UNIVERSITY OF ZULULAND

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

Bridget Valeria Zinhle Nkosi

201759818

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

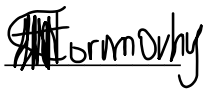
Supervisor: Prof Khajamohiddin Syed

Co-supervisor: Ms. Tiara Padayachee

KwaDlangezwa, September 2022

DECLARATION

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



Signed on the day of 26 September 2022

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

APPROVAL

I hereby approve the final submission of the following dissertation.

Prof K Syed

Supervisor

Ms. T Padayachee

Co-supervisor

ABSTRACT

Species belonging to the bacterial phyla *Bacteroidetes* and *Firmicutes* represent over 90% of the gastrointestinal microbiota. Changes in the ratio of these two bacterial groups were found to have contrasting health effects, including obesity and inflammatory diseases. Despite the availability of many bacterial genomes, comparative genomic studies on the gene pools of these two bacterial groups concerning cytochrome P450 monooxygenases (P450s), ferredoxins, and secondary metabolite biosynthetic gene clusters (smBGCs) are not reported. Recently, an analysis of P450s, ferredoxins, and smBGCs in *Firmicutes* species has been reported. However, such studies on *Bacteroidetes* species have not been performed. This study is aimed to address this research gap. In this study, a thorough comparative analysis of P450s in the phylum *Bacteroidetes* has been carried out. P450 data mining and annotation of P450s in this phylum displayed 98 P450s in 77 species. It consisted of 130 genera, the *Hymenobacter* genus having the most P450s. Twenty-one P450 families were discovered, with CYP1103 dominating. Cluster analysis revealed 1298 smBGCs, with terpene being the most dominant. Out of the 98 P450s found in 334 *Bacteroidetes* species, only eight P450s (8.2 %) of seven *Bacteroidetes* species were found as part of the secondary metabolite BGCs. Genome data mining and annotation of ferredoxins in 104 *Bacteroidetes* species revealed the presence of 269 ferredoxins in their genomes. Among the *Bacteroidetes* species, *Tenacibaculum jejuense* had the highest number of ferredoxins (six). The 269 ferredoxins found in *Bacteroidetes* species can be grouped into five iron-sulfur (Fe-S) cluster types: 2Fe-2S, 3Fe-4S, 4Fe-4S, 2[4Fe-4S], and 2[4Fe-4S]Alv. The 7Fe-8S cluster-type ferredoxins were not found in the *Bacteroidetes* species analyzed in this study. Based on the amino acid spacing pattern analysis between the cysteine amino acids of the Fe-S cluster binding motif, 136 2Fe-2S ferredoxins of *Bacteroidetes* can be grouped into five subtypes. Eleven 4Fe-4S ferredoxins found in *Bacteroidetes* species can be grouped into three subtypes. The study revealed the presence of diverse sets of P450s, ferredoxins, and smBGCs in *Bacteroidetes* species genomes. *Bacteroidetes* species have the highest number of P450 families, ferredoxin cluster-types, and smBGCs compared to *Firmicutes* species. Only four P450 families, three ferredoxin cluster types, and five smBGCs are commonly shared between these two bacterial groups. Considering the above facts, we propose that the contrasting effects of these two bacterial groups on the host are partly due to the distinct nature of secondary metabolites produced by these organisms. Thus, the cause of the contrasting health effects of these two bacterial groups lies in their gene pools.

DEDICATION

This work is dedicated to my family, my grandmother Reborn Nkosi, my mother Nompumelelo Nkosi, my siblings Zethu and Nduduzo, and my son Luhle Khumalo.

ACKNOWLEDGEMENTS

I want to thank the following people and funding for supporting, encouraging, and helping me achieve my goals.

1. Prof Khajamohiddin Syed. Thank you for seeing potential in me and allowing me to achieve my goals as a postgraduate student.
2. Miss Tiara Padayachee. Thank you for assisting me through my research and being a helping hand every time I needed you.
3. My family. Thank you for your unconditional love and support and for being a pillar of strength throughout this journey.
4. I thank National Research Foundation, South Africa, for funding my studies.

RESEARCH OUTPUTS

Research Articles

- The peer-reviewed article below has been published from my master's research work.

Nkosi, B.V.Z., Padayachee, T., Gront, D., Nelson, D.R. and Syed, K., 2022. Contrasting health effects of *Bacteroidetes* and *Firmicutes* lies in their genomes: Analysis of P450s, ferredoxins, and secondary metabolite clusters. **International journal of molecular sciences**, 23(9), p.5057.

- I supervised honors' projects and earned co-authorship in the articles below due to my contributions.

1. Msweli, S., Chonco, A., Msweli, L., Syed, P.R., Karpoormath, R., Chen, W., Gront, D., **Nkosi, B.V.Z.,** Nelson, D.R. and Syed, K., 2022. Lifestyles shape the cytochrome P450 repertoire of the bacterial phylum *Proteobacteria*. **International Journal of Molecular Sciences**, 23(10), p.5821.

2. Ngcobo, P.E., **Nkosi, B.V.Z.,** Nelson, D.R. and Syed. K. 2022. Evolution of cytochrome P450 enzymes in Archaea. **Molecular Biology and Evolution** (Under review).

Presentation

Nkosi, B.V.Z. and Syed, K. 2021. Genome data mining, annotation, and comparative analysis of P450s and their secondary metabolites in the bacterial phylum *Bacteroidetes*. Faculty of Science, Agriculture and Engineering Virtual Symposium, 20-21 October 2021, University of Zululand, KwaDlangezwa, KwaZulu-Natal, South Africa.

Convention attendance

Nkosi, B.V.Z. Bio Africa Convention, Durban ICC, 27-31 August 2022.

ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
CYP/P450	Cytochrome P450 monooxygenase
Fe-S	Iron-Sulfur
iTOL	Interactive Tree of Life
KEGG	Encyclopedia of Genes and Genomes
MAFFT	Multiple Alignment using Fast Fourier Transform
MeV	Multi Experiment Viewer
NCBI	National Centre for Biotechnology and Information
PKS	Polyketide Synthase
smBGCs	Secondary metabolite biosynthetic gene clusters
Transatpks-NRPS	Trans- acyltransferase PKS Non-Ribosomal Peptide Synthase
T-REX	Tree and Reticulogram Reconstruction

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1.Background and problem statement

The bacterial phylum *Bacteroidetes* consists of gram-negative bacteria (Lapébie et al., 2019, McKee et al., 2021). *Bacteroidetes* species are primary degraders of carbohydrates as they provide energy for their host by breaking down the polysaccharides their host cannot digest (Lapébie et al., 2019, McKee et al., 2021). The bacterial species can be found in all ecosystems, including fresh water and soil (Larsbrink and McKee, 2020, McKee et al., 2021). In humans, they are predominantly found in the gastrointestinal tract, inhabiting the distal gut (Johnson et al., 2017). They play a significant role in the gastrointestinal tract by interacting with the gut immune system, inhibiting the colonization of potential pathogens (Khan et al., 2021). Gut *Bacteroidetes* species produce acids, such as acetic acid, propionic acid, and succinic acid, as final products of their metabolism, which help kill pathogens (Khan et al., 2021).

This bacterial phylum consists of 130 genera (Kanehisa et al., 2019); *Bacteroides* are known for producing skimmed or low-fat milk and displaying postbiotic activities, such as promoting health and well-being (Carmen et al., 2021). *Bacteroidetes*, in association with other bacterial species belonging to the phylum *Firmicutes*, have been in focus for quite some time due to their impact on human health. These two bacterial groups represent 90% of the gut microbiota (Rinninella et al., 2019), indicating their importance in human health. As expected, due to their abundance in the human gut, the ratio of *Firmicutes* and *Bacteroidetes* was found to have contrasting effects on human health, including in obesity (high ratio) and inflammatory disease (low ratio) (Stojanov et al., 2020, Manor et al., 2020), albeit with some uncertainties (Magne et al., 2020). These two bacterial groups produce secondary metabolites that ultimately affect human health.

Despite the availability of many species' genomes belonging to these two phyla, comparative analyses of the gene pools responsible for these two bacterial groups' behavior and health effects on humans are scarcely reported. Genome-wide comparative analysis of a few species revealed that *Firmicutes* species have smaller genomes and a disproportionately smaller number of glycan degrading enzymes than *Bacteroidetes* species (Mahowald et al., 2009). A subsequent study involving 60 *Bacteroidetes* and 197 *Firmicutes* on host-synthesized mucin glycans revealed different glycosyl hydrolases patterns between *Bacteroidetes* and *Firmicutes*, indicating a preference for cleaved mucin glycans in the host (Ravcheev and Thiele, 2017). Analysis of chicken caecum gut microbiome revealed that *Bacteroidetes* and *Firmicutes* follow different strategies for colonization and coexistence in the intestinal tract

(Medvecky et al., 2018). A study detailing the genomic blueprint of the human gut microbiota revealed a large number of uncultured *Firmicutes* compared to *Bacteroidetes* (Almeida et al., 2019). This study also reported novel secondary metabolite biosynthetic gene clusters (smBGCs) coding for undiscovered natural compounds produced by the intestinal microbiota (Almeida et al., 2019).

In a recent study, *Firmicutes* species were found to have a large number of cytochrome P450 monooxygenases (CYPs/P450s) (Padayachee et al., 2020), and their redox proteins, ferredoxins (Nzuza et al., 2021a), in their genomes. P450s are heme–thiolate enzymes are known to play a role in an organism's primary and secondary metabolism. These enzymes are found in species across the biological kingdoms, including in nonliving entities such as viruses (Nelson, 2018, Lamb et al., 2019). It is now well-established that the P450 contingent of organisms is indicative of their lifestyle, because the lifestyle of an organism was found to affect the P450 gene pool in its genome (Syed et al., 2014, Padayachee et al., 2020, Khumalo, 2020, Kgosiemang et al., 2014, Ngwenya et al., 2018, Nzuza et al., 2021b, Msomi et al., 2021, Akapo et al., 2019, Qhanya et al., 2015, Malinga et al., 2022). The P450s were found to help organisms adapt to ecological niches, and organisms with parasitic, commensal, or adapted to living on simple carbon sources were found to have the lowest number of P450s in their genomes (Syed et al., 2014, Padayachee et al., 2020, Khumalo, 2020, Kgosiemang et al., 2014, Ngwenya et al., 2018, Nzuza et al., 2021b, Msomi et al., 2021, Akapo et al., 2019, Qhanya et al., 2015). P450s need electrons to perform their enzymatic reactions, supplied by redox proteins such as ferredoxins (Chiliza et al., 2020, Li et al., 2020). A recent study revealed many ferredoxins in *Firmicutes* species (Nzuza et al., 2021a). The study also suggested unique ferredoxins in *Firmicutes* species indicative of characteristics of the species in this phylum (Nzuza et al., 2021a). It is well-known that P450s, due to their regio- and stereo-specific oxidation capabilities, play a crucial role in producing secondary metabolites per se, contributing to the diversity of the secondary metabolites in an organism (Podust and Sherman, 2012, Greule et al., 2018). In a recent study, many P450s were part of smBGCs, indicating their role in secondary metabolite production in *Firmicutes* species (Padayachee et al., 2020).

Considering the above facts, especially the P450s and ferredoxins gene pools as characteristics of an organism's lifestyle and their role in the production of secondary metabolites, in this study, we selected these two sets of genes for comparative analysis of the gene pools of *Bacteroidetes* and *Firmicutes*. These two bacterial groups are known to produce secondary metabolites (Padayachee et al., 2020), and it is well-known that these secondary metabolites affect human health (Glowacki and Martens, 2020, Sharon et al., 2014, Man et al.,

2020). Thus, we performed a comprehensive comparative analysis of smBGCs to understand the rationale behind these two bacterial groups' distinct effects on human health regarding their secondary metabolism.

1.2. Novelty of the study

This study is the first to analyze P450s and ferredoxins in *Bacteroidetes* species. Furthermore, this study, for the first time, reports on a comparative analysis of P450s and ferredoxins between *Bacteroidetes* and *Firmicutes*. The study results were published in an accredited international journal with an impact factor of 6.2 (please see research outputs), highlighting the importance of the study.

1.3. Aims and Objectives

1.3.1. Study Aim

To perform a comparative analysis of P450s and their redox partners in the species of the bacterial phylum *Bacteroidetes* and identification of P450s involved in the biosynthesis of secondary metabolites.

1.3.2. Objectives

- To perform genome data mining and annotation of P450s in *Bacteroidetes* species.
- To perform data mining and analysis of redox proteins in *Bacteroidetes* species.
- To perform a phylogenetic analysis of P450s of *Bacteroidetes* species.
- To perform a P450 family conservation analysis in *Bacteroidetes* species.
- To perform genome data mining and identification of secondary metabolite biosynthetic gene clusters in *Bacteroidetes* species.
- To identify P450s involved in secondary metabolism in *Bacteroidetes* species.

1.4. Dissertation overview

Chapter 1:

Provides the background and problem statement, novelty, and aim and objectives of the study.

Chapter 2:

The literature review provides relevant background on P450s, ferredoxins, secondary metablolite gene clusters and *Bacteroidetes* species.

Chapter 3:

Provides methodologies related to P450s, ferredoxins, and secondary metabolite gene cluster analysis in *Bacteroidetes* species.

Chapter 4:

This chapter consists of results and discussion on P450s, ferredoxins, and secondary metabolite gene clusters in *Bacteroidetes* species. Furthermore, a comparative analysis of P450s, ferredoxins and secondary metabolite gene clusters between *Bacteroidetes* species with other bacterial species was also reported.

Chapter 5:

This chapter provides overall conclusions and future perspectives.

Please note that Chapters 3 to 5 are presented in the format of a research article published in the International Journal of Molecular Sciences (see research outputs section).

CHAPTER 2: LITERATURE REVIEW

2.1. *Bacteroidetes*

Bacteroidetes is a bacterial phylum that consists of gram-negative bacteria that are primarily degraders of carbohydrates (Lapébie et al., 2019, McKee et al., 2021). They provide energy for their host through fermentation by degrading polysaccharides that cannot be digested by their host (Lapébie et al., 2019, McKee et al., 2021). These bacterial species are found in all ecosystems, including fresh water and soil (Larsbrink and McKee, 2020, McKee et al., 2021). In humans, they are predominantly found in the gastrointestinal tract, inhabiting the distal gut (large intestine) (Johnson et al., 2017).

Species belonging to this phylum carry out beneficial activities in the human gut, such as transporting toxins and mutant compounds, breaking down biopolymers, especially polysaccharides, and breaking down enzymes targeting dietary polymers (Rios-Covian et al., 2015). They also play a significant role in the gastrointestinal tract by interacting with the gut immune system, inhibiting the colonization of potential pathogens (Khan et al., 2021). Gut *Bacteroidetes* species produce acids as final products of their metabolism, such as acetic acid, propionic acid, and succinic acid, that help kill pathogens (Khan et al., 2021).

Bacteroidetes and Firmicutes represent over 90% of the gastrointestinal microbiota (Rinninella et al., 2019). The ratio of these two groups of bacteria can be used as a marker for one's health concerning obesity and inflammatory bowel disease (Figure 2.1) (Stojanov et al., 2020, Manor et al., 2020, Magne et al., 2020). The increase in *Firmicutes* vs. *Bacteroidetes* was found to cause obesity, and the decrease in the ratio causes inflammatory disease (Figure 2.1). The ratio between these organisms is also significant in aging and colon cancer (Odamaki et al., 2016).

This bacterial phylum consists of 130 genera; *Bacteroides* are known for producing skimmed or low-fat milk in European countries, displaying postbiotic activities, such as promoting health and well-being (Carmen et al., 2021). *Zobellia galactanivorans*, a flavobacterium, degrade algal biomass (Thomas et al., 2011). *Spirosoma* species produce valuable human compounds such as terpenes, siderophores, arylpolyenes, bacteriocins, and resorcinols with neuroprotective, anti-tumorigenic, and anti-inflammatory properties, to name a few (Rojas et al., 2020). Apart from having many valuable bacteria within this phylum, species belonging to the *Empedobacter* genus were pathogenic to humans, causing diseases such as meningitis, endophthalmitis, urinary tract infection, and bacteremia (Bokhari et al., 2015, Sharma et al., 2016). Studies have shown that *Bacteroides fragilis* and other members of

the *Bacteroides* genus are the primary causes of anaerobic septicemia (Shanson, 2014). The genus *Elizabethkingia* consists of novel species such as *E. anophelis* and *E. meningosepta* which are gram-negative, non-spore-forming, non-motile, and rod-shaped and depict natural resistance to multiple known antibiotics (Kämpfer et al., 2011).

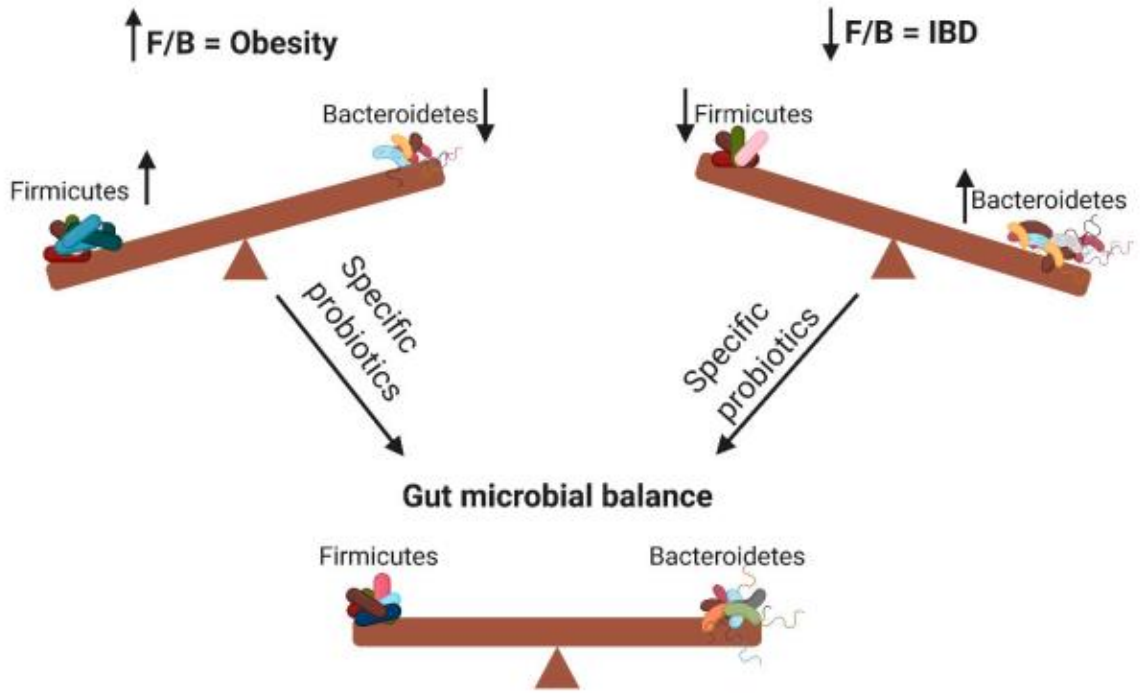


Figure 2.1. Schematic representation of *Firmicutes/ Bacteroidetes* ratio in relation to obesity and inflammatory bowel disease (Stojanov et al., 2020).

2.2. Secondary metabolites produced by *Bacteroidetes*

Secondary metabolites, also called specialized metabolites, toxins, secondary products, or natural products, are organic compounds produced by any lifeform, e.g., bacteria, fungi, animals, or plants, which are not directly involved in the normal growth, development, or reproduction of the organism. It is well-known that some *Bacteroidetes* species produce secondary metabolites of biotechnological potentials (Table 2.1). *Spirosoma* species produce valuable human compounds with neuroprotective, anti-tumorigenic, and anti-inflammatory properties, among others (Rojas et al., 2020). Some of the secondary metabolites produced by these species and their biological properties are listed in Table 2.1.

Table 2. 1. Secondary metabolites produced by *Bacteroidetes* species and their biological functions.

Species (Source)	**Secondary Metabolite**	**Biological Function**	**References**
Spirosoma spp. and *Sphingobacterium* sp. 21	Terpene	Neuroprotective, anti-tumorigenic, and anti-inflammatory	(Gershenzon and Dudareva, 2007, Rojas et al., 2020)
Flavobacterium spp.	Siderophore	Iron scavenger and transporter	(Belimov et al., 2005)
Spirosoma spp.; *Pedobacter cryoconitis*; *Sphingobacterium* sp. 21 and *Chitinophaga pinesis*	Type I PKS (Polyketide synthase) (T1PKS)	Useful as agrochemical and pharmaceutical compounds	(Katz and Baltz, 2016, Rojas et al., 2020)
Chinophaga pinesis; *Flavobacterium johnsoniae* UW10	Arylpolyene	Shields bacteria from reactive oxidation	(McBride et al., 2009, Schöner et al., 2014)
Spirosoma spp.; *Flavobacterium* spp.	Type III PKS (T3PKS)	Involved in hypocrellin and kanamycin synthesis	(Chaudhary et al., 2020, Rojas et al., 2020)
Spirosoma spp.; *Bacteroides* spp.	Bacteriocin	Anti-bacterial activity	(Walsh et al., 2015, Rojas et al., 2020)
Prevotella spp.	Resorcinol	Used to treat skin disorders	(Zhou et al., 2019)
Spirosoma spp.	Non-ribosomal peptide synthetase cluster (NRPs)	Antibiotic, immunosuppressant, and cytotoxic properties	(Rojas et al., 2020)
Spirosoma spp.	Lanthipeptide	Antimicrobial, antifungal, and antiviral activities	(Rojas et al., 2020)

2.3. P450s role in the production of secondary metabolites

P450s are versatile catalysts performing different reactions, including decarboxylation, desaturation, ring formation, and denitrification, to name a few (Figure 2.2) (Bernhardt, 2006). P450s play a significant role in producing secondary metabolites such as terpenes, carotenoids, polyketides, and macrolides (Podust and Sherman, 2012, Greule et al., 2018). In the mentioned reactions, the secondary metabolites produced include antibiotics, antimicrobial agents, and bio-reactive compounds (Kelly and Kelly, 2013). Some examples of P450s role in the production of different secondary metabolites include the oxidation of methylene and methyl groups by P450s to produce the antibiotic phenalinolactone (Figure 2.3) (Daum et al., 2010, Dürr et al., 2006, Xu et al., 2014); Synthesis of the diterpene isopimara-8, 15-dien-19-ol through hydroxylation of the methyl groups, and cyclooctatine stereo- and regiospecific hydroxylation by CYP1051A1 (Figure 2.3) (Xu et al., 2014, Kim et al., 2009). The P450s CYP175A1 and CYP287A1 catalyze carotenoid hydroxylation in different compounds (Figure 2.4) (Blasco et al., 2004, Zhou et al., 2015) and CYP106A2 hydrolyzing the steroid hormone progesterone to four products (Figure 2.5) (Lisurek et al., 2008, Zhang et al., 2020).

Hydrocarbon hydroxylation
Alkene epoxidation
Alkyne oxygenation
Arene epoxidation
Aromatic hydroxylation
N-Dealkylation
S- Dealkylation
O- Dealkylation
N-Hydroxylation
N-Oxidation
S-Oxidation
Oxidative deamination
Oxidative dehalogenation
Alcohol and aldehyde oxidations
Dehydrogenation
Dehydratations
Reductive dehalogenation
N-Oxide reduction
Epoxide reduction
Reductive β-scission of alkyl peroxides
NO reduction
Isomerizations
Oxidative C-C bond cleavage

Figure 2.2. Cytochrome P450s catalytic reactions (Bernhardt, 2006).

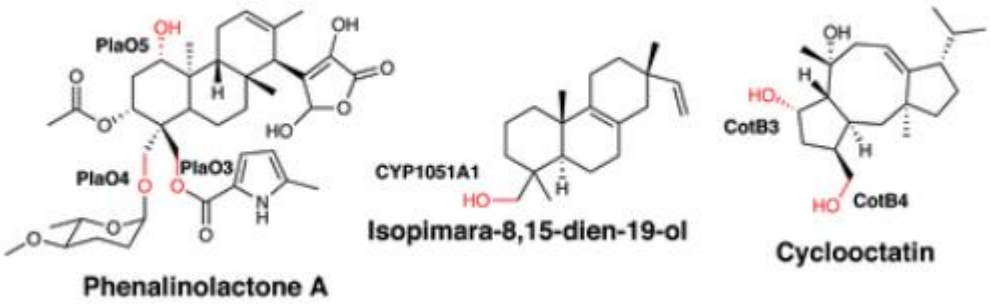


Figure 2.3. P450 hydroxylation of terpenes (Greule et al., 2018).

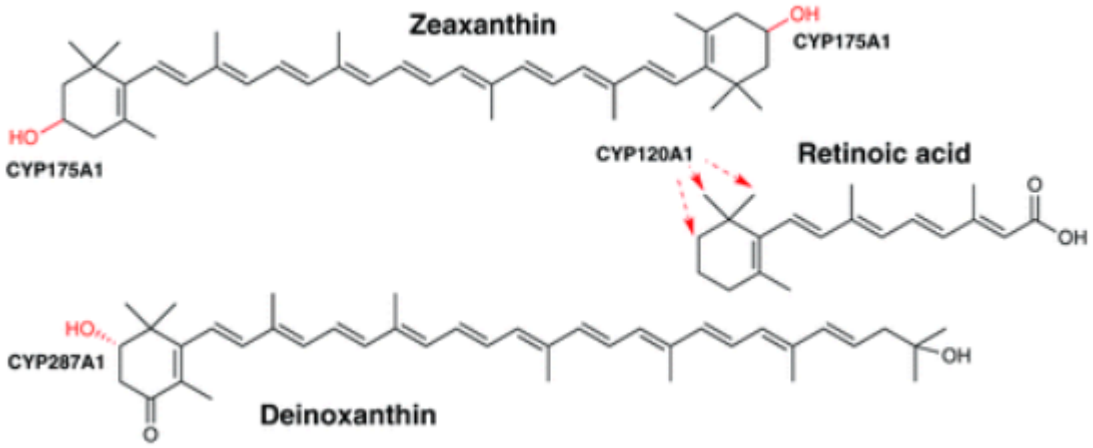


Figure 2.4. P450s role in carotenoid biosynthesis (Greule et al., 2018).

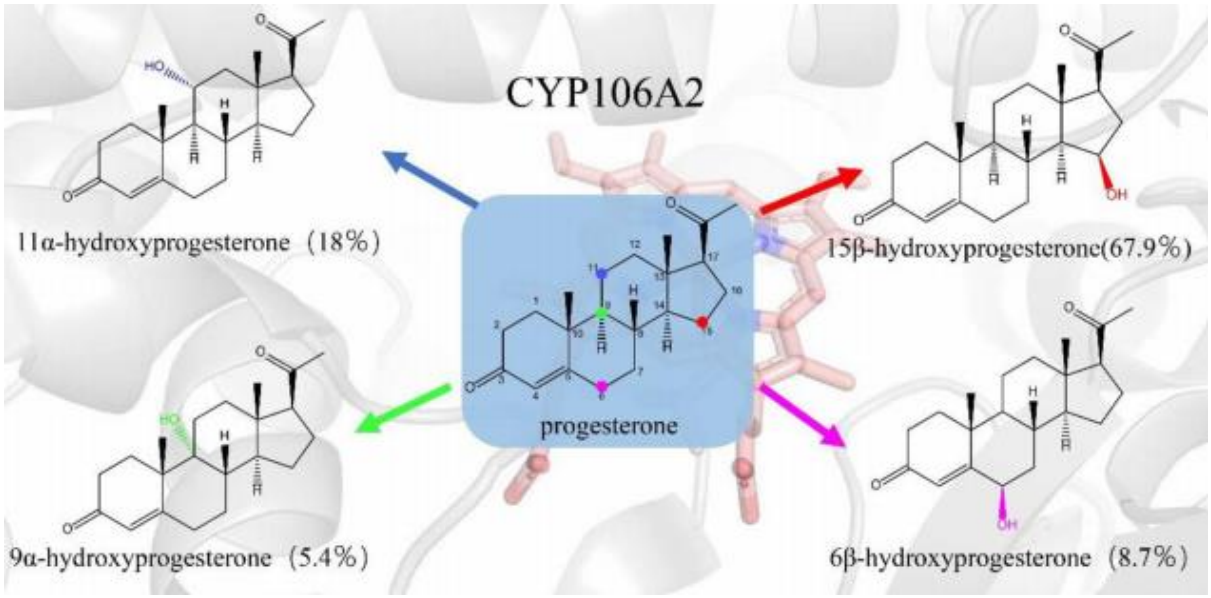


Figure 2.5. P450 mediated hydrolysis of progesterone into different steroids (Zhang et al., 2020).

2.4. P450s and their redox partners

P450s require a redox partner to perform their biological function (Hannemann et al., 2007). However, not all P450s need a redox partner, as some P450s are self-sufficient in such a way that they are fused to a redox partner (Guengerich and Munro, 2013, Lamb and Waterman, 2013, Sello et al., 2015). A protein partner is necessary to deliver one or more electrons to reduce the iron atom of the heme co-factor (Figure 2.6) (Werck-Reichhart and Feyereisen, 2000). These electrons are supplied by another protein known as redox proteins (Hannemann

et al., 2007). Many redox proteins can be found in organisms and have been classified into ten groups (Hannemann et al., 2007). These redox partners play an essential role as auxiliary proteins in protein transfer and the distribution of products (Guengerich, 2002). They do so by interacting with P450s, providing two electrons from NADPH or NADH to reduce the inert O_2 (Li et al., 2020). The electron transfer step is usually the rate-limiting step in the P450 catalytic cycle (Figure 2.6) (Li et al., 2020). An example of a redox protein is ferredoxin (Fdxs), which can be classified as [2Fe-2S], [3/4Fe-4S], [4Fe-4S], [3/4Fe-4S], high-potential iron-sulfur proteins and Rieske proteins (Hannemann et al., 2007, Zhang et al., 2020).

The regio-, chemo-, and stereo-selectivity of P450 reactions can be influenced by redox partners (Chiliza et al., 2020, Li et al., 2020). The versatility of the P450s *in vivo* is hypothesised to increase due to different redox partners having the ability to bind with different P450s (Chiliza et al., 2020, Li et al., 2020). Figure 2.7 illustrates the versatility of P450s and the various classifications of their P450-redox partner combinations (Li et al., 2020). There are ten classes of P450 systems, class I (A and B): mitochondrial and bacterial P450 system containing FAD-reductase and ferredoxin that reduces the P450 (Figure 2.7) (Bernhardt, 2006, Hannemann et al., 2007).

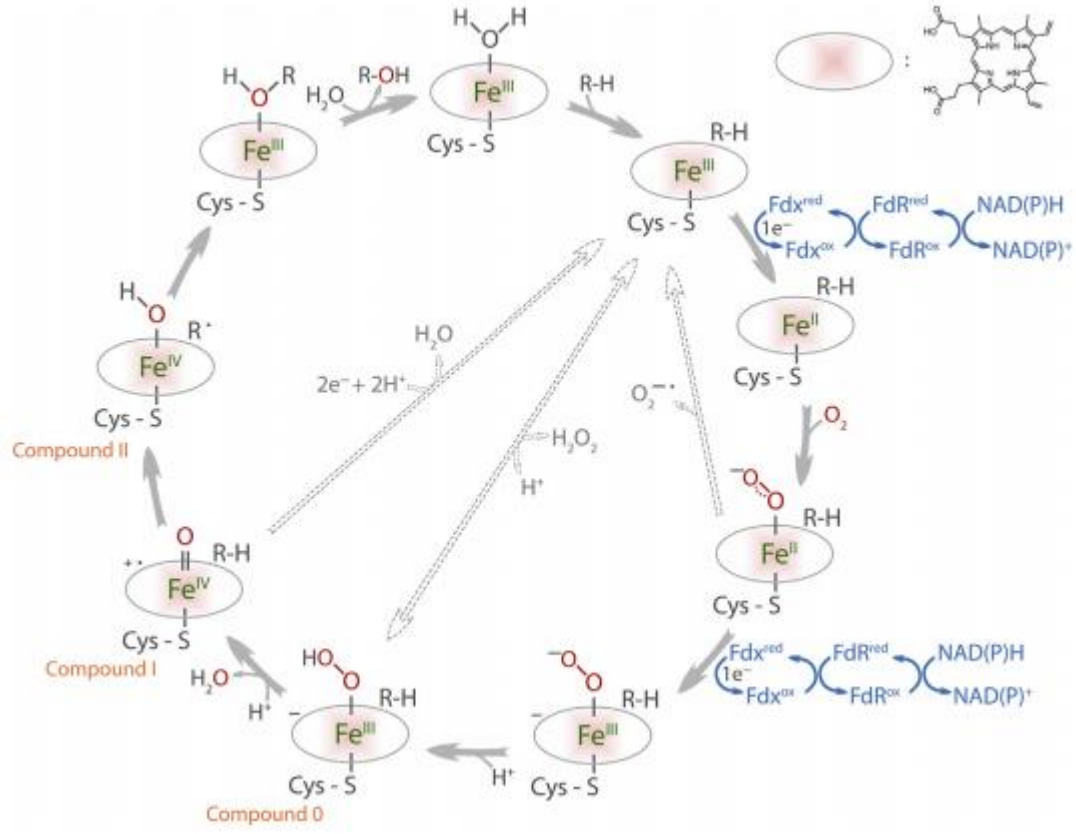


Figure 2.6. Cytochrome P450 cycle. The diagram also displays the relationship between P450s and ferredoxin redox partners (Li et al., 2020).

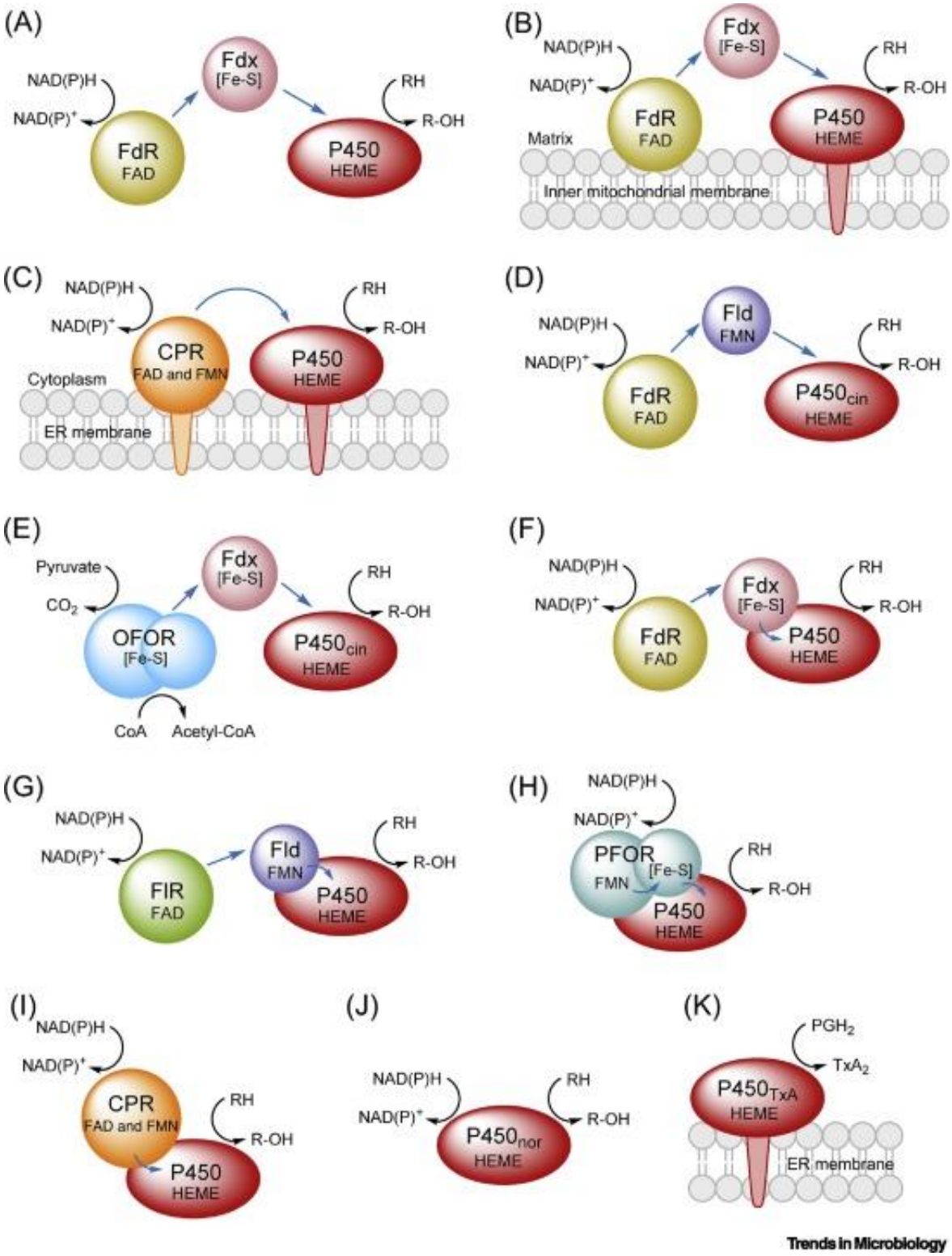


Figure 2.7. Schematic diagram representing P450 classification based on redox partners (Li et al., 2020).

CHAPTER 3: METHODOLOGY

3.1. Species and database

Genomes for 334 *Bacteroidetes* species, available for public use at Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2019), were used in the study for data mining of P450s, ferredoxins, and smBGCs. Information on genera, species names, species codes, and their genome IDs is presented in Table S1.

3.2. Genome data mining and annotation of P450s

Genome data mining and annotation of P450s were carried out using the standard procedure described previously by our laboratory (Msomi et al., 2021, Padayachee et al., 2020). Briefly, the proteome of each *Bacteroidetes* species was acquired from KEGG (Kanehisa et al., 2019) and submitted to the National Center for Biotechnology and Information (NCBI) Batch Web CD-Search Tool (Marchler-Bauer et al., 2017). The result was analyzed, and proteins that belong to the P450 superfamily were selected and searched for the presence of characteristic P450 motifs, EXXR, and CXG (Gotoh, 1992, Syed and Mashele, 2014). Proteins that were short in amino acid length and lacked one of the motifs were regarded as P450 fragments, and these P450 fragments were not considered for further analysis. The selected P450s were annotated (assigning the P450 family and P450 subfamily) following the International P450 Nomenclature Committee rules (Nelson et al., 1993, Nelson, 2006, Nelson, 1998). Proteins with a percentage identity greater than 55% were classified under the same subfamily, whereas those with a percentage identity greater than 40% were classified under the same family. Proteins with a percentage identity lower than 40% were classified under a new family.

3.3. Genome data mining and annotation of ferredoxins

Genome data mining and annotation of ferredoxins in *Bacteroidetes* species with P450s were carried out using our laboratory's recently published procedure (Nzuza et al., 2021a). Briefly, each of the *Bacteroidetes* species genomes was blasted using ferredoxins belonging to different Fe-S cluster types (Table S4), and the hit protein sequences were collected. The hit protein sequences were then subjected to the protein Basic Local Alignment Search Tool (BLAST) at the NCBI (Sayers et al., 2021) against the Protein Data Bank (PDB) database (consortium, 2019) and analyzed for the presence of characteristic motif of ferredoxins at the Pfam database (El-Gebali et al., 2019), InterPro database (Mitchell et al., 2019), and NCBI Conserved Domains Database (C.D.D.) (Lu et al., 2020). Proteins that had a hit against ferredoxins at the PDB database and have ferredoxin motifs, as indicated by different databases, were selected for further annotation. Annotation of ferredoxins (assigning Fe-S cluster subtypes) was carried

out based on the characteristic spacing patterns between cysteine amino acids of the Fe-S cluster-binding motif as described previously (Nzuza et al., 2021a). Ferredoxins belonging to the new subtypes were assigned a unique subtype number on par with the continuation of ferredoxin subtype numbers published for the species of *Alphaproteobacteria* and *Firmicutes* (Nzuza et al., 2021a). Some *Bacteroidetes* species ferredoxins were retrieved from the published article (Campbell et al., 2019) and annotated into different subtypes. These *Bacteroidetes* species names and their ferredoxins are indicated in Table S2.

3.4. Phylogenetic analysis of P450s

Phylogenetic analysis of P450s was carried out following the procedure described recently by our laboratory (Msomi et al., 2021, Padayachee et al., 2020). The phylogenetic tree of P450s was constructed using protein sequences (Supplementary Dataset S1). Firstly, the Multiple Alignment using Fast Fourier Transform (MAFFT) v6.864 (Katoh et al., 2005) was used to align the Trex web server's protein sequences (Boc et al., 2012). The alignments were then used to interpret the best tree by the Trex web server (Boc et al., 2012). Lastly, a web-based tool, VisuaLife, was used to create, visualize, and color the tree (Kryś and Gront, 2021).

3.5. Generation of P450 profile heat-maps

The generation of the heat map profile was carried out according to the method previously reported by our laboratory (Msomi et al., 2021, Padayachee et al., 2020). The data were represented as (−3) for P450 family/subtype absence (green) and (3) for P450 family/subtype presence (red). A tab-delimited file was imported into Multi-experiment viewer (Mev) (Howe et al., 2011). Hierarchical clustering using a Euclidean distance metric was used to cluster the data. P450 families formed the vertical axis, and *Bacteroidetes* species formed the horizontal axis.

3.6. smBGCs analysis and P450s identification

Secondary metabolite gene clusters (smBGCs) and the P450s part of the smBGCs were carried out following the procedure described by our laboratory (Padayachee et al., 2020, Nzuza et al., 2021b). Briefly, genome IDs of *Bacteroidetes* species (Table S1) were submitted to antibiotics & Secondary Metabolite Analysis Shell (anti-SMASH) (Blin et al., 2019) for the identification of secondary metabolite BGCs. Anti-SMASH results were downloaded in gene cluster sequences and Excel spreadsheets representing species-wise cluster information. P450s that formed part of a specific gene cluster were identified by manual data mining of gene cluster sequences. Standard gene cluster abbreviation terminology available in the anti-SMASH database (Blin et al., 2019) was maintained in this study.

3.7. Data analysis

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

3.8. Comparative analysis of P450s, ferredoxins, and smBGCs data

P450s, ferredoxins, and smBGCs data for *Firmicutes* species were retrieved from published articles (Nzuza et al., 2021a, Padayachee et al., 2020) and used for comparative analysis. Ferredoxins proteins used for data mining were retrieved from published articles (Saeki et al., 1991, Moulis et al., 1996, Dauter et al., 1997, Pochapsky et al., 1999, Macedo-Ribeiro et al., 2001, Bentley et al., 2002, Unciuleac et al., 2004, Koksharova et al., 2006, McLean et al., 2006, Frazão et al., 2008, Saridakis et al., 2009, Cai et al., 2017, Child et al., 2018, Ortega Ugalde et al., 2018, Lau et al., 2019).

CHAPTER 4: RESULTS AND DISCUSSION

4.1. Only a few *Bacteroidetes* species have P450s

Genome-wide analysis of P450s in 334 *Bacteroidetes* species belonging to 130 genera revealed the presence of P450s only in 77 *Bacteroidetes* species, indicating most of the species do not have P450s in their genome (Figure 4.1 and Table S1). This shows that only 23% of *Bacteroidetes* species have P450s in their genomes. Interestingly, 23% of *Firmicutes* species had P450s in their genomes (Padayachee et al., 2020), indicating that most of the species belonging to these two phyla do not have P450s. Analysis of *Bacteroidetes* genera disclosed that of the 130 genera, species belonging to 44 genera have P450s in their genomes (Figure 4.1 and Table S1). A point to be noted is that only a few species genomes are available in the 130 genera. Sometimes only a single species genome is available; thus, the future availability of more species genomes will provide more accurate information on P450s in these genera (Figure 4.1 and Table S1). However, this study analyzed a significant number of species belonging to genera such as *Bacteroides*, *Capnocytophaga*, and *Prevotella*, no P450s were found, suggesting that species in these genera probably do not have P450s (Figure 4.1 and Table S1).

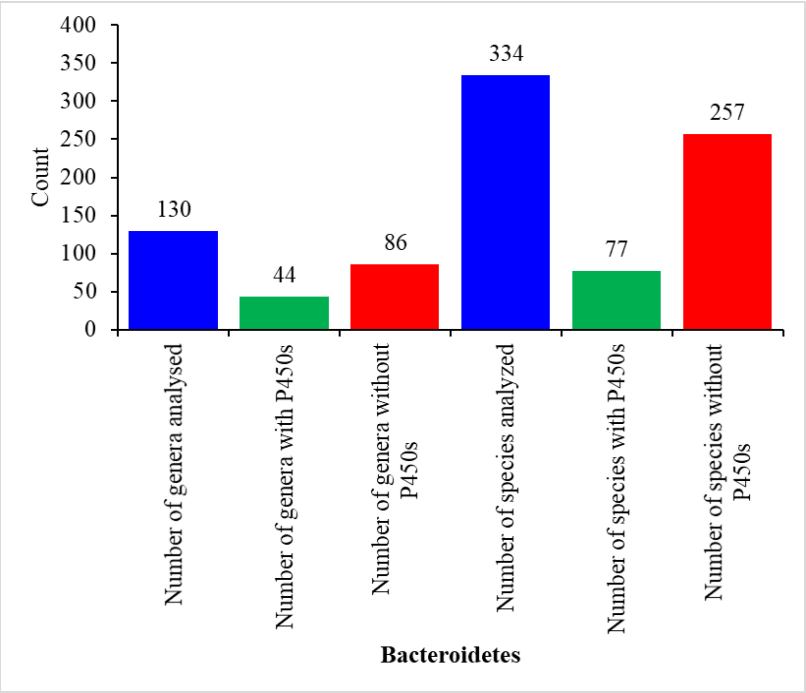


Figure 4.1. Analysis of P450s in *Bacteroidetes* species. The number next to the bar indicates the count for that category. Detailed analysis of the species, genera, and P450s are presented in Table S1.

Analysis of P450s in 77 *Bacteroidetes* species revealed the presence of 98 P450s in their genomes (Figure 4.2 and Table S1). The P450 count in the *Bacteroidetes* species ranged from a single P450 to three P450s; *Zunongwangia profunda* and *Flavivirga eckloniae* had the highest number of P450s (three P450s) in their genomes (Table S1). *Bacteroidetes* species were found to have the highest average number of P450s in their genomes compared to Gammaproteobacterial species but the lowest compared to other bacterial species (Table 4.1). *Bacteroidetes* species P450s identified in this study and their protein sequences and species are presented in Supplementary Dataset S1.

Table 4.1. Comparative analysis of key features of P450s and their association with secondary metabolism between *Bacteroidetes* species and different bacterial species. Abbreviations: No.: number, smBGCs: biosynthetic gene clusters.

	***Bacteroidetes* Species**	***Firmicutes* Species**	**Gammaproteobacterial Species**	***Streptomyces* Species**	**Mycobacterial Species**	**Cyanobacterial Species**
Total no. of species analyzed	334	972	1 261	203	60	114
No. of species with P450s	77	229	169	203	60	114
No. of P450s	98	712	277	5460	1784	341
No. of families	21	14	84	253	77	36
No. of subfamilies	28	53	105	698	132	79
Dominant P450 family	CYP1103	CYP107	CYP133	CYP107	CYP125	CYP110
Average no. of P450s	1	3	0.2	27	30	3
P450 diversity percentage	0.28	0.01	0.18	0.02	0.07	0.09
No. of P450s part of smBGCs	8	126	49	1231	204	27
Percentage of P450s part of BGCs	8	18	18	23	11	8
Reference(s)	This work	(Padayachee et al., 2020)	(Msomi et al., 2021)	(Senate et al., 2019, Mnguni et al., 2020)	(Senate et al., 2019, Parvez et al., 2016)	(Khumalo, 2020)

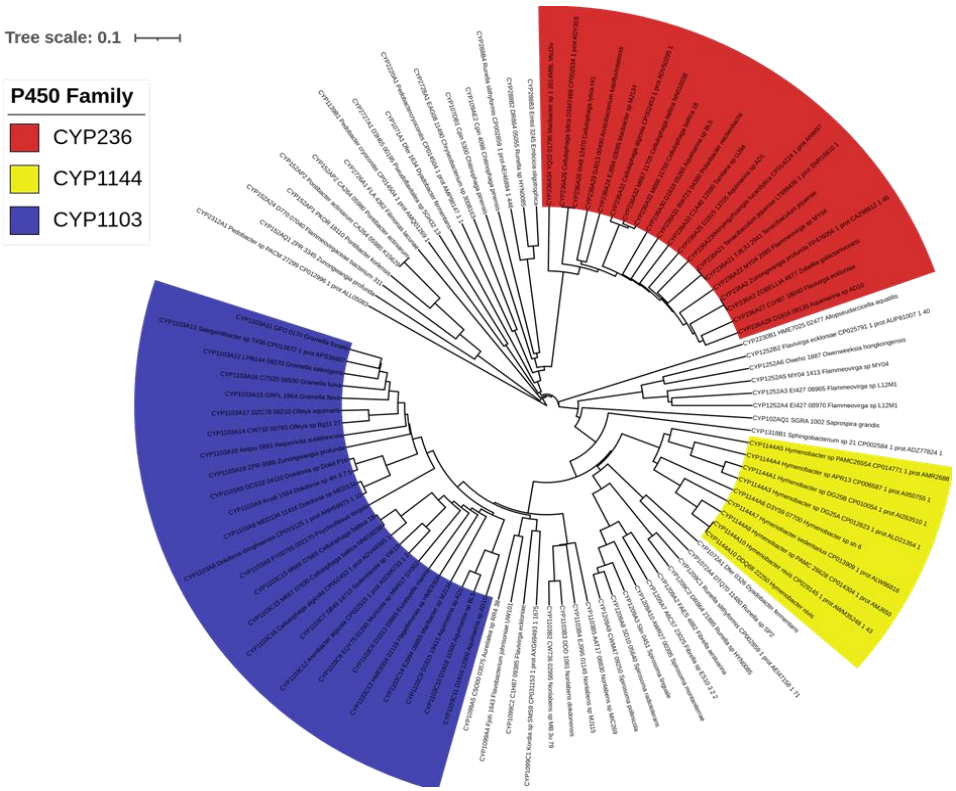


Figure 4.2. Phylogenetic analysis of *Bacteroidetes* species P450s. The P450 families that are expanded in these species are displayed in different colors.

4.2. *Bacteroidetes* species have the highest P450 diversity

Based on the International P450 Nomenclature Committee rules, i.e., percentage identity of >40% for a family and >55% for a subfamily (Nelson et al., 1993, Nelson, 2006, Nelson, 1998), and following the phylogenetic analysis of P450s, where P450s belonging to the same family are grouped (Figure 4.2), 98 P450s of *Bacteroidetes* species can be grouped into 21 P450 families and 28 P450 subfamilies (Table 4.2). The number of P450 families in *Bacteroidetes* species was higher than in *Firmicutes* (Table 4.1). However, the number of P450 families in *Bacteroidetes* species was the lowest compared to Gammaproteobacterial species, mycobacterial and cyanobacterial species, and *Streptomyces* species (Table 4.1). The comparative analysis of P450 diversity percentage among different bacterial groups revealed that *Bacteroidetes* species have the highest P450 diversity and *Firmicutes* species have the lowest P450 diversity (Table 4.1).

Table 4.2. Comparative analysis of P450 families and subfamilies in *Bacteroidetes* species.

P450 Family	Count	Percentage	P450 Subfamily	Count
CYP1103	29	30	A	13
			B	4
			C	12
CYP236	20	20	A	20
CYP1144	10	10	A	10
CYP1209	8	8	A	6
			C	2
CYP1252	5	5	A	4
			B	1
CYP152	5	5	A	1
			A.P.	3
			AQ	1
CYP1099	4	4	A	2
			C	2
CYP288	3	3	B	3
CYP1072	2	2	A	2
CYP102	1	1	AQ	1
CYP1071	1	1	A	1
CYP107	1	1	DB	1
CYP109	1	1	AE	1
CYP1139	1	1	B	1
CYP1318	1	1	B	1
CYP2220	1	1	A	1
CYP2230	1	1	B	1
CYP2312	1	1	A	1
CYP2726	1	1	A	1
CYP2727	1	1	A	1
CYP2728	1	1	A	1

Among P450 families, the CYP1103 has the highest number of members, with 29 P450s contributing 30% of total P450s in *Bacteroidetes* species (Table 4.2), followed by CYP236 (20 P450s contributing 20%) and CYP1144 (10 P450s contributing 10%) (Table 4.2). The number of members in the remaining 18 P450 families ranged from one to eight (Table 4.2). This indicates that the P450 families CYP1103, CYP236 and CYP1144 were expanded in *Bacteroidetes* species. The P450 family expansion was also observed in other bacterial species belonging to *Firmicutes* (Padayachee et al., 2020) and *Gammaproteobacteria* (Msomi et al., 2021). Comparative analysis of dominant P450 families across different bacterial species revealed that the CYP107 family was dominant in *Firmicutes* and *Streptomyces* species (Table 4.1). In contrast, different P450 families were prevalent in other bacterial groups (Table 4.1). Interestingly, P450 families such as CYP102, CYP107, and CYP109 had only one P450 each in *Bacteroidetes* species, although these are quite large families in other species (Padayachee et al., 2020, Senate et al., 2019, Mnguni et al., 2020, Khumalo, 2020, Parvez et al., 2016). The analysis of the P450 subfamilies revealed that 12 out of 21 P450 families had a single subfamily (Table 4.2). The P450 families with the most subfamilies were CYP1103 and CYP152, with three subfamilies each (Table 4.2). They were followed by P450 families such as CYP1099, CYP1209, and CYP1252, each having two subfamilies (Table 4.2). A particular subfamily was dominant when analyzing the P450 subfamilies in a specific family. In CYP1103 and CYP1252 families, the subfamily ‘A’ was predominant, and in the CYP152 family, the subfamily ‘AP’ was dominant (Table 4.2). The heat map analysis of P450 family profiles revealed that no P450 family was conserved across the *Bacteroidetes* species (Figure 4.3). However, based on the heat-map profile of P450 families, the P450 families CYP1103 and CYP236 were found to have co-presence in eight *Bacteroidetes* species (Figure 4.3).

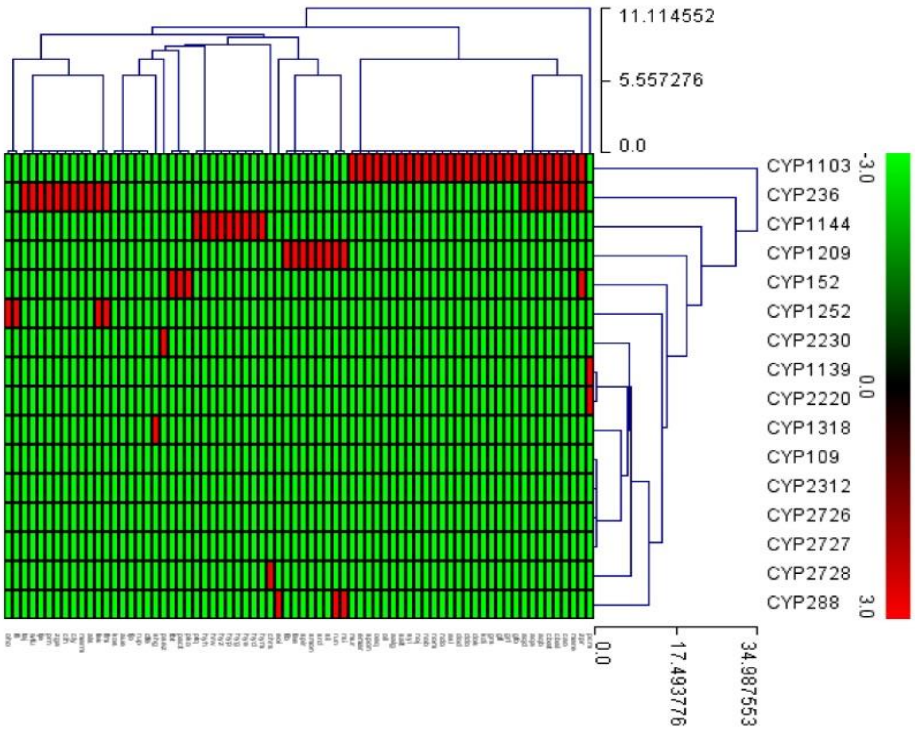


Figure 4.3. Analysis of P450 family absence/presence and co-occurrence in *Bacteroidetes* species. The data have been represented as −3 for family absence (green) and 3 for family presence (red). 77 *Bacteroidetes* species form the horizontal axis, and 21 P450 families form the vertical axis.

4.3. *Bacteroidetes*-, and *Firmicutes*-species have diverse P450 families in their genome

The P450 family level comparative analysis revealed that only four P450 families are commonly shared between *Bacteroidetes*- and *Firmicutes*-species (Figure 4.4), indicating these two bacterial groups have a diverse set of P450s in their genomes. In addition, the number of members in the commonly shared P450 families was found to be highly expanded in *Firmicutes* species. In contrast, in *Bacteroidetes* species, one (CYP102, CYP107, CYP109) to five members (CYP152) are present (Figure 4.4). This suggests that these two bacterial groups have different P450s in their genomes, indicating that P450s play different roles in their physiology, including producing different secondary metabolites.

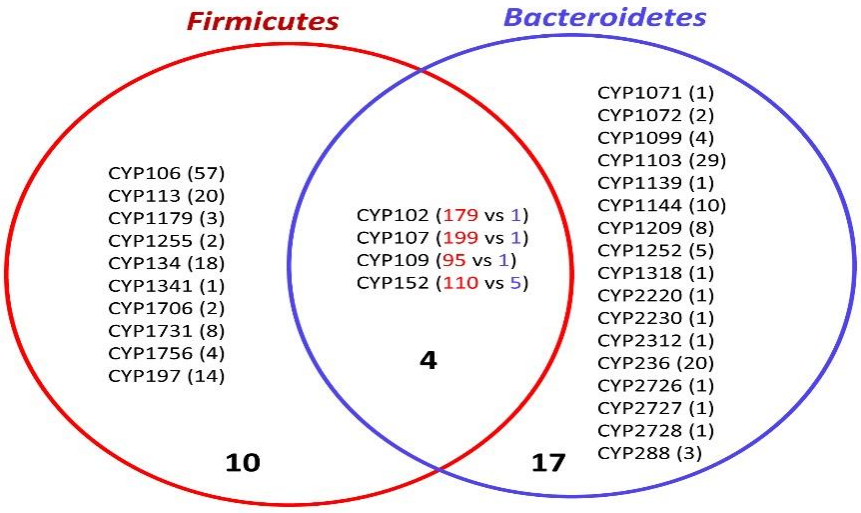


Figure 4.4. Comparative analysis of P450 families between *Bacteroidetes*- and *Firmicutes*-species. The number in parenthesis indicates the number of members in a P450 family. The numbers indicated with red and blue colors represent the P450 family count for *Firmicutes*- and *Bacteroidetes*-species, respectively. Numbers in bold indicate the number of P450 families.

4.4. *Bacteroidetes* species have a large and diverse number of secondary metabolite BGCs

The analysis of secondary metabolite biosynthetic gene clusters (smBGCs) revealed many smBGCs in *Bacteroidetes* species compared to *Firmicutes* species (Figure 4.5). In total, 269 *Bacteroidetes* species have 1297 smBGCs, with an average of 4.8 smBGCs in their genome, whereas 229 *Firmicutes* species have 126 smBGCs, with an average of 0.5 smBGCs in their genome, indicating the lowest number of smBGCs in *Firmicutes* species. This suggests that *Bacteroidetes* species produce more secondary metabolites compared to *Firmicutes* species. Analysis of smBGCs revealed the presence of 30 cluster types in *Bacteroidetes* species, compared to only 15 cluster types in *Firmicutes* species (Figure 4.5). This further indicates that *Bacteroidetes* species produce numerous highly diverse secondary metabolites compared to *Firmicutes*.

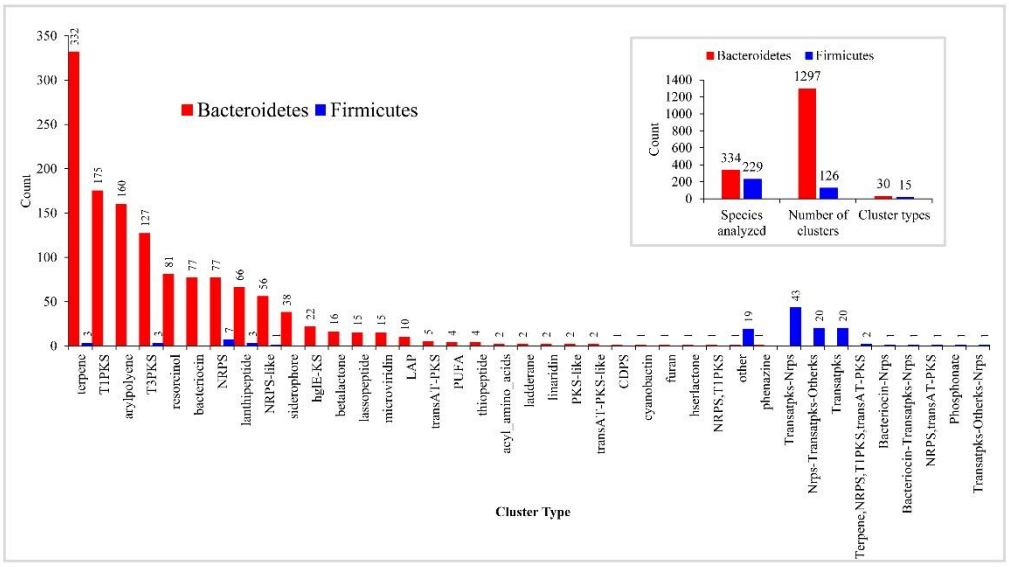


Figure 4.5. Comparative analysis of secondary metabolite biosynthetic gene clusters (smBGCs) between *Bacteroidetes*- and *Firmicutes*-species. The main panel compares cluster types, and the inset panel represents overall features between these two bacterial groups. The number next to the bar indicates the count for that category. The cluster type names and abbreviations used in the figure are the standard abbreviations proposed by anti-SMASH (Blin et al., 2019).

Among 30 types of smBGCs found in *Bacteroidetes* species, terpene was dominant, followed by Type I PKS (Polyketide synthase) (T1PKS), Aryl polyene cluster (arylpolyene), Type III PKS (T3PKS) (Figure 12). Altogether, these four cluster types contribute to 60% of smBGCs in *Bacteroidetes* species (Figure 4.5). Among 15 types of smBGCs found in *Firmicutes* species, Trans-AT PKS- Non-ribosomal peptide synthetase cluster (Transatpks-Nrps) was dominant, followed by Nrps-Transatpks-Otherks and Transatpks (Figure 4.5). The smBGCs abbreviations used here are the standard abbreviations proposed by anti-SMASH (Blin et al., 2019).

The comparative analysis of cluster types revealed that *Bacteroidetes* species and *Firmicutes* species only share six cluster types, indicating the distinct nature of smBGCs between these two bacterial groups (Figure 4.5). However, the number of smBGCs in these six cluster types significantly differed between these species. In the case of the terpene cluster type, 332 smBGCs were found in *Bacteroidetes* species, whereas only three smBGCs were found in *Firmicutes* species. The difference was also evident for cluster types T3PKS, NRPS, lanthipeptide, and NRPS-like, where many smBGCs were found in *Bacteroidetes* species. In contrast, smBGCs were limited to a single digit in *Firmicutes* species (Figure 4.5). Most of the

smBGCs have no similarity to known smBGCs, indicating *Bacteroidetes* species smBGCs encode novel secondary metabolites.

Considering the above facts, we propose that the contrasting effects of these two bacterial groups on hosts and organisms are partly due to the distinct nature of secondary metabolites produced by these organisms.

4.5. *Bacteroidetes* species P450s has a minor role in secondary metabolism

Analysis of the P450s part of smBGCs revealed that only eight P450s (8%) are part of these clusters (Table 4.3), indicating P450s play a minor role in secondary metabolism in *Bacteroidetes* species. In contrast to *Bacteroidetes* species' P450s, 18% of *Firmicutes* species' P450s were part of smBGCs (Table 4.1), indicating *Firmicutes* species P450s play a significant role in secondary metabolism. The percentage of the P450s part of smBGCs in *Bacteroidetes* species was the lowest compared to other bacterial groups (Table 4.1). Of the 21 P450 families, only 5 formed part of the smBGCs in *Bacteroidetes* species (Table 4.3). Among these families, four members were from the CYP1209 family. Only one member from the P450 families, CYP109, CYP109, CYP1139, and CYP1318, was part of smBGCs (Table 4.3). The connection between *Bacteroidetes* species' P450 families and secondary metabolite cluster type revealed that the P450 family CYP1209 is mainly associated with biosynthetic gene cluster terpene (Table 4.3). Two P450s, CYP109 and CYP107, from the same species, *Chitinophaga pinensis*, were part of different cluster types (Table 4.3), indicating their association in producing different secondary metabolites.

Table 4.3. Comparative analysis of P450s involved in secondary metabolism in *Bacteroidetes* species. The cluster type names/abbreviations used in the table are the standard abbreviations proposed by anti-SMASH [56].

Species Name	P450	Cluster Type
Chitinophaga pinensis	CYP109	terpene
	CYP107	NRPS
Pedobacter cryoconitis	CYP1139	lanthipeptide
Sphingobacterium sp. 21	CYP1318	NRPS-like
Spirosoma linguale	CYP1209	terpene
Spirosoma radiotolerans	CYP1209	terpene
Spirosoma montaniterrae	CYP1209	terpene
Spirosoma pollinicola	CYP1209	terpene

4.6. *Bacteroidetes*- and *Firmicutes*-species have highly diverse ferredoxins in their genomes

Genome data mining and annotation of ferredoxins in 104 *Bacteroidetes* species revealed the presence of 269 ferredoxins in their genomes (Figure 4.6 and Table S2). Among *Bacteroidetes* species, *Tenacibaculum jejuense* has the highest number of six ferredoxins (Table S2). *Bacteroidetes* species were found to have double the number of ferredoxins in their genomes compared to *Firmicutes* species, as the average number of ferredoxins was 2.6 in *Bacteroidetes* species compared to 1.2 in *Firmicutes* species (Padayachee et al., 2020). The 269 ferredoxins found in *Bacteroidetes* species can be grouped into five Fe-S cluster types: 2Fe-2S, 3Fe-4S, 4Fe-4S, 2[4Fe-4S], and 2[4Fe-4S]Alv (Figure 4.6 and Table S2). The 7Fe-8S cluster-type ferredoxins were not found in the *Bacteroidetes* species analyzed in this study. Of the five Fe-S cluster types found in *Bacteroidetes* species, the 2Fe-2S was the most abundant, with 136 ferredoxins, followed by 2[4Fe-4S]Alv, with 107 ferredoxins (Figure 4.6). In comparison to *Bacteroidetes* species, *Firmicutes* species had only four Fe-S cluster types, such as 2Fe-2S, 4Fe-4S, 7Fe-8S, 2[4Fe-4S], in their genomes (Figure 4.6), indicating the absence of 3Fe-4S and 2[4Fe-4S]Alv Fe-S cluster ferredoxins. Further differences were observed concerning the number of ferredoxins in the common Fe-S cluster types found in these two bacterial groups (Figure 4.6). *Bacteroidetes* species have more 2Fe-2S cluster ferredoxins, whereas *Firmicutes* species have more 4Fe-4S and 2[4Fe-4S] cluster ferredoxins (Figure 4.6).

Overall, 4Fe-4S cluster-type ferredoxins and 2[4Fe-4S]Alv cluster-type ferredoxins were most abundant in *Firmicutes* species and *Bacteroidetes* species, respectively (Figure 4.6). This suggests that these two bacterial groups have different preferences for Fe-S cluster type.

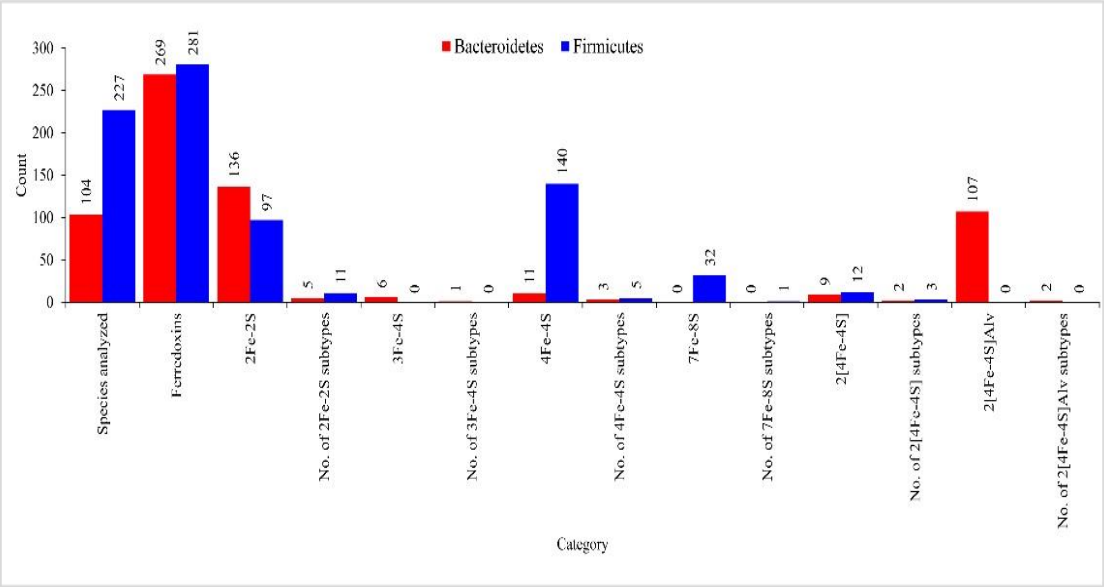


Figure 4.6. Comparative analysis of ferredoxins iron-sulfur (Fe-S) cluster features between *Bacteroidetes*- and *Firmicutes*-species. The number next to the bar indicates the count for that category.

Based on the amino acid spacing pattern analysis between the cysteine amino acids of the Fe-S cluster binding motif (Nzuza et al., 2021a), 136 and 97 2Fe-2S ferredoxins of *Bacteroidetes* and *Firmicutes* species can be grouped into 5 and 11 subtypes (Figure 4.6 and Table S3). Among *Bacteroidetes* species 2Fe-2S ferredoxin subtypes, subtype 18 has the most ferredoxins, followed by subtype 4 (Figure 4.6), indicating these species highly prefer subtype 18 ferredoxins. The comparative analysis revealed that three subtypes were shared between the *Bacteroidetes* species and the *Firmicutes* species (Figure 4.6), suggesting the common ancestral origin of these ferredoxin subtypes (Nzuza et al., 2021a). Six 3Fe-4S ferredoxins found in *Bacteroidetes* species can be grouped into a single subtype 8 (Table S3).

Eleven 4Fe-4S ferredoxins found in *Bacteroidetes* species can be grouped into three subtypes (Figure 4.6 and Table S3). Of the three subtypes, subtype 13 ferredoxins were found in higher numbers (Table S3). Contrary to the 2Fe-2S ferredoxin subtypes, no common 4Fe-4S subtypes were found between these two bacterial groups (Figure 4.6), indicating that 4Fe-4S ferredoxins are highly diverse in these two bacterial groups. Nine 2[4Fe-4S] ferredoxins

found in *Bacteroidetes* species can be grouped into two subtypes (Table S3). Of the two subtypes, subtype 34 ferredoxins were found in higher numbers (Table S3). There were no common 2[4Fe-4S] subtypes between these two bacterial groups (Table S3). The 107 2[4Fe-4S]Alv ferredoxins of *Bacteroidetes* species were grouped into two subtypes (Table S3). Of the two subtypes, subtype 11 has more ferredoxins than subtype 12 (Table S3). Ferredoxin sequences identified in this study and their subtypes were presented in Supplementary Dataset S2.

CHAPTER 5: CONCLUSION AND FUTURE PERSPECTIVES

Each organism belonging to a particular group is different because it has a characteristic gene pool that is ultimately responsible for its behavior. This study attempts to understand the gene pools of two different bacterial groups, *Bacteroidetes* and *Firmicutes*, that make up more than 90% of the human gut and exert distinct effects on human health. Based on their distinct health effects, one can expect diversity in their gene pools. As expected, these two bacterial groups were found to have a diverse set of cytochrome P450 monooxygenases (CYPs/P450s) and ferredoxins genes in their genome. Annotation and classification of P450s and ferredoxins revealed that *Bacteroidetes* species have more P450 families and ferredoxin subtypes than *Firmicutes* species. A point to note is that the Alvin ferredoxins (2[4Fe-4S]Alv) are expanded in the *Bacteroidetes* species, although this is not observed in *Firmicutes* and Alphaproteobacterial species. This indicates gene pool diversity in these two sets of genes in these organisms. Furthermore, very few P450s were found to be part of secondary metabolism in *Bacteroidetes* species compared to *Firmicutes* species.

This study strongly supports the hypothesis put forward by our laboratory that organisms' lifestyles influence the P450 contingent in their genomes. The commensal, pathogenic lifestyle of *Bacteroidetes* resulted in the loss of P450s in their genomes; a few species have P450s in this phylum. The same phenomenon was observed in *Firmicutes* species and Betaproteobacterial species. Analysis of secondary metabolites biosynthetic gene clusters (smBGCs) revealed that *Bacteroidetes* species have many cluster types compared to *Firmicutes* species, indicating that the former produces a more diverse array of secondary metabolites. Furthermore, the smBGCs in these two bacterial groups were distinct, indicating that these species produce different secondary metabolites and, as a result, distinct health effects on humans. A point to note is that, unlike *Firmicutes* species smBGCs (Padayachee et al., 2020), *Bacteroidetes* species smBGCs have less or almost no similarity to known smBGCs indicating these clusters encode novel secondary metabolites.

Results from this study serve as a reference for further analysis of the gene pools and characterization of secondary metabolites from these two bacterial groups. This study is the first report on a comparative analysis of P450s, ferredoxins, and smBGCs between *Bacteroidetes* and *Firmicutes* species.

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.

ALMEIDA, A., MITCHELL, A. L., BOLAND, M., FORSTER, S. C., GLOOR, G. B., TARKOWSKA, A., LAWLEY, T. D. & FINN, R. D. 2019. A new genomic blueprint of the human gut microbiota. *Nature,* 568**,** 499-504.

BELIMOV, A., HONTZEAS, N., SAFRONOVA, V., DEMCHINSKAYA, S., PILUZZA, G., BULLITTA, S. & GLICK, B. 2005. Cadmium-tolerant plant growth-promoting bacteria associated with the roots of Indian mustard (Brassica juncea L. Czern.). *Soil Biology and Biochemistry,* 37**,** 241-250.

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**,** 141-147.

BERNHARDT, R. 2006. Cytochromes P450 as versatile biocatalysts. *Journal of biotechnology,* 124**,** 128-145.

BLASCO, F., KAUFFMANN, I. & SCHMID, R. D. 2004. CYP175A1 from Thermus thermophilus HB27, the first beta-carotene hydroxylase of the P450 superfamily. *Appl Microbiol Biotechnol,* 64**,** 671-4.

BLIN, K., PASCAL ANDREU, V., DE LOS SANTOS, E. L. C., DEL CARRATORE, F., LEE, S. Y., MEDEMA, M. H. & WEBER, T. 2019. The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res,* 47**,** D625-d630.

BOC, A., DIALLO, A. B. & MAKARENKOV, V. 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res,* 40**,** W573-9.

BOKHARI, S., ABBAS, N., SINGH, M., CINDRICH, R. B. & ZEANA, C. 2015. Empedobacter brevis bacteremia in a patient infected with HIV: case report and review of literature. *Case reports in infectious diseases,* 2015.

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.

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**,** 79.

CARMEN, C. M., SEPPO, S. & VINDEROLA, G. 2021. Postbiotics: defining the impact of inactivated microbes and their metabolites on promotion of health. *The Human Microbiome in Early Life.* Elsevier.

CHAUDHARY, D. K., DAHAL, R. H., KIM, D.-U. & KIM, J. 2020. Flavobacterium sandaracinum sp. nov., Flavobacterium caseinilyticum sp. nov., and Flavobacterium hiemivividum sp. nov., novel psychrophilic bacteria isolated from Arctic soil. *International journal of systematic and evolutionary microbiology,* 70**,** 2269-2280.

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. *Biophysical Reviews***,** 1-6.

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

DAUM, M., SCHNELL, H. J., HERRMANN, S., GÜNTHER, A., MURILLO, R., MÜLLER, R., BISEL, P., MÜLLER, M. & BECHTHOLD, A. 2010. Functions of genes and enzymes involved in phenalinolactone biosynthesis. *Chembiochem,* 11**,** 1383-91.

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.

DÜRR, C., SCHNELL, H. J., LUZHETSKYY, A., MURILLO, R., WEBER, M., WELZEL, K., VENTE, A. & BECHTHOLD, A. 2006. Biosynthesis of the terpene phenalinolactone in Streptomyces sp. Tü6071: analysis of the gene cluster and generation of derivatives. *Chem Biol,* 13**,** 365-77.

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.

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.

GERSHENZON, J. & DUDAREVA, N. 2007. The function of terpene natural products in the natural world. *Nature chemical biology,* 3**,** 408-414.

GLOWACKI, R. W. & MARTENS, E. C. 2020. In sickness and health: Effects of gut microbial metabolites on human physiology. *PLoS Pathogens,* 16**,** e1008370.

GOTOH, O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry,* 267**,** 83-90.

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.

GUENGERICH, F. P. 2002. Rate-limiting steps in cytochrome P450 catalysis.

GUENGERICH, F. P. & MUNRO, A. W. 2013. Unusual cytochrome p450 enzymes and reactions. *J Biol Chem,* 288**,** 17065-73.

HANNEMANN, F., BICHET, A., EWEN, K. M. & BERNHARDT, R. 2007. Cytochrome P450 systems—biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1770**,** 330-344.

HOWE, E. A., SINHA, R., SCHLAUCH, D. & QUACKENBUSH, J. 2011. RNA-Seq analysis in MeV. *Bioinformatics,* 27**,** 3209-3210.

JOHNSON, E., HEAVER, S., WALTERS, W. & LEY, R. 2017. Microbiome and metabolic disease: revisiting the bacterial phylum Bacteroidetes. *Journal of Molecular Medicine,* 95.

KÄMPFER, P., MATTHEWS, H., GLAESER, S. P., MARTIN, K., LODDERS, N. & FAYE, I. 2011. Elizabethkingia anophelis sp. nov., isolated from the midgut of the mosquito Anopheles gambiae. *International Journal of Systematic and Evolutionary Microbiology,* 61**,** 2670-2675.

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.

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.

KATZ, L. & BALTZ, R. H. 2016. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology and Biotechnology,* 43**,** 155-176.

KELLY, S. L. & KELLY, D. E. 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20120476.

KGOSIEMANG, I. K. R., SYED, K. & MASHELE, S. S. 2014. Comparative genomics and evolutionary analysis of cytochrome P450 monooxygenases in fungal subphylum *Saccharomycotina*. *J Pure Appl Microbiol,* 8**,** 12.

KHAN, I., BAI, Y., ZHA, L., ULLAH, N., ULLAH, H., SHAH, S. R. H., SUN, H. & ZHANG, C. 2021. Mechanism of the Gut Microbiota Colonization Resistance and Enteric Pathogen Infection. *Frontiers in cellular and infection microbiology***,** 1273.

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

KIM, S. Y., ZHAO, P., IGARASHI, M., SAWA, R., TOMITA, T., NISHIYAMA, M. & KUZUYAMA, T. 2009. Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two p450 hydroxylases. *Chem Biol,* 16**,** 736-43.

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

KRYŚ, J. D. & GRONT, D. 2021. VisuaLife: Library for interactive visualization in rich web applications. *Bioinformatics*.

LAMB, D. C., FOLLMER, A. H., GOLDSTONE, J. V., NELSON, D. R., WARRILOW, A. G., PRICE, C. L., TRUE, M. Y., KELLY, S. L., POULOS, T. L. & STEGEMAN, J. J. 2019. On the occurrence of cytochrome P450 in viruses. *Proceedings of the National Academy of Sciences,* 116**,** 12343-12352.

LAMB, D. C. & WATERMAN, M. R. 2013. Unusual properties of the cytochrome P450 superfamily. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20120434.

LAPÉBIE, P., LOMBARD, V., DRULA, E., TERRAPON, N. & HENRISSAT, B. 2019. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nature Communications,* 10**,** 2043.

LARSBRINK, J. & MCKEE, L. S. 2020. Bacteroidetes bacteria in the soil: Glycan acquisition, enzyme secretion, and gliding motility. *Advances in applied microbiology,* 110**,** 63-98.

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.

LI, S., DU, L. & BERNHARDT, R. 2020. Redox partners: function modulators of bacterial P450 enzymes. *Trends in microbiology,* 28**,** 445-454.

LISUREK, M., SIMGEN, B., ANTES, I. & BERNHARDT, R. 2008. Theoretical and experimental evaluation of a CYP106A2 low homology model and production of mutants with changed activity and selectivity of hydroxylation. *ChemBioChem,* 9**,** 1439-1449.

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.

MACEDO-RIBEIRO, S., MARTINS, B. M., BARBOSA PEREIRA, P., 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.

MAGNE, F., GOTTELAND, M., GAUTHIER, L., ZAZUETA, A., PESOA, S., NAVARRETE, P. & BALAMURUGAN, R. 2020. The firmicutes/bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients? *Nutrients,* 12**,** 1474.

MAHOWALD, M. A., REY, F. E., SEEDORF, H., TURNBAUGH, P. J., FULTON, R. S., WOLLAM, A., SHAH, N., WANG, C., MAGRINI, V. & WILSON, R. K. 2009. Characterizing a model human gut microbiota composed of members of its two

dominant bacterial phyla. *Proceedings of the National Academy of Sciences,* 106**,** 5859-5864.

MALINGA, N. A., NZUZA, N., PADAYACHEE, T., SYED, P. R., KARPOORMATH, R., GRONT, D., NELSON, D. R. & SYED, K. 2022. An Unprecedented Number of Cytochrome P450s Are Involved in Secondary Metabolism in Salinispora Species. *Microorganisms,* 10**,** 871.

MAN, A. W., ZHOU, Y., XIA, N. & LI, H. 2020. Involvement of gut microbiota, microbial metabolites and interaction with polyphenol in host immunometabolism. *Nutrients,* 12**,** 3054.

MANOR, O., DAI, C. L., KORNILOV, S. A., SMITH, B., PRICE, N. D., LOVEJOY, J. C., GIBBONS, S. M. & MAGIS, A. T. 2020. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nature communications,* 11**,** 1-12.

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.

MCBRIDE, M. J., XIE, G., MARTENS, E. C., LAPIDUS, A., HENRISSAT, B., RHODES, R. G., GOLTSMAN, E., WANG, W., XU, J. & HUNNICUTT, D. W. 2009. Novel features of the polysaccharide-digesting gliding bacterium Flavobacterium johnsoniae as revealed by genome sequence analysis. *Applied and environmental microbiology,* 75**,** 6864-6875.

MCKEE, L. S., LA ROSA, S. L., WESTERENG, B., EIJSINK, V. G., POPE, P. B. & LARSBRINK, J. 2021. Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature. *Environmental Microbiology Reports,* 13**,** 559-581.

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**,** 8427-8443.

MEDVECKY, M., CEJKOVA, D., POLANSKY, O., KARASOVA, D., KUBASOVA, T., CIZEK, A. & RYCHLIK, I. 2018. Whole genome sequencing and function prediction

of 133 gut anaerobes isolated from chicken caecum in pure cultures. *BMC genomics,* 19**,** 1-15.

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.

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.

MSOMI, N. N., PADAYACHEE, T., NZUZA, N., SYED, P. R., KRYŚ, J. D., CHEN, W., GRONT, D., NELSON, D. R. & SYED, K. 2021. In silico analysis of P450s and their role in secondary metabolism in the bacterial class Gammaproteobacteria. *Molecules,* 26**,** 1538.

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.

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.

NZUZA, N., PADAYACHEE, T., CHEN, W., GRONT, D., NELSON, D. R. & SYED, K. 2021a. Diversification of Ferredoxins across Living Organisms. *Current Issues in Molecular Biology,* 43**,** 1374-1390.

NZUZA, N., PADAYACHEE, T., SYED, P. R., KRYŚ, J. D., CHEN, W., GRONT, D., NELSON, D. R. & SYED, K. 2021b. Ancient Bacterial Class Alphaproteobacteria

Cytochrome P450 Monooxygenases Can Be Found in Other Bacterial Species. *International journal of molecular sciences,* 22**,** 5542.

ODAMAKI, T., KATO, K., SUGAHARA, H., HASHIKURA, N., TAKAHASHI, S., XIAO, J.-Z., ABE, F. & OSAWA, R. 2016. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiology,* 16**,** 90.

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

PADAYACHEE, T., NZUZA, N., CHEN, W., NELSON, D. R. & SYED, K. 2020. impact of lifestyle on cytochrome P450 monooxygenase repertoire is clearly evident in the bacterial phylum Firmicutes. *Scientific reports,* 10**,** 1-12.

PARVEZ, M., QHANYA, L. B., MTHAKATHI, N. T., KGOSIEMANG, I. K., BAMAL, H. D., PAGADALA, N. S., XIE, T., YANG, H., CHEN, H., THERON, C. W., MONYAKI, R., RASELEMANE, S. C., SALEWE, V., MONGALE, B. L., MATOWANE, R. G., ABDALLA, S. M., BOOI, W. I., VAN WYK, M., OLIVIER, D., BOUCHER, C. E., NELSON, D. R., TUSZYNSKI, J. A., BLACKBURN, J. M., YU, J. H., MASHELE, S. S., CHEN, W. & SYED, K. 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Sci Rep,* 6**,** 33099.

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.

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.

RAVCHEEV, D. A. & THIELE, I. 2017. Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of

host-synthetized mucin glycans and utilization of mucin-derived monosaccharides. *Frontiers in genetics,* 8**,** 111.

RINNINELLA, E., RAOUL, P., CINTONI, M., FRANCESCHI, F., MIGGIANO, G. A. D., GASBARRINI, A. & MELE, M. C. 2019. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms,* 7**,** 14.

RIOS-COVIAN, D., SÁNCHEZ, B., SALAZAR, N., MARTÍNEZ, N., REDRUELLO, B., GUEIMONDE, M. & DE LOS REYES-GAVILÁN, C. G. 2015. Different metabolic features of Bacteroides fragilis growing in the presence of glucose and exopolysaccharides of bifidobacteria. *Frontiers in Microbiology,* 6.

ROJAS, J., AMBIKA MANIRAJAN, B., RATERING, S., SUAREZ, C. & SCHNELL, S. 2020. Draft Genome Sequences of Spirosoma agri KCTC 52727 and Spirosoma terrae KCTC 52035. *Microbiology Resource Announcements,* 9**,** e00317-20.

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.

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.

SCHÖNER, T. A., FUCHS, S. W., SCHÖNAU, C. & BODE, H. B. 2014. Initiation of the flexirubin biosynthesis in Chitinophaga pinensis. *Microbial Biotechnology,* 7**,** 232-241.

SELLO, M. M., JAFTA, N., NELSON, D. R., CHEN, W., YU, J. H., PARVEZ, M., KGOSIEMANG, I. K., 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*. *Sci Rep,* 5**,** 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.

SHANSON, D. C. 2014. *Microbiology in clinical practice*, Butterworth-Heinemann.

SHARMA, D., PATEL, A., SONI, P., SHARMA, P. & GUPTA, B. 2016. Empedobacter brevis meningitis in a neonate: a very rare case of neonatal meningitis and literature review. *Case Reports in Pediatrics,* 2016.

SHARON, G., GARG, N., DEBELIUS, J., KNIGHT, R., DORRESTEIN, P. C. & MAZMANIAN, S. K. 2014. Specialized metabolites from the microbiome in health and disease. *Cell metabolism,* 20**,** 719-730.

STOJANOV, S., BERLEC, A. & ŠTRUKELJ, B. 2020. The influence of probiotics on the firmicutes/bacteroidetes ratio in the treatment of obesity and inflammatory bowel disease. *Microorganisms,* 8**,** 1715.

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., SHALE, K., PAGADALA, N. S. & TUSZYNSKI, J. 2014. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS One,* 9**,** e86683.

THOMAS, F., BARBEYRON, T. & MICHEL, G. 2011. Evaluation of reference genes for real-time quantitative PCR in the marine flavobacterium Zobellia galactanivorans. *Journal of microbiological methods,* 84**,** 61-66.

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.

WALSH, C. J., GUINANE, C. M., HILL, C., ROSS, R. P., O'TOOLE, P. W. & COTTER, P. D. 2015. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. *BMC microbiology,* 15**,** 1-11.

WERCK-REICHHART, D. & FEYEREISEN, R. 2000. Cytochromes P450: a success story. *Genome Biology,* 1**,** reviews3003.1.

XU, M., HILLWIG, M. L., LANE, A. L., TIERNAN, M. S., MOORE, B. S. & PETERS, R. J. 2014. Characterization of an orphan diterpenoid biosynthetic operon from Salinispora arenicola. *J Nat Prod*, 77**,** 2144-7.

ZHANG, X., PENG, Y., ZHAO, J., LI, Q., YU, X., ACEVEDO-ROCHA, C. G. & LI, A. 2020. Bacterial cytochrome P450-catalyzed regio-and stereoselective steroid hydroxylation enabled by directed evolution and rational design. *Bioresources and Bioprocessing*, 7**,** 1-18.

ZHOU, K., BAO, Y. & ZHAO, G. 2019. Effects of dietary crude protein and tannic acid on rumen fermentation, rumen microbiota and nutrient digestion in beef cattle. *Archives of animal nutrition*, 73**,** 30-43.

ZHOU, Z., ZHANG, W., SU, S., CHEN, M., LU, W., LIN, M., MOLNÁR, I. & XU, Y. 2015. CYP287A1 is a carotenoid 2-β-hydroxylase required for deinoxanthin biosynthesis in Deinococcus radiodurans R1. *Appl Microbiol Biotechnol*, 99**,** 10539-46.

ANNEXURE

The following supporting information can be downloaded at:

https://www.mdpi.com/article/10.3390/ijms23095057/s1.

Table S1. Information on *Bacteroidetes* species and their respective genera was used in the study. Species abbreviations, their genome IDs (GenBank), the presence or absence of P450s in different species, and the number of P450s are presented in the table.

Table S2. Comparative analysis of ferredoxins in *Bacteroidetes* species.

Table S3. Subtype-level comparative analysis of ferredoxins between *Bacteroidetes*- and *Firmicutes*-species. Ferredoxins were classified into different subtypes following the procedure described elsewhere

Table S4. Information on ferredoxins used as reference proteins for datamining ferredoxins in *Bacteroidetes* species.

Supplementary Dataset S1: P450 sequences identified in *Bacteroidetes* species are presented along with their annotated name, followed by protein ID (in parenthesis) and species name.

Supplementary Dataset S2: Ferredoxin sequences identified and annotated in *Bacteroidetes* species are presented according to their cluster type and subtype.