proteins with less than 40% identity were

assigned to a new P450 family. P450s and gene cluster data for *Bacillus* species were retrieved from a published article and used in the study (Mthethwa et al., 2018).

3.3.3 Phylogenetic analysis of P450s.

The phylogenetic tree of *Firmicutes* species P450s was constructed following the method described elsewhere (Mthethwa et al., 2018, Senate et al., 2019, Parvez et al., 2016, Khumalo et al., 2020). Briefly, the *Firmicutes* species P450 protein sequences were aligned using the online database, MAFFT (Katoh et al., 2005). Then the alignments were automatically subjected to tree inferring and optimization by the Trex web server (Boc et al., 2012). Finally, the best-inferred trees were visualized, coloured, and generated by iTOL (Letunic and Bork, 2019).

3.3.4 Generation of P450 profile heat-maps.

Heat-map profiles for P450s were constructed following the method described elsewhere (Senate et al., 2019, Parvez et al., 2016). The heat-map profile shows the presence or absence of P450s in *Firmicutes* species. The data were represented as −3 for family absence (green) and 3 for family presence (red). A tab-delimited file was loaded into MeV (Multi-experiment viewer) using a two-colour array (Saeed et al., 2003). Hierarchical clustering using a Euclidean distance metric was used to cluster the data. Fourteen CYP families formed the vertical axis and 229 *Firmicutes* species formed the horizontal axis.

3.3.5 Secondary metabolite BGCs analysis and P450s identification.

Secondary metabolite BGCs analysis and P450 identification in *Firmicutes* species were carried out following the method described elsewhere (Senate et al., 2019, Parvez et al., 2016), Briefly, each *Firmicutes* species genome ID was submitted to anti-SMASH (Weber et al., 2015) for the identification of secondary metabolite BGCs. The results were downloaded both in the form of gene cluster sequences and Excel spreadsheets. All gene cluster sequences for each species were downloaded and captured in a separate Word file. Using these data, P450s that were part of a specific gene cluster were identified. Standard gene cluster abbreviation terminology available at the anti-SMASH database (Weber et al., 2015) was maintained in this study.

3.3.6 Bacterial P450s and gene cluster data.

P450s and their BGCs data for *Streptomyces* species (Senate et al., 2019, Mnguni et al., 2020), mycobacterial species (Parvez et al., 2016) and cyanobacterial species (Khumalo et al., 2020) were retrieved from published articles and used for comparative analysis.

3.3.7 Functional prediction of P450s.

Functional prediction of P450s was carried out based on their location in a particular BGC and the percentage sequence identity of BGCs with characterized gene clusters where more than 70% identity was set as a cut-of value. However, in the case of *Firmicutes* species most BGCs have 100% sequence identity with the characterized gene cluster, indicating P450s' definite role in the production of a specific secondary metabolite.

CHAPTER 4: COMPARATIVE ANALYSIS OF FERREDOXINS IN THE BACTERIAL PHYLUM *FIRMICUTES* WITH OTHER DOMAINS OF LIFE

4.1 Introduction

Iron sulfur (Fe-S) clusters are known as the simplest electron transfer groups and the first developed clusters during chemical evolution (Cammack, 1983). 4Fe-4S clusters consist of a mixture of iron sulfide and thiol of cysteine peptide (Bruschi and Guerlesquin, 1988). These proteins are involved in various processes such as steroid hydroxylation, carbon dioxide and hydrogen fixation, photosynthesis and oxidative phosphorylation (Bruschi and Guerlesquin, 1988). Such a protein was isolated from *Clostridium pasteurianum* belonging to the phylum *Firmicutes* and named ferredoxin (Mortenson et al., 1962). Ferredoxins are small proteins that bind to inorganic clusters, organized around two to four iron atoms and a number of sulfur atoms (Elsen et al., 2010). Many genes encoding these proteins are found in bacteria. Ferredoxins have similar characteristics, namely having the ability to transfer electrons in biological reactions with non-heme iron and sulfur being present in equivalent amounts (Bruschi and Guerlesquin, 1988).

Ferredoxins can be grouped according to their structure, namely eight iron, seven iron, four iron, three iron and two iron ferredoxins. Eight iron ferredoxins are found in sulfate-reducing bacteria and photosynthetic bacteria but are mainly found in anaerobic fermenting bacteria (Bruschi and Guerlesquin, 1988). Species belonging to the sub-phylum *Clostridia* in the phylum *Firmicutes* are anaerobic bacteria that require ferredoxins for their metabolism. These ferredoxins can be identified by the presence of eight cysteine residues binding to two 4Fe-4S clusters (Bruschi and Guerlesquin, 1988). Four iron ferredoxins are very similar to the above ferredoxin in terms of properties; however, they have one less cluster and are therefore characterized by the presence of four or six cysteine residues (Bruschi and Guerlesquin, 1988). These proteins have been isolated from species belonging to the bacterial phylum *Firmicutes* such as *Bacillus polymyxa* (Shethna et al., 1971), *Desulfovibrio gigas* (Bruschi and Couchoud, 1979) and *Desulfovibrio africanus* (Bruschi and Hatchikian, 1982). Three iron ferredoxins are linked to five cysteines and are found in acidic bacteria such as *Bacillus acidocaldarius* (Schlatter et al., 1985). Two iron ferredoxins are found in various bacteria and have a distinct structure consisting of four cysteines (Bruschi and Guerlesquin, 1988) (Fig 4.1).

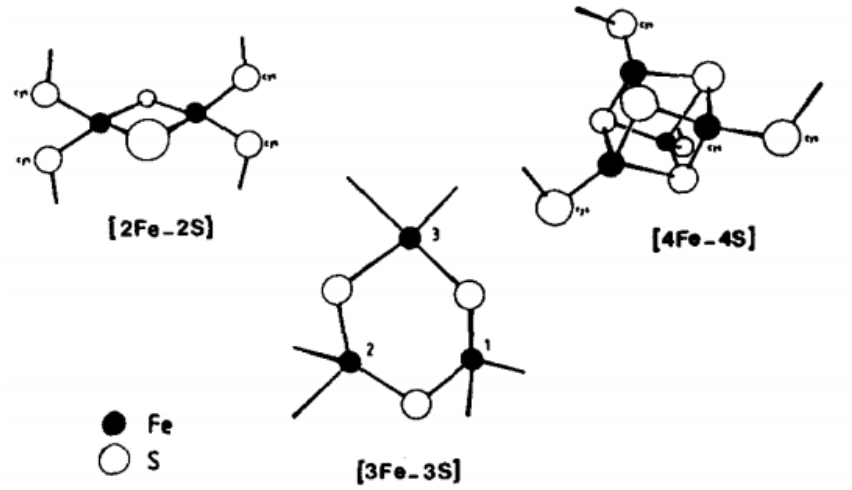


Figure 4.1. Structures of the two-, three- and four-iron cluster types found in bacterial ferredoxins (Bruschi and Guerlesquin, 1988)

Each of the Fe-S cluster types have their own characteristic Fe-S binding motif where the Fe-atom binds to the cysteine amino acid residue. The characteristic Fe-S binding motif includes four cysteines for 2Fe-2S, three cysteines and a proline that follows the third cysteine for 3Fe-4S, four cysteines and a proline that follows the fourth cysteine for 4Fe-4S and the 7Fe-8S ferredoxins having both 3Fe-4S and 4Fe-4S cluster features. The 2[4Fe-4S] ferredoxins have two 4Fe-4S motifs but the spacing between Fe-atom binding cysteines are different compared to the 4Fe-4S cluster type motif. 2[4Fe-4S] proteins are divided into two sub-families with small proteins (approximately 55 amino acids) having isopotential Fe-S clusters and larger proteins known as Alvin (Alv) ferredoxins having clusters with different potentials (Saridakis et al., 2009). Sequence analysis revealed the presence of an extra cysteine in 2[4Fe-4S]Alv ferredoxins exactly after three amino acids following the last cysteine of the second 4Fe-4S binding cluster (Saridakis et al., 2009). Based on the arrangement of cysteines in the Fe-S binding motif, 2Fe-2S cluster proteins are also classified as plant-type, mitochondrial-type, bacterial-type and thioredoxin-type (Zanetti and Pandini, 2013).

It is believed that among Fe-S cluster types, 4Fe-4S clusters are the first to be evolved as the conditions that mimic the primitive Earth resulted in the formation of 4Fe-4S clusters (Beinert et al., 1997; Meyer, 2008; Venkateswara Rao and Holm, 2004). Studies indicated that the 4Fe-4S cluster is sensitive to oxygen compared to the 2Fe-2S cluster (Hsueh et al., 2013, Imlay, 2006, Jagannathan and Golbeck, 2009) and thus, it is predicted that after the Great Oxidation Event organisms preferred 2Fe-2S clusters due to their oxygen tolerance.

This phenomenon is also observed in a study where comparative analysis of 2Fe-2S and 4Fe-4S cluster proteins and flavodoxin proteins across the domains of life was carried out and it has been found that 4Fe-4S cluster type proteins were abundantly present in anaerobic organisms whereas 2Fe-2S cluster type proteins are abundant in aerobic organisms (Campbell et al., 2019). A study dating back to 1966 on observing the symmetrical origin of 2[4Fe-4S] ferredoxins from *C. pasteurianum* by Eck and Dayhoff (1966) led to the proposal that all proteins emerged through the duplication of an even shorter and simpler ancestral peptide (Eck and Dayhoff, 1966). Recently, it has been shown that indeed 2[4Fe-4S] ferredoxins have drifted from their symmetric roots *via* gene duplication followed by mutations (Mutter et al., 2019). Numerous studies reporting the gene duplication events as the source of the growth and diversification of Fe-S proteins have been listed in a recent article (Campbell et al., 2019).

Electron transfer is essential for all living organisms. This transfer is required in a number of biological processes such as building and breaking down large molecules and cellular metabolism (Child et al., 2018). Monooxygenases are specialized enzymes involved in such processes and require electrons to function. Cytochrome P450 monooxygenase enzymes (CYPs/P450s) are highly selective monooxygenases and require two electrons which are generally derived from NADPH and delivered by electron carrier proteins (Child et al., 2018, White and Coon, 1980). Bacterial P450s typically use the class 1 electron transfer system consisting of a NADPH dependent ferredoxin reductase and an iron-sulfur ferredoxin (Child et al., 2018). Self-sufficient P450s are fused to various types of redox proteins and therefore do not require electrons from an external electron-carrier (Hannemann et al., 2007, Hlavica, 2015, Lamb and Waterman, 2013, Sello et al., 2015).

In each bacterium, electron transfer systems can support multiple P450s (Ewen et al., 2009). Therefore, the number of genes in bacteria generally decrease in this order: P450>Ferredoxin>Ferredoxin reductase (Child et al., 2018). It is interesting to know that studies have shown that P450 activity is dependent on choosing the correct ferredoxin (Child et al., 2018). Various factors may affect the interactions between a P450 and its chosen ferredoxin such as electrostatic interactions, redox potentials, distance and evolution (Chiliza et al., 2020). *Mycobacterium Marinum* is used as a model organism for ferredoxin studies. P450s in this organism are found within clusters that encode a ferredoxin reductase and a ferredoxin which displays the class 1 electron transfer system (Child et al., 2018). The ferredoxins in this species do not follow the standard motifs associated with ferredoxins,

namely CXXCXXC(X)nCP for 4Fe-4S and CXXA/GXXC(X)nCP for 3Fe-3S (Conover et al., 1990) (Fig. 4.2). This shows variations in this species and the associated motifs.

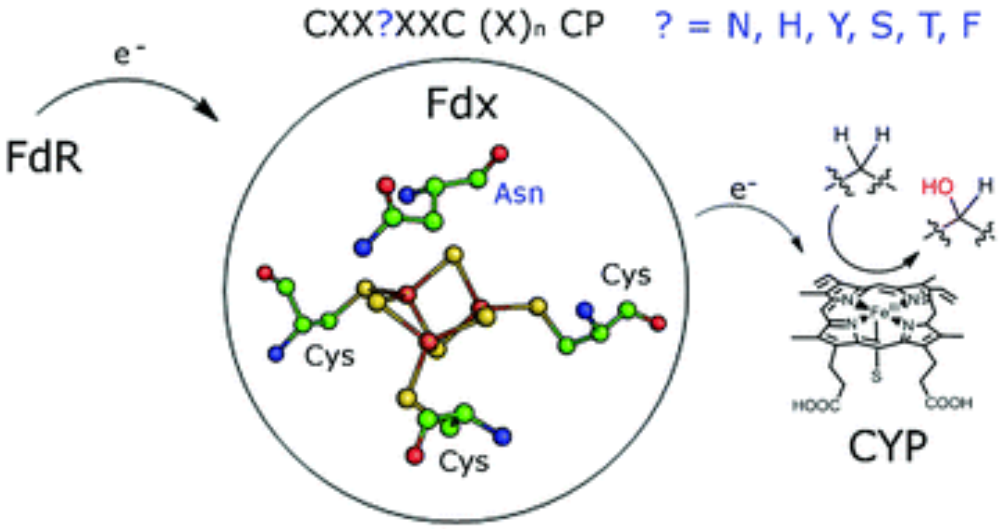


Figure 4.2. Unusual cluster binding motifs found in *Mycobacterium marinum* (Child et al., 2018)

Flavodoxins are small proteins that are similar to ferredoxins in terms of their electron transfer abilities; however, these proteins contain one molecule of non-covalently, tightly bound flavin mononucleotide (FMN) (Sancho, 2006). FMN acts as the redox active component in flavodoxins. They were discovered in the 1960s in *Cyanobacteria* (Smillie, 1965) and *Clostridia* (Knight et al., 1966). In these species flavodoxins were found to replace the iron-containing ferredoxins in various reactions.

Over 2.4 billion years ago, oxygenic photosynthesis evolved on Earth and led to the rise of oxygen levels in the atmosphere (Lyons et al., 2014). Soluble Fe^{2+} was oxidized to Fe^{3+}. This caused a drastic decrease in the bioavailability of iron and thus led to iron-poor environments. Iron can be highly toxic in its unbound state as it leads to the production of dangerous hydroxyl radicals from hydrogen peroxide (Kehrer, 2000). Flavodoxins became more popular during these times as they use FMN as a co-factor rather than iron (Ilbert and Bonnefoy, 2013; Pierella Karlusich et al., 2014; Goñi et al., 2008).

Apart from being used as a model for protein folding and binding (Sancho, 2006), flavodoxins are used to determine low iron environments which are critical in organisms such

as phytoplankton (Inda and Luisa Peleato, 2003) and they can be used as potential drug targets in pathogenic bacteria (Cremades et al., 2005).

The transfer of electrons from electron carrier proteins/redox partners to a P450 is a key step in the P450 catalytic cycle (Sevrioukova et al., 1999). As mentioned above, few ferredoxins have been isolated from species belonging to the bacterial phylum *Firmicutes* Apart from these studies, ferredoxin profiles in *Firmicutes* species have not been reported. This study is aimed to address this research gap by performing genome data mining, annotation and phylogenetic analysis of ferredoxins in *Firmicutes* species. In addition to this, comparative analysis of *Firmicutes* species ferredoxins with species belonging to other domains of life was also carried out.

4.2 Materials and methods

4.2.1 Species and database

In this study, 227 *Firmicutes* species genomes that are available for public use at Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2019) were used. These species are known to have P450s in their genome (Padayachee et al., 2020) and thus selected for ferredoxins analysis considering P450s need ferredoxins for their function and so as one can expect the presence of a greater number of ferredoxins in these species.

4.2.2 Genome data mining and annotation of ferredoxins

Ferredoxins belonging to different Fe-S cluster types reported in the literature (Table 4.1) were used as reference proteins. Protein BLAST (Altschul et al., 1990) was performed using these reference proteins against each of the bacterial species genomes at KEGG (Kanehisa et al., 2019). The hit proteins were manually checked for the presence of ferredoxin Fe-S cluster type canonical motifs and the proteins that had the Fe-S binding motifs were selected. The selected proteins were then subjected to protein BLAST (Altschul et al., 1990) at National Center for Biotechnology and Information (NCBI) (Sayers et al., 2021) against Protein Data Bank (PDB) database (consortium, 2019) and analyzed for the presence of ferredoxins characteristic motif at Pfam database (El-Gebali et al., 2019), InterPro database (Mitchell et al., 2019) and NCBI Conserved Domains Database (CDD) (Lu et al., 2020). Proteins that had a hit against ferredoxins at PDB database and had the ferredoxin motifs as indicated by different databases were selected for further analysis. The rationale for using the PDB database as the priority to sort hit proteins is that the BLAST against this database not only helped in identifying ferredoxin proteins but also helped in accurately identifying Fe-atom

binding cysteine amino acids based on the alignment of hit protein sequences with crystallized ferredoxin sequences. In addition to this, in some cases, BLAST against PDB database also helped to sort proteins accurately whether they are 3Fe-4S or 4Fe-4S cluster proteins and 7Fe-8S or 2[4Fe-4S] cluster proteins. Hit proteins that were found to be dehydrogenases, oxidases or reductases with Fe-S binding motif was not considered as ferredoxins as the Fe-S clusters in these proteins exceeded more than two and they belong to a different protein family. Furthermore, sequences that had hits to ferredoxins but not in full-length were also not considered for further analysis. I also noted that manual sorting of Fe-S cluster proteins into different types by consulting as many databases as possible such as PDB database (consortium, 2019), Pfam database (El-Gebali et al., 2019), InterPro database (Mitchell et al., 2019) and NCBI CDD (Lu et al., 2020) is as important for accurate assigning of Fe-S proteins to different types and also accurately identifying the Fe-atom binding cysteine amino acids. In this study, utmost care has been taken to select only ferredoxins for further analysis.

Table 4.1. Information on ferredoxins that are used as reference proteins in the study

2Fe-2S		
GenBank accession number or PDB code	**Species name**	**Reference**
NP_004100.1 (Adrenodoxin)	*Homo sapiens*	(Cai et al., 2017)
1PDX (Putidaredoxin)	*Pseudomonas putida*	(Pochapsky et al., 1999)
ABB56370.1	*Synechococcus elongatus* PCC 7942 = FACHB-805	(Koksharova et al., 2006)
ABB56928.1	*Synechococcus elongatus* PCC 7942 = FACHB-805	(Koksharova et al., 2006)
WP_013424358.1 (FraEuI1c_3227)	*Frankia* sp. EuI1c (*Frankia inefficax* sp.)	(Lau et al., 2019)
WP_012394830.1 (mmi:MMAR_3155)	*Mycobacterium marinum*	(Child et al., 2018)
3Fe-4S/4Fe-4S		
CAB59502.1	*Streptomyces coelicolor* A3(2)	(Bentley et al., 2002)
WP_013425251.1 (FraEuI1c_4132)	*Frankia* sp. EuI1c (*Frankia inefficax* sp.)	(Lau et al., 2019)
WP_013426476.1 (FraEuI1c_5370)	*Frankia* sp. EuI1c (*Frankia inefficax* sp.)	(Lau et al., 2019)
WP_011740769.1 (Mmar_2879)	*Mycobacterium marinum*	(Child et al., 2018)
WP_012395565.1 (Mmar_3973)	*Mycobacterium marinum*	(Child et al., 2018)
WP_012396301.1 (Mmar_4763)	*Mycobacterium marinum*	(Child et al., 2018)

NP_215277.1 (FdX-Rv0763c)	*Mycobacterium tuberculosis* H37Rv	(Ortega Ugalde et al., 2018, McLean et al., 2006)
NP_216302.1 (FdxE-Rv1786)	*Mycobacterium tuberculosis* H37Rv	(Ortega Ugalde et al., 2018, McLean et al., 2006)
ABB57779.1	*Synechococcus elongatus* PCC 7942 = FACHB-805	(Koksharova et al., 2006)
7Fe-8S		
NP_215693 (FdxC-Rv1177),	*Mycobacterium tuberculosis* H37Rv	(Ortega Ugalde et al., 2018)
NP_216523.1 (FdxA-Rv2007c)	*Mycobacterium tuberculosis* H37Rv	(Ortega Ugalde et al., 2018)
2VKR	*Acidianus ambivalens*	(Frazão et al., 2008)
1H98	*Thermus thermophilus*	(Macedo-Ribeiro et al., 2001)
ABB56846.1	*Synechococcus elongatus* PCC 7942 = FACHB-805	(Koksharova et al., 2006)
2[4Fe-4S]		
2ZVS	*Escherichia coli* K-12	(Saridakis et al., 2009)
2FDN	*Clostridium acidurici*	(Dauter et al., 1997)
WP_013068980.1 (FDI)	*Rhodobacter capsulatus*	(Saeki et al., 1991)
2[4Fe-4S]Alv		
1BLU	*Allochromatium vinosum*	(Moulis et al., 1996)
2FGO	*Pseudomonas aeruginosa*	(Dauter et al., 1997)
WP_023923722.1 (FDIII)	*Rhodobacter capsulatus*	(Saeki et al., 1991)

1RGV	*Thauera aromatica* K172	(Unciuleac et al., 2004)

4.2.3 Ferredoxins subtype classification and nomenclature

Ferredoxins selected under different Fe-S cluster types were then subjected to multiple sequence alignment (MSA) at Clustal Omega database (Madeira et al., 2019). Based on the MSA patterns proteins were seperated into different groups to such an extent where there will be no amino acid gaps between the cysteine amino acids of the Fe-S cluster binding motif. Ferredoxins that are not aligned with other proteins were then subjected to analysis of spacing between cysteine amino acids of the Fe-S cluster binding motif individually. Then the spacing patterns between cysteine amino acids of the Fe-S cluster binding motif was presented as the characteristic signature of a subtype. Analysis of ferredoxins revealed that a proline amino acid is not always conserved in 2[4Fe-4S] ferredoxins and thus is not included in the Fe-S binding motif signature for this type of ferredoxins. Although proline is included for 7Fe-8S cluster proteins, only cysteine residues and the amino acid spacing between these residues can be taken as a signature as non-conservation of proline will be expected when a greater number of ferredoxins are analyzed. To represent different subtypes in a type, a nomenclature system was developed such that each protein starts with its Fe-S cluster type followed by abbreviation "ST" for "subtype" and then a numerical value indicating its subtype number in a type (Fig. 4.3). Subtype numbers are assigned following the occurrence of new ferredoxins along the study. Ferredoxins identified in the bacterial species of *Firmicutes* is represented in Supplementary dataset 2.

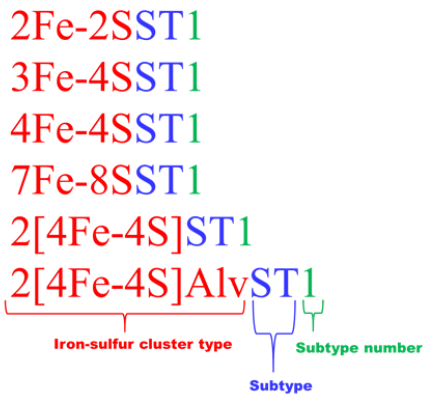


Figure 4.3. Ferredoxin nomenclature based on the spacing between the cysteine amino acids of Fe-S cluster binding motif. Ferredoxins start with their Fe-S cluster type, followed by ST indicating subtype and then a numeric indicating its subtype number in that type. Proteins

grouped into different subtypes have the same characteristic spacing between the cysteine amino acids of the Fe-S cluster binding motif

4.2.4 Assigning Fe-S cluster subtypes to the ferredoxins retrieved from literature

Ferredoxins that are reported in the literature were collected by going through the published articles and also data mining at PDB database (consortium, 2019). Most of the ferredoxins belonging to the species of Archaea and some Eukarya is retrieved from articles published by Campbell and co-workers. The search at PDB database was carried out using Fe-S cluster type names and the ferredoxin sequences were collected after manually looking at the Fe-S clusters of the ferredoxin. These ferredoxins were then subjected to subtype nomenclature as described in the section 4.2.3. Ferredoxins retrieved from literature and data mined at PDB database were presented along with their nomenclature in Supplementary dataset 3.

4.2.5 Phylogenetic analysis of ferredoxins

Phylogenetic analysis of ferredoxins was carried out following the procedure described recently by our laboratory (Khumalo et al., 2020, Padayachee et al., 2020). The phylogenetic tree of ferredoxins was constructed using protein sequences. Firstly, the MAFFT v6.864 program (Katoh et al., 2005) was used to align the protein sequences that are part of the Trex web server (Boc et al., 2012). The alignments were then subjected to interpret the best tree by the Trex web server (Boc et al., 2012). Lastly, a web-based tool, iTOL, was used to create, visualize and colour the tree (Letunic and Bork, 2016).

4.2.6 Generation of ferredoxin subtype profile heat maps

Ferredoxin subtype profile heat maps were generated following the procedure described recently by our laboratory (Khumalo et al., 2020, Padayachee et al., 2020). The heat map was generated using the ferredoxin subtype data to show the presence or absence of ferredoxin subtypes in three domains of life, Archaea, Bacteria and Eukarya. The data was represented as (−3) for subtype absence (green) and (3) for subtype presence (red). A tab-delimited file was imported into MeV (Multi-experiment viewer) (Howe et al., 2011). Hierarchical clustering using a Euclidean distance metric was used to cluster the data. Ferredoxin subtypes formed the horizontal axis, and three domains of life formed the vertical axis.

4.3 Results and discussion

4.3.1 *Firmicutes* species have different Fe-S cluster type ferredoxins in their genomes

Genome data mining of 227 *Firmicutes* species revealed the presence of 281 ferredoxins in their genomes (Figure 4.4). Among the *Firmicutes* species *Kyrpidia spormannii* have the

highest number of six ferredoxins. Four Fe-S cluster type ferredoxins were found in *Firmicutes* species (Fig. 4.4). The 3Fe-4S and 2[4Fe-4S]Alv cluster type ferredoxins are not identified in the *Firmicutes* species analyzed in this study.

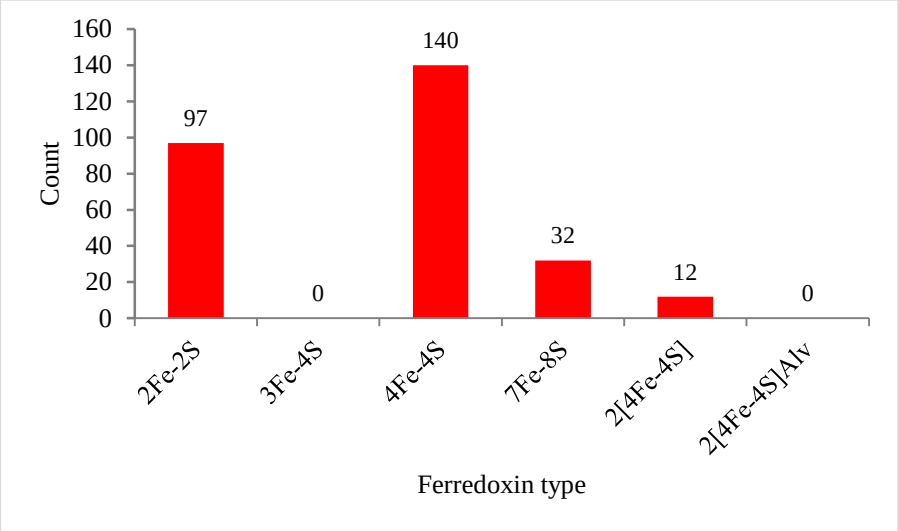


Figure 4.4. Comparative analysis of ferredoxins iron-sulfur (Fe-S) cluster in *Firmicutes* species. The number on top of the bar indicates the count for that ferredoxin type

Analysis of ferredoxins Fe-S cluster types in *Firmicutes* species revealed that 4Fe-4S was the most abundant ferredoxin type with 140 ferredoxins followed by 2Fe-2S with 97 ferredoxins, 7Fe-8S with 32 ferredoxins and 2[4Fe-4S] with 12 ferredoxins (Fig. 4.4). A point to be noted is that this studies results on the presence of a large number of 4Fe-4S cluster type ferredoxins in *Firmicutes* species corroborate with a previous study where only 2Fe-2S and 4Fe-4S cluster types were analyzed as part of global protein electron carrier proteins analysis across the domain of life (Campbell et al., 2019).

4.3.2. Ferredoxins diversity analysis in *Firmicutes* species

In this study, the amino acid spacing pattern between the cysteine amino acids of Fe-S cluster binding motif as a signature for ferredoxins subtype classification has been tested to understand the diversity of ferredoxins using *Firmicutes* species as model organisms. Analysis of ferredoxin Fe-S cluster subtypes revealed that *Firmicutes* species have highly diverse ferredoxins in their genomes especially 2Fe-2S and 4Fe-4S types. This suggests that the subtype classification clarifies the diversity of ferredoxins within a cluster type. A detailed analysis on ferredoxin Fe-S cluster subtypes in *Firmicutes* species is presented below:

4.3.2.1 2Fe-2S

Based on the amino acid spacing pattern between the cysteine amino acids of the Fe-S cluster binding motif, 97 2Fe-2S ferredoxins of *Firmicutes* species can be grouped into 11 subtypes (Table 4.2). *Firmicutes* species have the highest number of subtype 20 ferredoxins in their genomes followed by subtype 18 (Table 4.2). 2Fe-2S proteins are less oxygen sensitive compared to 4Fe-4S redox proteins (Imlay, 2006, Jagannathan and Golbeck, 2009). 4Fe-4S redox proteins are more susceptible to destruction due to an exposed sulfido atom, however 2Fe-2S proteins contain a hydrophobic patch that covers the metallocluster and are protected from degradation (Pierella Karlusich et al., 2014). Therefore, 2Fe-2S redox proteins are highly favoured. This may explain the large number of 2Fe-2S proteins found in *Firmicutes*.

Table 4.2. Subtype classification of ferredoxins across the domains of life. Ferredoxins collected from species belonging to different biological kingdoms that are reported in the literature and data mined at PDB are also presented in the table as per their subtypes under the domains: Archaea, Bacteria and Eukarya

Subtypes	Cysteine spacing signature	No. AA	No of ferredoxins			
			Firmicutes	From literature		
				Archaea	Bacteria	Eukarya
2Fe-2S						
Subtype 1	$CX_5CX_2CX_{36}C$	47			2	64
Subtype 2	$CX_5CX_2CX_{37}C$	48	1		4	22
Subtype 3	$CX_4CX_2CX_{29}C$	39	1	7	11	55
Subtype 4	$CX_5CX_2CX_{35}C$	46	3		2	
Subtype 5	$CX_5CX_2CX_{38}C$	49	1			
Subtype 6	$CX_4CX_2CX_{34}C$	44	11	7	1	
Subtype 7	$CX_4CX_2CX_{51}C$	61			1	
Subtype 8	$CX_4CX_2CX_{31}C$	41	6	1	1	
Subtype 9	$CX_4CX_2CX_{33}C$	43	3	19		2
Subtype	$CX_5CX_2CX_{39}C$	50			1	

10						
Subtype 18	$CX_5CX_2CX_{34}C$	45	20	2	1	1
Subtype 20	$CX_5CX_2CX_{32}C$	43	44		1	5
Subtype 21	$CX_4CX_{31}CX_3C$	42	5			
Subtype 22	$CX_2CX_{41}CX_3C$	50	2			
Subtype 23	$CX_7CX_{38}CX_3C$	52			1	
Subtype 24	$CX_4CX_2CX_{30}C$	40		70		1
Subtype 25	$CX_5CX_2CX_{30}C$	41				1
Subtype 31	$CX_4CX_{36}CX_3C$	47		1		
Subtype 36	$CX_3CX_1CX_{38}C$	46			1	
Subtype 37	$CX_{12}CX_{32}CX_3C$	51			1	
Subtype 38	$CX_4CX_2CX_{28}C$	38		40		3
Subtype 39	$CX_4CX_2CX_{46}C$	56		2		
Subtype 40	$CX_4CX_2CX_{49}C$	59		3		
Subtype 41	$CX_4CX_2CX_{45}C$	55		2		
Subtype 42	$CX_4CX_2CX_{65}C$	75		1		
Subtype 43	$CX_4CX_2CX_{50}C$	60		2		
Subtype 44	$CX_4CX_2CX_{47}C$	57		2		
Subtype 45	$CX_4CX_2CX_{48}C$	58		1		
Subtype 46	$CX_5CX_2CX_{52}C$	63		2		
Subtype 47	$CX_5CX_2CX_{31}C$	42		1		
Subtype 48	$CX_5CX_2CX_{28}C$	39		2		
Subtype 49	$CX_5CX_2CX_{27}C$	38		10		
Subtype 50	$CX_5CX_2CX_{82}C$	93		2		
Subtype 51	$CX_5CX_2CX_{29}C$	40		1		

Subtype 52	$CX_4CX_2CX_{32}C$	42		1		
Subtype 53	$CX_5CX_2CX_{42}C$	53				1
Subtype 54	$CX_4CX_2CX_{22}C$	32				2
Subtype 55	$CX_4CX_2CX_{29}C$	39		2		
4Fe-4S						
Subtype 2	$CX_2CX_2CX_{43}CP$	52	107			
Subtype 3	$CX_2CX_2CX_{45}CP$	54	24		1	
Subtype 4	$CX_2CX_2CX_{37}CP$	46	6			
Subtype 5	$CX_2CX_2CX_{44}CP$	53	2			
Subtype 6	$CX_2CX_2CX_{39}CP$	48	1			
Subtype 7	$CX_2CX_2CX_{36}CP$	45			2	
Subtype 8	$CX_2CX_2CX_{34}CP$	43			1	
Subtype 9	$CX_2CX_2CX_{38}CP$	46		1		1
Subtype 10	$CX_5CX_3CX_{32}CP$	45		1		
Subtype 11	$CX_5CX_3CX_{30}CP$	43		12		
Subtype 12	$CX_5CX_3CX_{31}CP$	44		2		
7Fe-8S						
Subtype 1	$CX_7CX_3CPX_{17}CX_2CX_2CX_3CP^*$	43	32		13	
Subtype 5	$CX_5CX_3CPX_{26}CX_2CX_2CX_3CP$	50		1		
Subtype 6	$CX_5CX_3CPX_{24}CX_2CX_2CX_3CP$	48			1	1
Subtype 7	$CX_{10}CX_3CPX_{17}CX_2CX_2CX_3CP$	46			1	
Subtype 8	$CX_5CX_3CPX_{22}CX_2CX_2CX_3CP$	46		9		
Subtype 9	$CX_5CX_3CPX_{18}CX_2CX_2CX_3C$	42		1		
Subtype 10	$CX_3CX_3CPX_{22}CX_2CX_2CX_3CP$	44		2		
2[4Fe-4S]						
Subtype 2	$CX_2CX_2CX_3CX_{18}CX_2CX_8CX_3C$	46			4	

Subtype 3	$CX_2CX_2CX_3CX_{20}CX_2CX_2CX_3C$	42	2	52	1	
Subtype 8	$CX_2CX_2CX_3CX_{24}CX_2CX_2CX_3C$	46		2		
Subtype 9	$CX_2CX_2CX_3CX_{18}CX_2CX_2CX_3C$	40	6	78	3	1
Subtype 10	$CX_2CX_2CX_3CX_{21}CX_2CX_2CX_3C$	43		2		
Subtype 11	$CX_2CX_2CX_3CX_{18}CX_3CX_2CX_3C$	41		2		
Subtype 12	$CX_2CX_2CX_3CX_{28}CX_2CX_2CX_3C$	50		1		2
Subtype 15	$CX_2CX_2CX_3CX_{19}CX_2CX_2CX_3C$	41	4	58		
Subtype 17	$CX_2CX_2CX_3CX_{29}CX_2CX_2CX_3C$	51				2
Subtype 18	$CX_4CX_2CX_3CX_{18}CX_2CX_2CX_3C$	42		1	1	
Subtype 19	$CX_3CX_2CX_3CX_{20}CX_2CX_2CX_3C$	43			2	
Subtype 20	$CX_2CX_2CX_3CX_{17}CX_2CX_2CX_3C$	39		24	1	
Subtype 21	$CX_3CX_3CX_3CX_{37}CX_1CX_3CX_3C$	61			1	
Subtype 22	$CX_2CX_2CX_3CX_{26}CX_2CX_2CX_3C$	48				6
Subtype 23	$CX_2CX_2CX_3CX_{30}CX_2CX_2CX_3C$	52				1
Subtype 24	$CX_2CX_2CX_3CX_{33}CX_2CX_2CX_3C$	55		22		
Subtype 25	$CX_2CX_2CX_3CX_{32}CX_2CX_2CX_3C$	54		23		
Subtype 26	$CX_2CX_2CX_3CX_{23}CX_2CX_2CX_3C$	45		2		
Subtype 27	$CX_2CX_2CX_3CX_{34}CX_2CX_2CX_3C$	56		1		
Subtype 28	$CX_2CX_2CX_3CX_{14}CX_2CX_2CX_3C$	36		2		
Subtype 29	$CX_2CX_2CX_3CX_{22}CX_2CX_2CX_3C$	44		2		
Subtype 30	$CX_2CX_2CX_2CX_{38}CX_2CX_2CX_3C$	59		1		
Subtype 31	$CX_4CX_2CX_3CX_{19}CX_2CX_2CX_3C$	43		24		
Subtype 32	$CX_5CX_2CX_3CX_{19}CX_2CX_2CX_3C$	44		3		
Subtype 33	$CX_2CX_2CX_3CX_{16}CX_2CX_2CX_3C$	38		16		

4.3.2.2 4Fe-4S

The 140 4Fe-4S ferredoxins found in *Firmicutes* species can be grouped into five different subtypes (Table 4.2). Of the five subtypes, subtype 2 ferredoxins were the dominant subtype in *Firmicutes* species indicating their preference by these species (Table 4.2). Literature suggests that 4Fe-4S was the first ferredoxin found in organisms (Sousa et al., 2013, Campbell et al., 2019). These ferredoxins were found in ancestral anaerobic species such as prokaryotes (Major et al., 2004). *Clostridia* species are known to harbor large amounts of 4Fe-4S proteins as they are anaerobic species and therefore the 4Fe-4S is not susceptible to damage (Campbell et al., 2019), a fair amount of *Clostridia* species was used in this study and can explain the large quantity of these ferredoxins in this phylum.

4.3.2.3 7Fe-8S

All the 32 *Firmicutes* species ferredoxins belongs to the subtype 1 (Table 4.2). This shows that this specific subtype is favoured in these species.

4.3.2.4 2[4Fe-4S]

12 2[4Fe-4S] ferredoxins of *Firmicutes* species can be grouped into three different subtypes (Table 4.2). With the most dominant subtype being subtype 6 followed by subtype 15 and 3 respectively. As previously mentioned, many ferredoxins are substituted with flavodoxins in iron poor environments. *Clostridia* species are known to possess 2[4Fe-4S] proteins (Guerrini et al., 2008). However, in glycolytic *Clostridia* species these proteins are substituted and could explain the low number (Schönheit et al., 1979, Elsen et al., 2010).

4.4 Evolutionary linkage of ferredoxins subtype classification

The subtype classification of ferredoxins proposed in this study is solely based on the arrangement of cysteine amino acids in the Fe-S cluster binding motif and to use this subtype classification as a criterion for assessing other aspects of ferredoxins one should evaluate the evolutionary linkage of subtypes, if any. Thus, here evolutionary analysis of ferredoxins was performed (Fig. 4.5). As shown in the Figure 4.5, except for some ferredoxins, most of the ferredoxins were grouped as per their types and as per their subtypes suggesting the subtype classification criteria certainly followed the evolutionary trend. This indicates that the amino acid spacing pattern between the cysteine amino acids of Fe-S cluster binding motif is evolutionarily conserved. Some interesting aspects that can be drawn from the evolutionary analysis is that, although 4Fe-4S ferredoxins aligned independently on the tree, their placement with 2[4Fe-4S] suggests the domains of 4Fe-4S ferredoxins certainly passed to

2[4Fe-4S] ferredoxins. This observation also strongly supports the hypothesis proposed by Eck and Dayhoff that all proteins emerged through the duplication of an even shorter and simpler ancestral peptide (Eck and Dayhoff, 1966). Some of the ferredoxin Fe-S cluster types branched on the tree and aligned with other Fe-S cluster types indicating they are evolutionarily linked to each other by sharing high sequence similarity in certain parts of the Fe-S cluster domain.

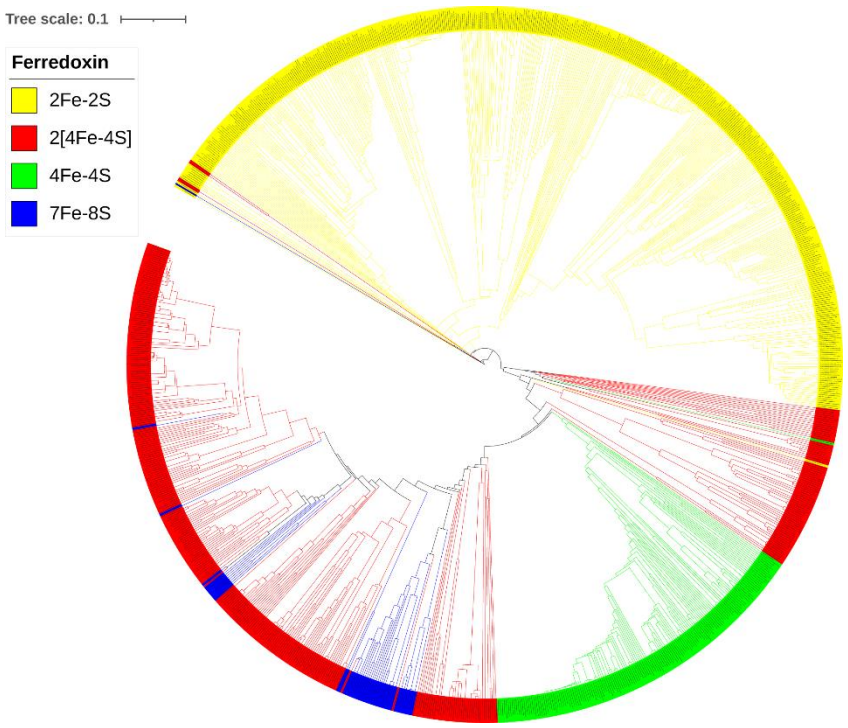


Figure 4.5. Phylogenetic analysis of ferredoxins. Ferredoxins belonging to different Fe-S cluster types were highlighted in different colours and indicated in the figure

4.5 Common ancestral linkage and lateral gene transfer of ferredoxins between *Firmicutes* and other domains of life

The current genomic era resulted in genome sequencing of many organisms and thus the availability of ferredoxin sequences from all domains of life. This has given us an opportunity to investigate common ancestral linkage and lateral gene transfer (LGT) of ferredoxins. Ferredoxins that are common to the *Firmicutes* species and other bacterial species will be a good indication of their common ancestral origin and ferredoxins that are common between prokaryotes (Archaea/Bacteria) and eukaryotes are a good indication of LGT. The subtype classification of ferredoxins as proposed in this study certainly helped in understanding these two aspects on ferredoxins. To understand the evolutionary linkage and

LGT of ferredoxins across the living organisms, 538, 95 and 171 ferredoxins from Archaea, Bacteria (excluding ferredoxins from *Firmicutes*) and Eukarya were collected and annotated, respectively (Table 4.2). Comparative analysis of ferredoxin subtypes revealed common ferredoxins between *Firmicutes* and other domains of life (Fig. 4.6). Amongst the 2Fe-2S subtypes, subtypes 3 and 18 were found in all categories and subtype 9 was found in all categories for 2[4Fe-4S] ferredoxins indicating their common ancestral origin (between Archaea and Bacteria) and LGT from prokaryotes (Archaea and Bacteria) to eukaryotes (Fig. 4.6). Common subtypes were found amongst *Firmicutes* and Archaea in 2[4Fe-2S] subtype 3 and 15 and 2Fe-2S subtype 6, 8 and 9, indicating their common ancestral origin (Fig. 4.6).

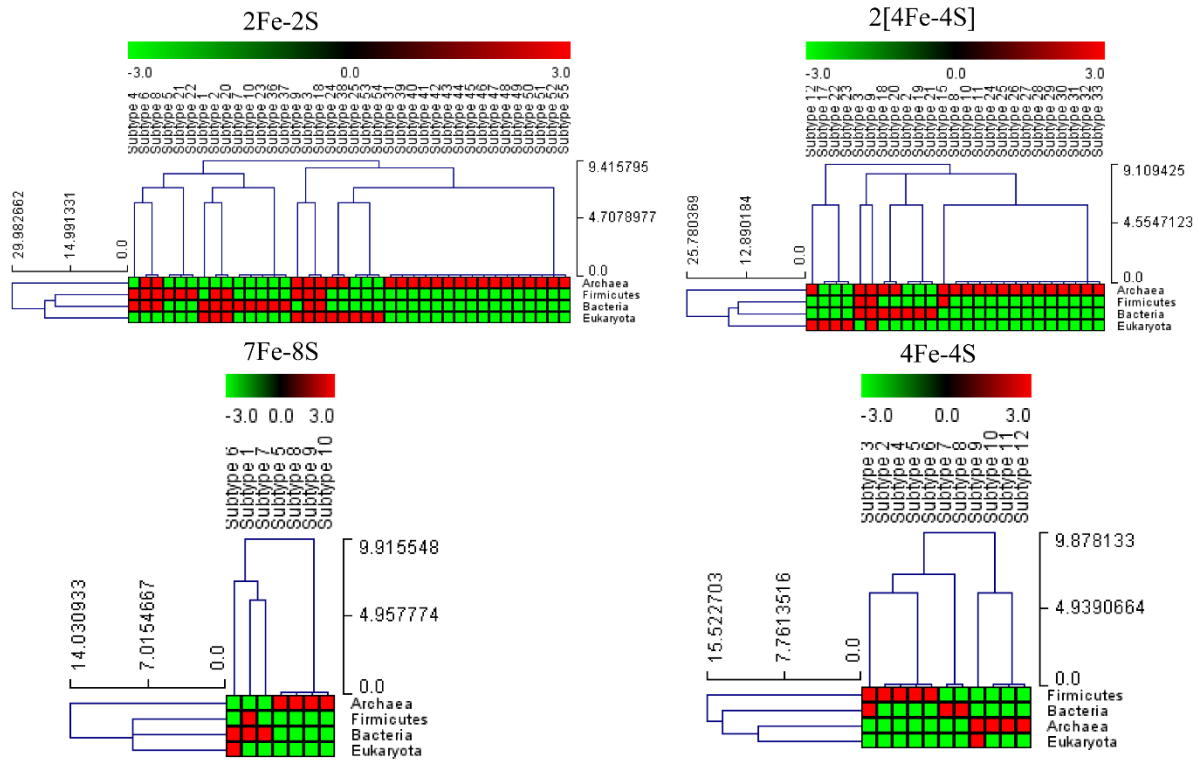


Figure 4.6. Heat map figure representing the presence (red) and absence (green) of ferredoxin subtypes in *Firmicutes*, Archaea, Bacteria and Eukarya. Ferredoxin subtypes form the horizontal axis and *Firmicutes* and other domains of life forms the vertical axis

CHAPTER 5: Conclusions and future perspectives

Cytochrome P450 monooxygenases (CYPs/P450s) have been well known to biologists for nearly six decades. These enzymes perform enzymatic reactions in stereo- and regio-specific manners and have thus gained attention in the production of chemicals valuable to human beings. Studies indicated that P450s performing such enzymatic reactions help organisms to adapt to specific ecological niches. This includes loss of P450s if an organism adapts to a lifestyle where abundant simple carbon sources are present. In this study, we examined such a phenomenon where *Firmicutes* species that are pathogens and adapted to simple lifestyles lost P450s in their genomes. This strongly supports the earlier hypothesis put forward by our laboratory that the impact of lifestyle shapes P450 content in an organism. Analysis of P450s in *Firmicutes* species revealed the presence of the lowest number of P450s and the lowest P450 diversity in these species compared to bacterial species belonging to the genera *Streptomyces* and Mycobacterium and from the phylum *Cyanobacteria*. The lowest P450 diversity is due to the P450 family blooming/expansion, especially in P450 families such as CYP107, CYP102, CYP152, CYP109 and CYP106. *Firmicutes* species were found to have the second highest number of BGCs in their genomes after *Streptomyces*. Most of the P450s belonging to the expanded families were found to be part of BGCs indicating that these P450s were favored by these species, possibly to gain some advantage, as observed in other organisms. Interestingly, the preference of certain P450 subfamilies for being part of BGCs was found in *Firmicutes* species despite these P450 subfamilies not being expanded. Based on the presence of ortholog P450s in *Firmicutes* species and the highest percentage of sequence identity with characterized BGCs, in this study we successfully predicted 89 P450s' function in the synthesis of different secondary metabolites. Interestingly, most of these P450s are involved in the synthesis of bacillaene and difficidin (polyketides) and fengycin (lipopeptide). Based on the P450 profiles among species belonging to the same phylum and their lifestyles, one can connect P450 profiles to their lifestyles and this concept has been successfully demonstrated in eukaryotes and in some prokaryotes. It is worth mentioning that the phylogenetic age of bacteria from old to young is *Firmicutes*, *Streptomyces*, *Cyanobacteria* and *Mycobacterium* (Battistuzzi et al., 2004), indicating that there is no link with the number and diversity of P450s and the age of the bacteria and thus P450 profiles may be more strongly related to the lifestyle. Studies are in progress to determine the impact of lifestyle on other bacterial species.

Iron-sulfur (Fe-S) cluster proteins such as ferredoxins are the simplest known electron transfer proteins in biology and they evolved in such a way that they tune both reduction potential and partner binding abilities (Campbell et al., 2019; Mutter et al., 2019; Hosseinzadeh and Lu, 2016; Kim et al., 2016). A total of 281 ferredoxins were found in 227 *Firmicutes* species. 4Fe-4S was the dominant ferredoxin found in this phylum. This is the earliest ferredoxin found in anaerobic species such as *Clostridia* species which was used in this study. In this study, for the first time, ferredoxins subtype classification has been proposed based on the amino acid spacing pattern between the cysteine amino acids of the Fe-S cluster binding motif as a subtype signature to understand the diversity and lateral (or horizontal) gene transfer of ferredoxins across the domains of life. Ferredoxins subtype classification helped in understanding their common ancestral origin between Archean and Bacteria and LGT from prokaryotes (Archaea/Bacteria) to eukaryotes. In this study, for easy identification of ferredoxins belonging to different subtypes, a nomenclature system was also developed. This study results serves as a reference on ferredoxin subtype classification and nomenclature. A complete analysis of ferredoxins of the entire *Firmicutes* phylum could be done to get a broader picture on the various types of ferredoxins. Further studies could be done in determining the criteria that P450s use to identify their preferred ferredoxin and P450s role in the generation of secondary metabolites.

REFERENCES

AKAPO, O. O., PADAYACHEE, T., CHEN, W., KAPPO, A. P., YU, J. H., NELSON, D. R. & SYED, K. 2019. Distribution and diversity of cytochrome P450 monooxygenases in the fungal class *Tremellomycetes*. *Int J Mol Sci,* 20.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ARUMUGAM, M., RAES, J., PELLETIER, E., LE PASLIER, D., YAMADA, T., MENDE, D. R., FERNANDES, G. R., TAP, J., BRULS, T., BATTO, J. M., BERTALAN, M., BORRUEL, N., CASELLAS, F., FERNANDEZ, L., GAUTIER, L., HANSEN, T., HATTORI, M., HAYASHI, T., KLEEREBEZEM, M., KUROKAWA, K., LECLERC, M., LEVENEZ, F., MANICHANH, C., NIELSEN, H. B., NIELSEN, T., PONS, N., POULAIN, J., QIN, J., SICHERITZ-PONTEN, T., TIMS, S., TORRENTS, D., UGARTE, E., ZOETENDAL, E. G., WANG, J., GUARNER, F., PEDERSEN, O., DE VOS, W. M., BRUNAK, S., DORÉ, J., ANTOLÍN, M., ARTIGUENAVE, F., BLOTTIERE, H. M., ALMEIDA, M., BRECHOT, C., CARA, C., CHERVAUX, C., CULTRONE, A., DELORME, C., DENARIAZ, G., DERVYN, R., FOERSTNER, K. U., FRISS, C., VAN DE GUCHTE, M., GUEDON, E., HAIMET, F., HUBER, W., VAN HYLCKAMA-VLIEG, J., JAMET, A., JUSTE, C., KACI, G., KNOL, J., LAKHDARI, O., LAYEC, S., LE ROUX, K., MAGUIN, E., MÉRIEUX, A., MELO MINARDI, R., M'RINI, C., MULLER, J., OOZEER, R., PARKHILL, J., RENAULT, P., RESCIGNO, M., SANCHEZ, N., SUNAGAWA, S., TORREJON, A., TURNER, K., VANDEMEULEBROUCK, G., VARELA, E., WINOGRADSKY, Y., ZELLER, G., WEISSENBACH, J., EHRLICH, S. D. & BORK, P. 2011. Enterotypes of the human gut microbiome. *Nature,* 473**,** 174-80.

BARSBY, T., KELLY, M. T. & ANDERSEN, R. J. 2002. Tupuseleiamides and basiliskamides, new acyldipeptides and antifungal polyketides produced in culture by a *Bacillus laterosporus* isolate obtained from a tropical marine habitat. *J Nat Prod,* 65**,** 1447-51.

BATTISTUZZI, F. U., FEIJAO, A. & HEDGES, S. B. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol,* 4**,** 44.

BEINERT, H., HOLM, R. H. & MÜNCK, E. 1997. Iron-sulfur clusters: nature's modular, multipurpose structures. *Science,* 277**,** 653-659.

BENTLEY, S. D., CHATER, K. F., CERDEÑO-TÁRRAGA, A.-M., CHALLIS, G. L., THOMSON, N., JAMES, K. D., HARRIS, D. E., QUAIL, M. A., KIESER, H. & HARPER, D. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature,* 417**,** 141.

BINTSIS, T. 2018. Lactic acid bacteria as starter cultures: An update in their metabolism and genetics. *AIMS Microbiol,* 4**,** 665-684.

BOC, A., DIALLO, A. B. & MAKARENKOV, V. 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Research,* 40**,** W573-W579.

BORISOVA, S. A., CIRCELLO, B. T., ZHANG, J. K., VAN DER DONK, W. A. & METCALF, W. W. 2010. Biosynthesis of rhizocticins, antifungal phosphonate oligopeptides produced by *Bacillus subtilis* ATCC6633. *Chem Biol,* 17**,** 28-37.

BOWER, S., PERKINS, J. B., YOCUM, R. R., HOWITT, C. L., RAHAIM, P. & PERO, J. 1996. Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon. *J Bacteriol,* 178**,** 4122-30.

BRUSCHI, M. & COUCHOUD, P. 1979. Amino acid sequence of *Desulfovibrio gigas* ferredoxin: revisions. *Biochem Biophys Res Commun,* 91**,** 623-8.

BRUSCHI, M. & GUERLESQUIN, F. 1988. Structure, function and evolution of bacterial ferredoxins. *FEMS Microbiol Rev,* 4**,** 155-75.

BRUSCHI, M. & HATCHIKIAN, E. C. 1982. Non-heme iron proteins of *Desulfovibrio*: the primary structure of ferredoxin I from *Desulfovibrio africanus*. *Biochimie,* 64**,** 503-7.

BUTCHER, R. A., SCHROEDER, F. C., FISCHBACH, M. A., STRAIGHT, P. D., KOLTER, R., WALSH, C. T. & CLARDY, J. 2007. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences,* 104**,** 1506-1509.

CAI, K., TONELLI, M., FREDERICK, R. O. & MARKLEY, J. L. 2017. Human mitochondrial ferredoxin 1 (FDX1) and ferredoxin 2 (FDX2) both bind cysteine desulfurase and donate electrons for iron–sulfur cluster biosynthesis. *Biochemistry,* 56**,** 487-499.

CAMMACK, R. 1983. Evolution and diversity of the iron sulfur proteins. . *Chem. Scripta,* 21**,** 87-95.

CAMPBELL, I. J., BENNETT, G. N. & SILBERG, J. J. 2019. Evolutionary relationships between low potential ferredoxin and flavodoxin electron carriers. *Frontiers in energy research,* 7**,** 10.3389/fenrg.2019.00079.

CANANI, R. B., COSTANZO, M. D., LEONE, L., PEDATA, M., MELI, R. & CALIGNANO, A. 2011. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J Gastroenterol,* 17**,** 1519-28.

CASTANER, O., GODAY, A., PARK, Y.-M., LEE, S.-H., MAGKOS, F., SHIOW, S.-A. T. E. & SCHRÖDER, H. 2018. The Gut Microbiome Profile in Obesity: A Systematic Review. *International journal of endocrinology,* 2018**,** 4095789-4095789.

CAULIER, S., NANNAN, C., GILLIS, A., LICCIARDI, F., BRAGARD, C. & MAHILLON, J. 2019. Overview of the antimicrobial compounds produced by members of the *Bacillus subtilis* group. *Frontiers in microbiology,* 10**,** 302-302.

CHAKRABORTI, C. K. 2015. New-found link between microbiota and obesity. *World journal of gastrointestinal pathophysiology,* 6**,** 110-119.

CHEN, X. H., VATER, J., PIEL, J., FRANKE, P., SCHOLZ, R., SCHNEIDER, K., KOUMOUTSI, A., HITZEROTH, G., GRAMMEL, N., STRITTMATTER, A. W., GOTTSCHALK, G., SÜSSMUTH, R. D. & BORRISS, R. 2006. Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J Bacteriol,* 188**,** 4024-36.

CHILD, S. A., BRADLEY, J. M., PUKALA, T. L., SVISTUNENKO, D. A., LE BRUN, N. E. & BELL, S. G. 2018. Electron transfer ferredoxins with unusual cluster binding motifs support secondary metabolism in many bacteria. *Chemical science,* 9**,** 7948-7957.

CHILIZA, Z. E., MARTÍNEZ-OYANEDEL, J. & SYED, K. 2020. An overview of the factors playing a role in cytochrome P450 monooxygenase and ferredoxin interactions. *Biophys Rev,* 12**,** 1217-1222.

CIMERMANCIC, P., MEDEMA, M. H., CLAESEN, J., KURITA, K., WIELAND BROWN, L. C., MAVROMMATIS, K., PATI, A., GODFREY, P. A., KOEHRSEN, M., CLARDY, J., BIRREN, B. W., TAKANO, E., SALI, A., LININGTON, R. G. & FISCHBACH, M. A. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell,* 158**,** 412-421.

CONOVER, R. C., KOWAL, A. T., FU, W. G., PARK, J. B., AONO, S., ADAMS, M. W. & JOHNSON, M. K. 1990. Spectroscopic characterization of the novel iron-sulfur cluster in *Pyrococcus furiosus* ferredoxin. *J Biol Chem,* 265**,** 8533-41.

CONSORTIUM, W. 2019. Protein data bank: the single global archive for 3D macromolecular structure data. *Nucleic acids research,* 47**,** D520-D528.

CREMADES, N., BUENO, M., TOJA, M. & SANCHO, J. 2005. Towards a new therapeutic target: *Helicobacter pylori* flavodoxin. *Biophysical Chemistry,* 115**,** 267-276.

CRYLE, M. J., BELL, S. G. & SCHLICHTING, I. 2010. Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: a cyclo-L-leucyl-L-leucyl dipeptide oxidase. *Biochemistry,* 49**,** 7282-96.

CRYLE, M. J. & SCHLICHTING, I. 2008. Structural insights from a P450 carrier protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. *Proc Natl Acad Sci U S A,* 105**,** 15696-701.

DAUTER, Z., WILSON, K. S., SIEKER, L. C., MEYER, J. & MOULIS, J.-M. 1997. Atomic resolution (0.94 Å) structure of *Clostridium acidurici* ferredoxin. Detailed geometry of [4Fe-4S] clusters in a protein. *Biochemistry,* 36**,** 16065-16073.

DIETERICH, W., SCHINK, M. & ZOPF, Y. 2018. Microbiota in the gastrointestinal tract. *Med Sci (Basel),* 6.

ECK, R. V. & DAYHOFF, M. O. 1966. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science,* 152**,** 363-366.

EL-GEBALI, S., MISTRY, J., BATEMAN, A., EDDY, S. R., LUCIANI, A., POTTER, S. C., QURESHI, M., RICHARDSON, L. J., SALAZAR, G. A. & SMART, A. 2019. The Pfam protein families database in 2019. *Nucleic acids research,* 47**,** D427-D432.

ELSEN, S., EFTHYMIOU, G., PETEINATOS, P., DIALLINAS, G., KYRITSIS, P. & MOULIS, J. M. 2010. A bacteria-specific 2[4Fe-4S] ferredoxin is essential in *Pseudomonas aeruginosa*. *BMC Microbiol,* 10**,** 271.

EWEN, K. M., HANNEMANN, F., KHATRI, Y., PERLOVA, O., KAPPL, R., KRUG, D., HÜTTERMANN, J., MÜLLER, R. & BERNHARDT, R. 2009. Genome mining in *Sorangium cellulosum* So ce56: identification and characterization of the homologous electron transfer proteins of a myxobacterial cytochrome P450. *The Journal of biological chemistry,* 284**,** 28590-28598.

FEYEREISEN, R. 2012. *Insect Molecular Biology and Biochemistry* Elsevier, Amsterdam.

FILIPPIDOU, S., WUNDERLIN, T., JUNIER, T., JEANNERET, N., DORADOR, C., MOLINA, V., JOHNSON, D. R. & JUNIER, P. 2016. A combination of extreme environmental conditions favor the prevalence of endospore-forming *Firmicutes*. *Front Microbiol,* 7**,** 1707.

FRAZÃO, C., ARAGÃO, D., COELHO, R., LEAL, S. S., GOMES, C. M., TEIXEIRA, M. & CARRONDO, M. A. 2008. Crystallographic analysis of the intact metal centres

[3Fe–4S] 1+/0 and [4Fe–4S] 2+/1+ in a Zn2+-containing ferredoxin. *FEBS letters,* 582**,** 763-767.

GIRVAN, H. M., PODDAR, H., MCLEAN, K. J., NELSON, D. R., HOLLYWOOD, K. A., LEVY, C. W., LEYS, D. & MUNRO, A. W. 2018. Structural and catalytic properties of the peroxygenase P450 enzyme CYP152K6 from *Bacillus methanolicus*. *J Inorg Biochem,* 188**,** 18-28.

GOÑI, G., SERRANO, A., FRAGO, S., HERVÁS, M., PEREGRINA, J. R., DE LA ROSA, M. A., GÓMEZ-MORENO, C., NAVARRO, J. A. & MEDINA, M. 2008. Flavodoxin-mediated electron transfer from photosystem I to ferredoxin-NADP+ reductase in anabaena: role of flavodoxin hydrophobic residues in protein-protein interactions. *Biochemistry,* 47**,** 1207-17.

GOSALBES, M. J., DURBAN, A., PIGNATELLI, M., ABELLAN, J. J., JIMENEZ-HERNANDEZ, N., PEREZ-COBAS, A. E., LATORRE, A. & MOYA, A. 2011. Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One,* 6**,** e17447.

GRAHAM, S. E. & PETERSON, J. A. 1999. How similar are P450s and what can their differences teach us? *Arch Biochem Biophys,* 369**,** 24-9.

GRANGEMARD, I., WALLACH, J., MAGET-DANA, R. & PEYPOUX, F. 2001. Lichenysin: a more efficient cation chelator than surfactin. *Appl Biochem Biotechnol,* 90**,** 199-210.

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.

GU, Q., YANG, Y., YUAN, Q., SHI, G., WU, L., LOU, Z., HUO, R., WU, H., BORRISS, R. & GAO, X. 2017. Bacillomycin D produced by *Bacillus amyloliquefaciens* is involved in the antagonistic interaction with the plant-pathogenic fungus *Fusarium graminearum*. *Appl Environ Microbiol,* 83.

GUERRINI, O., BURLAT, B., LÉGER, C., GUIGLIARELLI, B., SOUCAILLE, P. & GIRBAL, L. 2008. Characterization of two 2[4Fe4S] ferredoxins from *Clostridium acetobutylicum*. *Curr Microbiol,* 56**,** 261-7.

HAMBERGER, B. & BAK, S. 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20120426.

HANNEMANN, F., BICHET, A., EWEN, K. M. & BERNHARDT, R. 2007. Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta,* 1770**,** 330-44.

HERTLEIN, G., MÜLLER, S., GARCIA-GONZALEZ, E., POPPINGA, L., SÜSSMUTH, R. D. & GENERSCH, E. 2014. Production of the catechol type siderophore bacillibactin by the honey bee pathogen *Paenibacillus larvae*. *PLoS One,* 9**,** e108272.

HLAVICA, P. 2015. Mechanistic basis of electron transfer to cytochromes P450 by natural redox partners and artificial donor constructs. *Adv Exp Med Biol,* 851**,** 247-97.

HOSSEINZADEH, P. & LU, Y. 2016. Design and fine-tuning redox potentials of metalloproteins involved in electron transfer in bioenergetics. *Biochimica et Biophysica Acta (BBA)-Bioenergetics,* 1857**,** 557-581.

HOWE, E. A., SINHA, R., SCHLAUCH, D. & QUACKENBUSH, J. 2011. RNA-Seq analysis in MeV. *Bioinformatics,* 27**,** 3209-10.

HSUEH, K.-L., YU, L.-K., CHEN, Y.-H., CHENG, Y.-H., HSIEH, Y.-C., KE, S.-C., HUNG, K.-W., CHEN, C.-J. & HUANG, T.-H. 2013. FeoC from *Klebsiella pneumoniae* contains a [4Fe-4S] cluster. *Journal of bacteriology,* 195**,** 4726-4734.

ILBERT, M. & BONNEFOY, V. 2013. Insight into the evolution of the iron oxidation pathways. *Biochim Biophys Acta,* 1827**,** 161-75.

IMLAY, J. A. 2006. Iron-sulphur clusters and the problem with oxygen. *Mol Microbiol,* 59**,** 1073-82.

INDA, L. A. & LUISA PELEATO, M. 2003. Development of an ELISA approach for the determination of flavodoxin and ferredoxin as markers of iron deficiency in phytoplankton. *Phytochemistry,* 63**,** 303-8.

JAGANNATHAN, B. & GOLBECK, J. H. 2009. Understanding of the binding interface between PsaC and the PsaA/PsaB heterodimer in photosystem I. *Biochemistry,* 48**,** 5405-16.

JAWALLAPERSAND, P., MASHELE, S. S., KOVAČIČ, L., STOJAN, J., KOMEL, R., PAKALA, S. B., KRAŠEVEC, 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.

KANEHISA, M., SATO, Y., FURUMICHI, M., MORISHIMA, K. & TANABE, M. 2019. New approach for understanding genome variations in KEGG. *Nucleic Acids Res,* 47**,** D590-d595.

KANEHISA, M., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research,* 44**,** D457-D462.

KANG, H. K., SEO, C. H. & PARK, Y. 2015. Marine peptides and their anti-infective activities. *Mar Drugs,* 13**,** 618-54.

KATOH, K., KUMA, K., TOH, H. & MIYATA, T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res,* 33**,** 511-8.

KEHRER, J. P. 2000. The Haber-Weiss reaction and mechanisms of toxicity. *Toxicology,* 149**,** 43-50.

KGOSIEMANG, I. K. R., SYED, K. & MASHELE, S. Comparative genomics and evolutionary analysis of cytochrome P450 monooxygenases in fungal subphylum *Saccharomycotina*. 2014.

KHO, Z. Y. & LAL, S. K. 2018. The Human gut microbiome - A potential controller of wellness and disease. *Front Microbiol,* 9**,** 1835.

KHUMALO, M. J., NZUZA, N., PADAYACHEE, T., CHEN, W., YU, J.-H., NELSON, D. R. & SYED, K. 2020. Comprehensive analyses of cytochrome P450 monooxygenases and secondary metabolite biosynthetic gene clusters in *Cyanobacteria*. *International journal of molecular sciences,* 21**,** 656.

KIM, J. Y., NAKAYAMA, M., TOYOTA, H., KURISU, G. & HASE, T. 2016. Structural and mutational studies of an electron transfer complex of maize sulfite reductase and ferredoxin. *The Journal of Biochemistry,* 160**,** 101-109.

KNIGHT, C. A., BOWMAN, M. J., FREDERICK, L., DAY, A., LEE, C. & DUNLAP, C. A. 2018. The first report of antifungal lipopeptide production by a *Bacillus subtilis* subsp. *inaquosorum* strain. *Microbiol Res,* 216**,** 40-46.

KNIGHT, E., JR., D'EUSTACHIO, A. J. & HARDY, R. W. 1966. Flavodoxin: a flavoprotein with ferredoxin activity from *Clostrium pasteurianum*. *Biochim Biophys Acta,* 113**,** 626-8.

KOKSHAROVA, O. A., KLINT, J. & RASMUSSEN, U. 2006. The first protein map of *Synechococcus* sp. strain PCC 7942. *Microbiology,* 75**,** 664-672.

KOLIADA, A., SYZENKO, G., MOSEIKO, V., BUDOVSKA, L., PUCHKOV, K., PEREDERIY, V., GAVALKO, Y., DOROFEYEV, A., ROMANENKO, M., TKACH, S., SINEOK, L., LUSHCHAK, O. & VAISERMAN, A. 2017. Association between body mass index and *Firmicutes*/*Bacteroidetes* ratio in an adult Ukrainian population. *BMC Microbiol,* 17**,** 120.

LAMB, D. C. & WATERMAN, M. R. 2013. Unusual properties of the cytochrome P450 superfamily. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 368**,** 20120434-20120434.

LAU, I. C., FEYEREISEN, R., NELSON, D. R. & BELL, S. G. 2019. Analysis and preliminary characterisation of the cytochrome P450 monooxygenases from *Frankia* sp. EuI1c (*Frankia inefficax* sp.). *Archives of biochemistry and biophysics,* 669**,** 11-21.

LEE, D. S., YAMADA, A., MATSUNAGA, I., ICHIHARA, K., ADACHI, S., PARK, S. Y. & SHIRO, Y. 2002. Crystallization and preliminary X-ray diffraction analysis of fatty-acid hydroxylase cytochrome P450BSbeta from *Bacillus subtilis*. *Acta Crystallogr D Biol Crystallogr,* 58**,** 687-9.

LEE, D. S., YAMADA, A., SUGIMOTO, H., MATSUNAGA, I., OGURA, H., ICHIHARA, K., ADACHI, S., PARK, S. Y. & SHIRO, Y. 2003. Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from *Bacillus subtilis*. Crystallographic, spectroscopic, and mutational studies. *J Biol Chem,* 278**,** 9761-7.

LETUNIC, I. & BORK, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res,* 44**,** W242-5.

LETUNIC, I. & BORK, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res,* 47**,** W256-w259.

LEWIS, D. F. V. & HLAVICA, P. 2000. Interactions between redox partners in various cytochrome P450 systems: functional and structural aspects. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1460**,** 353-374.

LEY, R. E., PETERSON, D. A. & GORDON, J. I. 2006a. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell,* 124**,** 837-48.

LEY, R. E., TURNBAUGH, P. J., KLEIN, S. & GORDON, J. I. 2006b. Microbial ecology: human gut microbes associated with obesity. *Nature,* 444**,** 1022-3.

LI, H. & POULOS, T. L. 1997. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat Struct Biol,* 4**,** 140-6.

LOPETUSO, L. R., SCALDAFERRI, F., PETITO, V. & GASBARRINI, A. 2013. Commensal *Clostridia*: leading players in the maintenance of gut homeostasis. *Gut Pathog,* 5**,** 23.

LU, S., WANG, J., CHITSAZ, F., DERBYSHIRE, M. K., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., MARCHLER, G. H. & SONG, J. S. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic acids research,* 48**,** D265-D268.

LYONS, T. W., REINHARD, C. T. & PLANAVSKY, N. J. 2014. The rise of oxygen in Earth's early ocean and atmosphere. *Nature,* 506**,** 307-315.

MACEDO-RIBEIRO, S., MARTINS, B. M., PEREIRA, P. B., BUSE, G., HUBER, R. & SOULIMANE, T. 2001. New insights into the thermostability of bacterial ferredoxins: high-resolution crystal structure of the seven-iron ferredoxin from *Thermus thermophilus*. *JBIC Journal of Biological Inorganic Chemistry,* 6**,** 663-674.

MADEIRA, F., PARK, Y. M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., TIVEY, A. R., POTTER, S. C. & FINN, R. D. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research,* 47**,** W636-W641.

MAJOR, T. A., BURD, H. & WHITMAN, W. B. 2004. Abundance of 4Fe–4S motifs in the genomes of methanogens and other prokaryotes. *FEMS Microbiology Letters,* 239**,** 117-123.

MANICHANH, C., RIGOTTIER-GOIS, L., BONNAUD, E., GLOUX, K., PELLETIER, E., FRANGEUL, L., NALIN, R., JARRIN, C., CHARDON, P., MARTEAU, P., ROCA, J. & DORE, J. 2006. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut,* 55**,** 205-11.

MARCHLER-BAUER, A., BO, Y., HAN, L., HE, J., LANCZYCKI, C. J., LU, S., CHITSAZ, F., DERBYSHIRE, M. K., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., LU, F., MARCHLER, G. H., SONG, J. S., THANKI, N., WANG, Z., YAMASHITA, R. A., ZHANG, D., ZHENG, C., GEER, L. Y. & BRYANT, S. H. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res,* 45**,** D200-d203.

MARCIANO, F. & VAJRO, P. 2017. Oxidative Stress and Gut Microbiota * *Conflict of interest: None. *Gastrointestinal Tissue.*

MARIAT, D., FIRMESSE, O., LEVENEZ, F., GUIMARAES, V., SOKOL, H., DORE, J., CORTHIER, G. & FURET, J. P. 2009. The *Firmicutes*/*Bacteroidetes* ratio of the human microbiota changes with age. *BMC Microbiol,* 9**,** 123.

MATOWANE, R. G., WIETESKA, L., BAMAL, H. D., KGOSIEMANG, I. K. R., VAN WYK, M., MANUME, N. A., ABDALLA, S. M. H., MASHELE, S. S., GRONT, D.

& SYED, K. 2018. *In silico* analysis of cytochrome P450 monooxygenases in chronic granulomatous infectious fungus *Sporothrix schenckii*: Special focus on CYP51. *Biochim Biophys Acta Proteins Proteom,* 1866**,** 166-177.

MCLEAN, K. J., WARMAN, A. J., SEWARD, H. E., MARSHALL, K. R., GIRVAN, H. M., CHEESMAN, M. R., WATERMAN, M. R. & MUNRO, A. W. 2006. Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions. *Biochemistry,* 45**,** 8427-43.

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., DUDDELA, S., DÜSTERHUS, 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., JUNGMANN, K., KEGLER, C., KIM, H. U., KÖTTER, P., KRUG, D., MASSCHELEIN, J., MELNIK, A. V., MANTOVANI, S. M., MONROE, E. A., MOORE, M., MOSS, N., NÜTZMANN, 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., BALSKUS, E. P., BARONA-GÓMEZ, F., BECHTHOLD, A., BODE, H. B., BORRISS, 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., HÖFTE, 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.

MEYER, J. 2008. Iron–sulfur protein folds, iron–sulfur chemistry, and evolution. *JBIC Journal of Biological Inorganic Chemistry,* 13**,** 157-170.

MITCHELL, A. L., ATTWOOD, T. K., BABBITT, P. C., BLUM, M., BORK, P., BRIDGE, A., BROWN, S. D., CHANG, H.-Y., EL-GEBALI, S. & FRASER, M. I. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research,* 47**,** D351-D360.

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.

MOLDENHAUER, J., CHEN, X. H., BORRISS, R. & PIEL, J. 2007. Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. *Angew Chem Int Ed Engl,* 46**,** 8195-7.

MORTENSON, L. E., VALENTINE, R. C. & CARNAHAN, J. E. 1962. An electron transport factor from *Clostridium pasteurianum*. *Biochem Biophys Res Commun,* 7**,** 448-52.

MOULIS, J. M., SIEKER, L. C., WILSON, K. S. & DAUTER, Z. 1996. Crystal structure of the 2 [4Fe-4S] ferredoxin from *Chromatium vinosum*: Evolutionary and mechanistic inferences for [3/4Fe-4S] ferredoxins. *Protein science,* 5**,** 1765-1775.

MTHETHWA, B. C., CHEN, W., NGWENYA, M. L., KAPPO, A. P., SYED, P. R., KARPOORMATH, R., YU, J. H., NELSON, D. R. & SYED, K. 2018. Comparative analyses of cytochrome P450s and those associated with secondary metabolism in *Bacillus* species. *Int J Mol Sci,* 19.

MUNRO, A. W., GIRVAN, H. M. & MCLEAN, K. J. 2007. Cytochrome P450--redox partner fusion enzymes. *Biochim Biophys Acta,* 1770**,** 345-59.

MUNRO, A. W., LEYS, D. G., MCLEAN, K. J., MARSHALL, K. R., OST, T. W., DAFF, S., MILES, C. S., CHAPMAN, S. K., LYSEK, D. A., MOSER, C. C., PAGE, C. C. & DUTTON, P. L. 2002. P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem Sci,* 27**,** 250-7.

MUTTER, A. C., TYRYSHKIN, A. M., CAMPBELL, I. J., POUDEL, S., BENNETT, G. N., SILBERG, J. J., NANDA, V. & FALKOWSKI, P. G. 2019. De novo design of symmetric ferredoxins that shuttle electrons in vivo. *Proceedings of the National Academy of Sciences,* 116**,** 14557-14562.

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. 2018. Cytochrome P450 diversity in the tree of life. *Biochim Biophys Acta Proteins Proteom,* 1866**,** 141-154.

NELSON, D. R., KAMATAKI, T., WAXMAN, D. J., GUENGERICH, F. P., ESTABROOK, R. W., FEYEREISEN, R., GONZALEZ, F. J., COON, M. J., GUNSALUS, I. C., GOTOH, O. & ET AL. 1993. The P450 superfamily: update on new sequences, gene

mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol,* 12**,** 1-51.

NELSON, D. R. & STROBEL, H. W. 1987. Evolution of cytochrome P-450 proteins. *Mol Biol Evol,* 4**,** 572-93.

NGWENYA, M. L., CHEN, W., BASSON, A. K., SHANDU, J. S., YU, J. H., NELSON, D. R. & SYED, K. 2018. Blooming of unusual cytochrome P450s by tandem duplication in the pathogenic fungus *Conidiobolus coronatus*. *Int J Mol Sci,* 19.

NOBLE, M. A., MILES, C. S., CHAPMAN, S. K., LYSEK, D. A., MACKAY, A. C., REID, G. A., HANZLIK, R. P. & MUNRO, A. W. 1999. Roles of key active-site residues in flavocytochrome P450 BM3. *Biochem J,* 339 (Pt 2)**,** 371-9.

NOTONIER, S., ALEXANDER, M. & JAYAKODY, L. N. 2016. An overview of P450 enzymes: opportunity and challenges in industrial applications. *Enzyme Engineering,* 05.

ORTEGA UGALDE, S., DE KONING, C. P., WALLRAVEN, K., BRUYNEEL, B., VERMEULEN, N. P. E., GROSSMANN, T. N., BITTER, W., COMMANDEUR, J. N. M. & VOS, J. C. 2018. Linking cytochrome P450 enzymes from *Mycobacterium tuberculosis* to their cognate ferredoxin partners. *Appl Microbiol Biotechnol,* 102**,** 9231-9242.

OTTMAN, N., SMIDT, H., DE VOS, W. M. & BELZER, C. 2012. The function of our microbiota: who is out there and what do they do? *Front Cell Infect Microbiol,* 2**,** 104.

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**,** 13982.

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.

PAUL R. ORTIZ DE MONTELLANO 2015. Cytochrome P450. 4 ed.: Springer International Publishing.

PEYPOUX, F., BONMATIN, J. M. & WALLACH, J. 1999. Recent trends in the biochemistry of surfactin. *Appl Microbiol Biotechnol,* 51**,** 553-63.

PIERELLA KARLUSICH, J. J., LODEYRO, A. F. & CARRILLO, N. 2014. The long goodbye: the rise and fall of flavodoxin during plant evolution. *J Exp Bot,* 65**,** 5161-78.

POCHAPSKY, T. C., JAIN, N. U., KUTI, M., LYONS, T. A. & HEYMONT, J. 1999. A refined model for the solution structure of oxidized putidaredoxin. *Biochemistry,* 38**,** 4681-4690.

PODUST, L. M. & SHERMAN, D. H. 2012. Diversity of P450 enzymes in the biosynthesis of natural products. *Nat Prod Rep,* 29**,** 1251-66.

PROCÓPIO, R. E., SILVA, I. R., MARTINS, M. K., AZEVEDO, J. L. & ARAÚJO, J. M. 2012. Antibiotics produced by *Streptomyces*. *Braz J Infect Dis,* 16**,** 466-71.

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.

RAMAKRISHNA, B. S. 2013. Role of the gut microbiota in human nutrition and metabolism. *J Gastroenterol Hepatol,* 28 Suppl 4**,** 9-17.

ROMERO-TABAREZ, M., JANSEN, R., SYLLA, M., LÜNSDORF, H., HÄUßLER, S., SANTOSA, D. A., TIMMIS, K. N. & MOLINARI, G. 2006. Malonyl macrolactin A, a new macrolactin antibiotic from *Bacillus subtilis* active against methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant enterococci, and a small-colony variant of *Burkholderia cepacia*. *Antimicrobial Agents and Chemotherapy,* 50**,** 1701-1709.

ROOD, J. 2016. General physiological and virulence properties of the pathogenic *Clostridia*.

SAEED, A. I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN, M., STURN, A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V. & QUACKENBUSH, J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques,* 34**,** 374-8.

SAEKI, K., SUETSUGU, Y., TOKUDA, K.-I., MIYATAKE, Y., YOUNG, D., MARRS, B. & MATSUBARA, H. 1991. Genetic analysis of functional differences among distinct ferredoxins in *Rhodobacter capsulatus*. *Journal of Biological Chemistry,* 266**,** 12889-12895.

SANCHO, J. 2006. Flavodoxins: Sequence, folding, binding, function and beyond. *Cellular and molecular life sciences : CMLS,* 63**,** 855-64.

SARIDAKIS, E., GIASTAS, P., EFTHYMIOU, G., THOMA, V., MOULIS, J.-M., KYRITSIS, P. & MAVRIDIS, I. M. 2009. Insight into the protein and solvent contributions to the reduction potentials of [4Fe–4S] 2+/+ clusters: crystal structures of the *Allochromatium vinosum* ferredoxin variants C57A and V13G and the homologous *Escherichia coli* ferredoxin. *JBIC Journal of Biological Inorganic Chemistry,* 14**,** 783-799.

SAYERS, E. W., BECK, J., BOLTON, E. E., BOUREXIS, D., BRISTER, J. R., CANESE, K., COMEAU, D. C., FUNK, K., KIM, S. & KLIMKE, W. 2021. Database resources of the national center for biotechnology information. *Nucleic acids research,* 49**,** D10.

SCHLATTER, D., WALDVOGEL, S., ZÜLLI, F., SUTER, F., PORTMANN, W. & ZUBER, H. 1985. Purification, amino-acid sequence and some properties of the ferredoxin isolated from *Bacillus acidocaldarius*. *Biol Chem Hoppe Seyler,* 366**,** 223-31.

SCHÖNHEIT, P., BRANDIS, A. & THAUER, R. K. 1979. Ferredoxin degradation in growing *Clostridium pasteurianum* during periods of iron deprivation. *Arch Microbiol,* 120**,** 73-6.

SELLO, M. M., JAFTA, N., NELSON, D. R., CHEN, W., YU, J.-H., PARVEZ, M., KGOSIEMANG, I. K. R., MONYAKI, R., RASELEMANE, S. C., QHANYA, L. B., MTHAKATHI, N. T., SITHENI MASHELE, S. & SYED, K. 2015. Diversity and evolution of cytochrome P450 monooxygenases in *Oomycetes*. *Scientific reports,* 5**,** 11572-11572.

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.

SEONG, C. N., KANG, J. W., LEE, J. H., SEO, S. Y., WOO, J. J., PARK, C., BAE, K. S. & KIM, M. S. 2018. Taxonomic hierarchy of the phylum *Firmicutes* and novel *Firmicutes* species originated from various environments in Korea. *J Microbiol,* 56**,** 1-10.

SEVRIOUKOVA, I. F., LI, H., ZHANG, H., PETERSON, J. A. & POULOS, T. L. 1999. Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci U S A,* 96**,** 1863-8.

SHETHNA, Y. I., STOMBAUGH, N. A. & BURRIS, R. H. 1971. Ferredoxin from *Bacillus polymyxa*. *Biochem Biophys Res Commun,* 42**,** 1108-16.

SMILLIE, R. M. 1965. Isolation of two proteins with chloroplast ferredoxin activity from a blue-green alga. *Biochem Biophys Res Commun,* 20**,** 621-9.

SOUSA, F. L., THIERGART, T., LANDAN, G., NELSON-SATHI, S., PEREIRA, I. A. C., ALLEN, J. F., LANE, N. & MARTIN, W. F. 2013. Early bioenergetic evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20130088.

SUR, S., ROMO, T. D. & GROSSFIELD, A. 2018. Selectivity and mechanism of fengycin, an antimicrobial lipopeptide, from molecular dynamics. *J Phys Chem B,* 122**,** 2219-2226.

SUZUKI, H., MACDONALD, J., SYED, K., SALAMOV, A., HORI, C., AERTS, A., HENRISSAT, B., WIEBENGA, A., VANKUYK, P. A., BARRY, K., LINDQUIST, E., LABUTTI, K., LAPIDUS, A., LUCAS, S., COUTINHO, P., GONG, Y., SAMEJIMA, M., MAHADEVAN, R., ABOU-ZAID, M., DE VRIES, R. P., IGARASHI, K., YADAV, J. S., GRIGORIEV, I. V. & MASTER, E. R. 2012. Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and P. *chrysosporium,* to elucidate the genetic basis of the distinct wood types they colonize. *BMC Genomics,* 13**,** 444.

SYED, K. & MASHELE, S. S. 2014. Comparative analysis of P450 signature motifs EXXR and CXG in the large and diverse kingdom of fungi: identification of evolutionarily conserved amino acid patterns characteristic of P450 family. *PLoS One,* 9**,** e95616.

SYED K, S. K., PAGADALA NS, TUSZYNSKI J 2014. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS ONE,* 9.

SYED, P. R., CHEN, W., NELSON, D. R., KAPPO, A. P., YU, J.-H., KARPOORMATH, R. & SYED, K. 2019. Cytochrome P450 monooxygenase CYP139 family involved in the synthesis of secondary metabolites in 824 Mycobacterial species. *International journal of molecular sciences,* 20**,** 2690.

THURSBY, E. & JUGE, N. 2017. Introduction to the human gut microbiota. *Biochem J,* 474**,** 1823-1836.

TRAN, P. N., YEN, M. R., CHIANG, C. Y., LIN, H. C. & CHEN, P. Y. 2019. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl Microbiol Biotechnol,* 103**,** 3277-3287.

UNCIULEAC, M., BOLL, M., WARKENTIN, E. & ERMLER, U. 2004. Crystallization of 4-hydroxybenzoyl-CoA reductase and the structure of its electron donor ferredoxin. *Acta Crystallographica Section D: Biological Crystallography,* 60**,** 388-391.

URLACHER, V. B. & GIRHARD, M. 2012. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol,* 30**,** 26-36.

VENKATESWARA RAO, P. & HOLM, R. 2004. Synthetic analogues of the active sites of iron− sulfur proteins. *Chemical reviews,* 104**,** 527-560.

VERMA, D. K., AMI R. PATEL, P. & PREM PRAKASH SRIVASTAV, P. 2018. *Bioprocessing Technology in Food and Health: Potential Applications and Emerging Scope,* Apple Academic Press, Burlington.

WEBER, T., BLIN, K., DUDDELA, S., KRUG, D., KIM, H. U., BRUCCOLERI, R., LEE, S. Y., FISCHBACH, M. A., MÜLLER, R., WOHLLEBEN, W., BREITLING, R., TAKANO, E. & MEDEMA, M. H. 2015. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res,* 43**,** W237-43.

WELLS, C. L. & WILKINS, T. D. 1996. *Clostridia*: sporeforming snaerobic *Bacilli*. *In:* BARON, S. (ed.) *Medical Microbiology.* Galveston (TX): University of Texas Medical Branch at Galveston

WHITE, R. E. & COON, M. J. 1980. Oxygen activation by cytochrome P-450. *Annu Rev Biochem,* 49**,** 315-56.

WILSON, M. R., ZHA, L. & BALSKUS, E. P. 2017. Natural product discovery from the human microbiome. *J Biol Chem,* 292**,** 8546-8552.

WU, L., WU, H., CHEN, L., YU, X., BORRISS, R. & GAO, X. 2015. Difficidin and bacilysin from *Bacillus amyloliquefaciens* FZB42 have antibacterial activity against *Xanthomonas oryzae* rice pathogens. *Sci Rep,* 5**,** 12975.

YANG, J., ZHU, X., CAO, M., WANG, C., ZHANG, C., LU, Z. & LU, F. 2016. Genomics-inspired discovery of three antibacterial active metabolites, aurantinins B, C, and D from compost-associated *Bacillus subtilis* fmb60. *J Agric Food Chem,* 64**,** 8811-8820.

YUAN, J., ZHAO, M., LI, R., HUANG, Q., RENSING, C., RAZA, W. & SHEN, Q. 2016. Antibacterial compounds-macrolactin alters the soil bacterial community and abundance of the gene encoding PKS. *Frontiers in microbiology,* 7**,** 1904-1904.

ZANETTI, G. & PANDINI, V. 2013. Ferredoxin. *In:* LENNARZ, W. J. & LANE, M. D. (eds.) *Encyclopedia of Biological Chemistry (Second Edition).* Waltham: Academic Press.

ZHOU, M., LIU, F., YANG, X., JIN, J., DONG, X., ZENG, K. W., LIU, D., ZHANG, Y., MA, M. & YANG, D. 2018. Bacillibactin and bacillomycin analogues with cytotoxicities against human cancer cell lines from marine *Bacillus* sp. PKU-MA00093 and PKU-MA00092. *Mar Drugs,* 16.

ANNEXURE

Supplementary dataset 1: List of P450s identified in different *Firmicutes* species. Each P450 is presented with its given name followed by protein ID in parenthesis and species name.

Data can be found at: https://www.nature.com/articles/s41598-020-70686-8

Supplementary dataset 2: Ferredoxin protein sequences identified in 227 *Firmicutes* species. Ferredoxins were sorted as per their Fe-S cluster type and subtypes. Name of the ferredoxins were followed as mentioned in methodology. Name of the ferredoxins includes their nomenclature name followed by protein ID from KEGG database and species name.

Data can be found at: https://www.mdpi.com/1467-3045/43/3/98/htm

Supplementary dataset 3: Ferredoxin protein sequences identified by datamining of published articles and Protein Data Bank (PDB). Ferredoxins were sorted as per their Fe-S cluster type and subtypes. Name of the ferredoxins includes their nomenclature name as described in the methodology followed by protein ID (either GenBank accession number or PDB ID) database and species name.

Data can be found at: https://www.mdpi.com/1467-3045/43/3/98/htm