ISSN: 0974-8571 Vol.14 No. 1 June , 2022

International Journal of Computational Intelligence in Control

Intergenic Distances Analysis Reveals the Distribution Pattern of Cytochromes P450 Class I Redox System Genes in Kingdom Bacteria

Muhammad Farukh[1, 2,*], Adam Khan[3], Javed Iqbal[4], NasiruddunShaikh[4], Fatima tul Zahra[2]

[1]Department of Biotechnology, School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China.

[2]Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan

[3]Department of Botany, University of Lakki Marwat, KP, Pakistan

[4]Department of Botany, Government College University, Hyderabad, Pakistan

*Corresponding Author: Muhammad Farukh, E-mail: dr.muhammadfarukh@gmail.com

Date of Submission: 21st October 2021 Revised: 21st December 2021 Accepted: 09th February 2022

How to Cite: **Muhammad Farukh et.al. , 2022. Intergenic Distances Analysis Reveals the Distribution Pattern of Cytochromes P450 Class I Redox System Genes in Kingdom Bacteria. International Journal of Computational Intelligence in Control, 14(1).**

ABSTRACT: **Cytochromes P450 (P450s) catalyze the diverse range of oxidation reactions, during which they receive electrons from various redox partners. In bacteria, the P450 redox protein system is very diverse, but most of them use ferredoxin and ferredoxin reductase as electron transfer proteins. Due to the linear nature of bacterial genomes, short intergenic distances between P450s and putative redox proteins hint possible redox partnership. Therefore, it is of importance to look for their potential redox partner in the native genomes. The complete bacterial genomes were used to analyze the distribution of P450 class I redox, ferredoxin, and ferredoxin reductase genes in bacteria. The product features of coding regions were searched against the gene names, and the new homologs were predicted. The distances between the identified genes were calculated as the difference between the coordinates of corresponding loci. Moreover, the bacterial genomes were used to identify the gene clusters of P450 class I redox system genes. A total of 4,849 genomes were found containing P450, 12,613 genomes contain ferredoxin and 10,759 genomes contain ferredoxin reductase genes. In about 32% of P450-containing genomes, P450 genes and ferredoxin genes are located very close to each other, while the ferredoxin reductase genes are located at some distance apart in most cases. Gene cluster analysis shows that P450 genes are found in clusters with corresponding ferredoxin and ferredoxin reductase genes, mostly in phylum of *Actinobacteria*. Overall, our study provides insights into the distribution pattern of P450 genes along with corresponding ferredoxin and ferredoxin reductase genes in bacteria.**

Keywords: Bioinformatics, Cytochromes P450, Redox system genes, Class I, Ferredoxin reductase, Intergenic distance, Gene cluster

INTRODUCTION

Cytochromes P450 (P450s) are the largest and most widely studied superfamily of heme-thiolate-containing proteins [1]. Typically, these proteins are monooxygenases. The terminology P450 derives from the unusual spectral properties of heme- containing red pigments, which display a typical absorption band at 450 nm because of their reduced CO-bound complex [2]. P450s are very diverse and found throughout all five biological kingdoms of life, and they are involved in a vast range of the oxidation reactions which require in electron transfer chains [3-6]. The reactions carried out by cytochromes P450 are extremely diverse, which includes the bioconversion of xenobiotics, the bioactivation of chemical carcinogens, the biotransformation of drugs, the biosynthesis of physiologically important compounds such as fatty acids, steroids, fat-soluble vitamins and bile acids, eicosanoids, the conversion of alkanes, aromatics and terpenes compounds as well as the degradation of insecticides and herbicides [7-15]. The typical P450 reaction is mono-oxygenation in which one of the oxygen atoms of molecular oxygen is inserted into an organic substrate, while the second oxygen atom undergoes reduction to water. However, there are other P450-catalyzed reactions, including heteroatom oxidation and epoxidation [16]. Most P450s use NAD(P)H-driven redox proteins system for its catalytic reactions.

In eukaryotes, cytochromes P450 are bound to membranes with an N-terminal transmembrane helix, mostly are attached to the endoplasmic reticulum, and a subset is linked to the inner membrane of adrenal gland mitochondria. Generally, they contain either adrenodoxin-adrenodoxin reductase (mitochondrial-type) redox proteins system, where P450s obtain electron from NADPH via a FAD- containing, NADPH-dependent adrenodoxin reductase and an adrenodoxin [17, 18], or cytochrome P450 reductase (CPR) (microsomal-type) protein system, where CPR required to transfer the electron from NADPH to P450s [17, 19], which contains the prosthetic group's FAD and FMN. In prokaryotes, cytochrome P450s are the soluble proteins that lack the N-terminal membrane anchor. And most of the prokaryotes contain the ferredoxin-ferredoxin reductase (bacterial-type) redox system genes, where P450s receive an electron from NADH via an NADH-dependent FAD-containing ferredoxin reductase (FdR) and a mitochondrial-type ferredoxin (Fdx), which both are also soluble proteins [9]. While, there is much more in the redox protein apparatus that drives catalysis in prokaryotic P450 enzymes [20].

The functionally associated group of genes, found within the DNA of an organism, are termed as gene cluster and often located within a few thousands of base pairs. Intergenic distance, the distance between two genes on the DNA strand, tends to be shorter if they belong to the same gene cluster/operon. Intergenic distance is considered to be the most basic informative feature for gene cluster/operon prediction. In bacteria, the distribution of P450s is very diverse, with many bacteria have numerous P450s, and some are having no P450s, such as *Escherichia* coli [21, 22]. The P450s catalytic activity depends on their associated individual ferredoxin and ferredoxin reductase. It was reported that in many bacteria, P450 genes are arranged in operons or gene clusters with their redox protein genes, ferredoxin and ferredoxin reductase [23-31]. The *in vitro* of P450 is usually performed with spinach ferredoxin and spinach ferredoxin reductase. The activity of P450 in such combination is not as good as the P450 with its native redox partner [32]. Also, many P450s cannot be characterized due to the lack of suitable electron transfer protein and corresponding reductases. Therefore, it is of importance to look for their potential redox partner in the native genomes, which may be hinted by the intergenic distance between them.

So far, the distribution patterns of P450 genes with their potential redox partner genes have not been studied at the whole bacterial kingdom scale. In this study, to gain the overall picture of intergenic distances between P450 and their potential redox partner, we analyzed the complete genomes data of bacteria, a total of 13,565 bacterial genomes in NCBI,

to illustrate the distribution pattern of P450 class I redox system genes (P450, ferredoxin and ferredoxin reductase). Our results show that mostly P450 genes and ferredoxin genes are located close to each other, while many ferredoxin reductase genes are not located close to P450 genes. Moreover, the P450, ferredoxin, and ferredoxin reductase genes are found in gene clusters, mostly in phylum *Actinobacteria*. The analysis of the spatial relationship of bacterial P450 class I redox system genes may shed light on the distribution and interaction of these proteins.

MATERIALS AND METHODS

The sources of genomic sequencing data

The whole-genome assembled data of all available complete bacterial genomes were downloaded from the FTP (File Transfer Protocol) website ftp://ftp.ncbi.nlm.nih.gov of NCBI (National Center for Biotechnology Information) till June 11th, 2019. This dataset contains 13,565 complete bacterial genomes of all available bacterial strains in GenBank format.

Genome mining of cytochrome P450 and its class I redox system genes in bacteria

The GenBank files of all complete bacterial genomes were searched for the words; P450, ferredoxin, and ferredoxin reductase genes, against the 'product' feature of CDS (coding regions), with a custom python script, based on the gene annotation associated with the sequencing data. For P450, all the CDS were searched and selected if 'P450' but no 'reductase' word is present in their 'product' feature. For ferredoxin, we searched the 'product' feature having 'ferredoxin' word without 'reductase' word. Moreover, for ferredoxin reductase, we searched the gene 'product' feature having 'ferredoxin' and 'reductase' words. The coordinates of these genes on the genomes were also collected with the genes.

P450 gene prediction in genomes of bacteria

To find the P450 genes in more genomes, we used all complete bacterial genomes for the P450 gene prediction. We collected all the CDS of each genome to search for the P450 protein homologs. For the P450 reference sequences, all the prokaryotes identical protein groups (IPG) sequences of P450 proteins were downloaded from the NCBI website (https://www.ncbi.nlm.nih.gov/ipg) in FASTA format till July 19th, 2019 and this dataset contain 128,255 P450 IPG sequences. First, these IPG sequences of P450s were used as a database for the DIAMOND software. The DIAMOND BLASTP (version 0.9.24.125) [33] was used for P450 protein sequence search with parameters; minimum identity 40%, minimum 60% of subject coverage, maximum target sequences set as one, and with the sensitive parameters. The CDS of each bacterial genome were used as queries for the DIAMOND BLASTP search against the database of IPG sequences of P450s. Those CDS, which DIAMOND found as matches in the P450 IPG database, were selected for further confirmation. The HMMER software (Version 3.2.1) (http://hmmer.org/) was used for additional verification of hit CDS as P450 homologs. The same P450 IPG sequences, which were used as the database for DIAMOND, were used to build the Hidden Markov Model (HMM) profiles database using the HMMBUILD program of HMMER package. The hit CDS, which we extracted from DIAMOND search results, were used as queries. HMMSCAN program was used to search the P450 homolog model with the default parameters. The CDS, which we found matches in the P450 homolog model, were collected for downstream analysis.

Identification of P450 rich genera in bacteria

To find the P450 rich genera in bacteria, the genomes having all three-set of P450 class I redox system genes were used to normalize the occurrence of P450 genes against the number of coding genes of the genome. The values of the total number of coding genes of each genome were collected from GenBank files of the genomes. For the normalization, we have used the P450 gene counts per thousand coding genes in a genome formula, given below:

$$\frac{Number\ of\ identified\ P450\ genes\ in\ a\ genome}{Total\ number\ of\ coding\ genes\ in\ a\ genomes} * 1000$$

Correlation Coefficient analysis between cytochrome P450 class I redox system genes

The bacterial genomes, which contains all three of P450 class I redox system genes were used for the analysis of correlation coefficient between the numbers of the genes. The genera which had more than ten genomes were selected for the analysis. The correlation coefficient was analyzed between the number of P450 and ferredoxin genes, P450 and ferredoxin reductase genes, and ferredoxin and ferredoxin reductase genes.

Intergenic distances calculation between the genes of P450 class I redox system

The genomes which have three-member P450 class I redox system genes were used for the intergenic distances calculation analysis. The distances between the P450 class I redox system genes (P450/ferredoxin, P450/ferredoxin reductase and ferredoxin/ferredoxin reductase) were calculated as the difference between the coordinates of the corresponding loci, using a custom python script. The genes loci coordinates were collected from the GenBank files of the genomes, alongside the genes CDS feature. There were maybe multiple combinations of P450, ferredoxin, and ferredoxin reductase, only the distances between the closest pairs were selected.

Identification of gene clusters containing P450 class I redox system genes

The dataset of genomes which had the complete set of P450 class I redox system genes (P450, ferredoxin and ferredoxin reductase) were used to identify the biosynthetic gene clusters. For gene cluster identification, we used antiSMASH 4.2.0 software [34] with its border predict and inclusive parameters. A custom python script program was used to inspect the genes in different type of gene clusters.

RESULTS

Identification and general features of cytochromesP450 class I redox system genes in bacteria

A dataset was constructed with all available complete bacterial genomes from the NCBI website, which consists of 13,565 bacterial genomes in total. The P450 class I redox system genes (P450, ferredoxin and ferredoxin reductase) were identified according to the annotation of the genomic data. From these NCBI annotated genomes, we found P450 genes only in ~32.7% genomes of bacteria, which are relatively less number of genomes having P450 genes. To find the P450 genes homologs in more genomes, we used all bacterial genomes for the P450 gene prediction.

In P450 gene prediction analysis of all complete bacterial genomes, we found 4,849 bacterial genomes (35.7% of the total genomes) have cytochrome P450 genes. Among these genomes, 59.3% belong to *Proteobacteria*, 17.6% of them belong to *Firmicutes*, and 17.2% classified as *Actinobacteria*. Among these P450-containing genomes, around 43% genomes have only one P450 gene, and the rest have more than one P450s, ranging from one to 81 genes in a genome. The genome of strain *Nonomuraea sp. ATCC 55076* (Accession#; NZ_CP017717.1) has 81 P450 genes while strain *Mycobacterium sp. EPa45* (Accession#; NZ_CP011773.1) has 74 P450 genes. From the whole bacterial genome dataset, we identified 12,613 (93% of the total) bacterial genomes having ferredoxin genes, among which almost 59.9% are classified as *Proteobacteria*, 21.7% are classified as *Firmicutes*, and only 10.6% belong to *Actinobacteria*. Among these ferredoxin-containing genomes, some have one, and some have multiple ferredoxin genes, ranging from 1 to 41 genes in one genome. The genome of *Moorella thermoacetica ATCC 39073* strain (Accession# NC_007644.1) has 41 ferredoxin genes. Among these ferredoxin containing genomes, 4,447 genomes have P450 genes. A total of 10,759 (79.3% of the total) bacterial genomes, from whole bacterial genome dataset, were found having ferredoxin reductase genes, from which 65.4% belong to *Proteobacteria*, 18.9% belong to *Firmicutes*, and 9.2% belong to *Actinobacteria*. These ferredoxin reductase containing genomes were found to have one to 25 ferredoxin reductase genes each. The genomes of *Desulfobacula toluolica Tol2* (Accession#; NC_018645.1) has 25 ferredoxin reductase genes. The overall distribution of P450 class I redox system genes in bacteria is demonstrated in Figure 1.

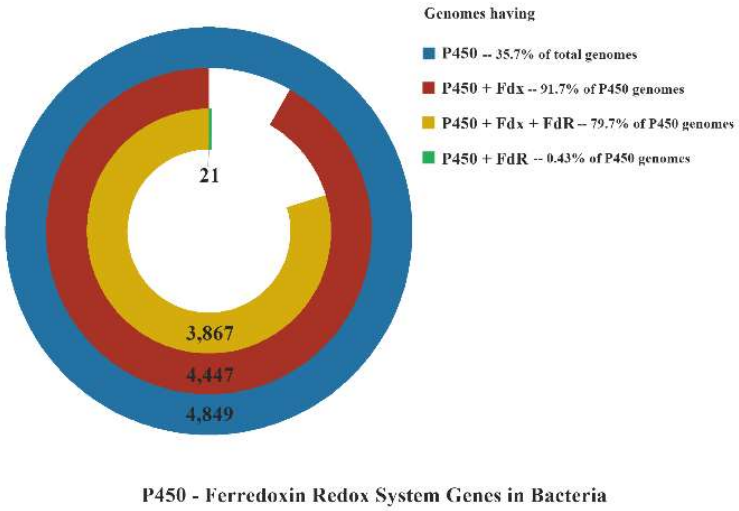

Figure 1: Distribution of cytochromes P450 class I redox system genes in bacteria

It shows the overall picture of the distribution of cytochromes P450 class I redox system genes in bacteria. Color schemes were used to represent the proportion of bacterial genomes which contain P450 class I redox system in the complete dataset. Here Fdx, and FdR represent ferredoxin and ferredoxin reductase respectively.

Interestingly, not all of the bacteria with P450 genes contain a complete redox proteins system. After examining the whole dataset, we found 3,867 bacterial genomes (around 79.7% of the genomes containing P450genes) have the three-member P450 class I redox system genes: P450, ferredoxin and ferredoxin reductase (Figure 1). Among these genomes with the complete set of P450 class I redox system genes, 67.6% of them belong to phylum *Proteobacteria*, 15.5% are classified as *Actinobacteria,* and 11.1% belong to *Firmicutes* (Figure 2). It suggests that most bacteria containing the P450/ferredoxin system belong to *Proteobacteria*.

P450 rich genera of bacteria and correlation coefficient analysis between P450 class I redox system genes

The genomes of genera *Mycobacterium* and *Mycolicibacterium* have the highest number of P450 genes which are 12.9 and 12.1 P450 genes per thousand coding genes in a genome, respectively. While among the genomes of *Frankia*, *Streptomyces*, and *Mycobacteroides*, the highest number of P450 genes are *Mycobacteroides*, the highest number of P450 genes are 9.9, 8.9, and 8.8 P450 genes per thousand coding genes in a genome, respectively (Figure 3). Among the genomes of genus Mycobacterium, 57 genomes have more than eight P450 genes per thousand coding genes, which mostly belong to species *Mycobacterium avium*, *Mycobacterium chimaera*, and *Mycobacterium intracellulare*. Among them, 25 genomes have more than 10 P450 genes. Among the genomes of genus *Mycolicibacterium*, eight genomes have more than eight P450 genes per thousand coding genes, among them, only three genomes have more than 10 P450 genes, which belong to species *Mycolicibacterium chitae*, *Mycolicibacterium hassiacum,* and *Mycolicibacterium flavescens*. From these *Mycolicibacterium* genomes, the most number of genomes (six genomes) belong to *Mycolicibacterium smegmatis*. Among the genomes of genus Streptomyces, only three genomes have more than eight P450 genes per thousand coding genes, two of them belong to *Streptomyces albulus* and one belongs to *Streptomyces clavuligerus*. Only one genome of genus *Mycobacteroides* and one genome of genus *Nocardia* have more the eight P450 genes per thousand coding genes.

All the P450-rich genera, shown in Figure 3, belong to

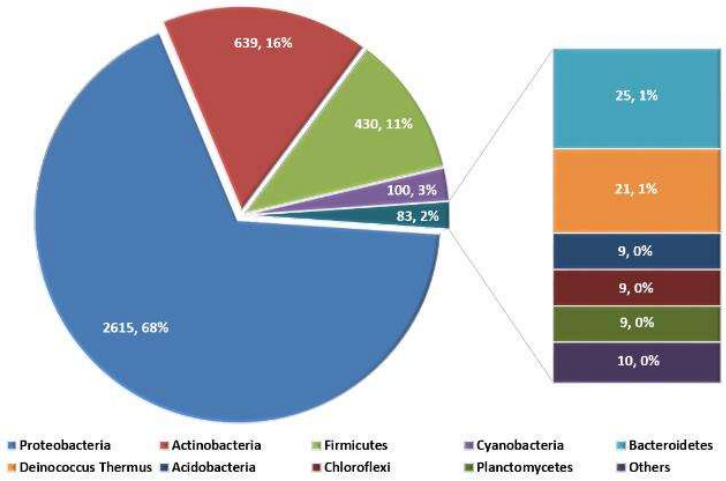

Figure 2: Bacterial genomes having P450 class I redox system genes from different phyla

Bacterial phyla containing P450, ferredoxin and ferredoxin reductase genes. Colors schemes were used to represent different phyla. Blue color represents the number of bacterial genomes that belongs to *Proteobacteria*. Red color shows the number of genomes belongs to *Actinobacteria*. Green color shows the number of genomes belongs to *Firmicutes*. Light Purple color represents the number of genomes belongs to phylum Cyanobacteria. Sky blue shows the number of genomes belongs to *Bacteroidetes*. Orange, dark blue, maroon, dark green and dark purple colors represent the number of genomes belongs to phyla *Deinococus-Thermus*, *Acidobacteria*, *Chloroflexi*, *Planctomycetes* and others, respectively. Phyla *Proteobacteria* and *Actinobacteria* were found the most abundant phylum.

the *Actinomycetes* (order *Actinomycetales*). The function of these bacteria is very well known. They play an essential role in the production of secondary metabolites. They are known to produce a wide variety of industrially and therapeutically significant compounds, i.e., antibiotics, chemotherapeutics, fungicides, herbicides, and immunosuppressants. Species of genus *Mycobacterium* comprise pathogens which are well known to cause severe diseases [35]. Usually, the mycobacteria are free-living saprophytes [36] and are found in a diverse type of habitats, such as aquatic [37] and soil environments [38]. Mycobacteria are well known causative agents of a broad spectrum of diseases in humans and animals. The species of genus *Mycobacterium* comprise a vast number of P450 genes, and many of them are involved in the biosynthesis of secondary metabolites and many other pathways [39-43]. The genus *Mycolicibacterium* comprises all of the major human pathogens [44, 45]. Like *Mycobacterium*, the species of genus *Mycolicibacterium* hold many P450 genes, and many P450 genes of *Mycolicibacterium* species are conserved in sequence and function with the many species of *Mycobacterium*, and many of them conserved with P450s of *Mycobacterium tuberculosis* [46-49]. *Streptomyces* are the most abundant natural source of various secondary metabolites and especially antibiotics [50]. The *Streptomyces* species contain a large number of P450 genes, and a significant number of them are associated with gene clusters for secondary metabolites, and many of them clustered with genes for ferredoxins. The P450s are involved in the biosynthesis of many Macrolide Antibiotic Agents [51-55].

The correlation coefficient analysis shows that the three-member genes of P450 class I redox system display the strong correlation between them in the genomes of genus *Mycobacterium*, where the correlation coefficient value between P450 and ferredoxin genes is 0.962, between P450 and ferredoxin reductase genes is 0.718 and between ferredoxin and ferredoxin reductase genes is 0.723 (Figure 4). Among the genomes of genus *Nostoc*, the correlation coefficient value between P450 and ferredoxin genes, is 0.709, between P450 and ferredoxin reductase genes, is 0.729 and between ferredoxin and ferredoxin reductase genes, is 0.655. Among the genomes of genus

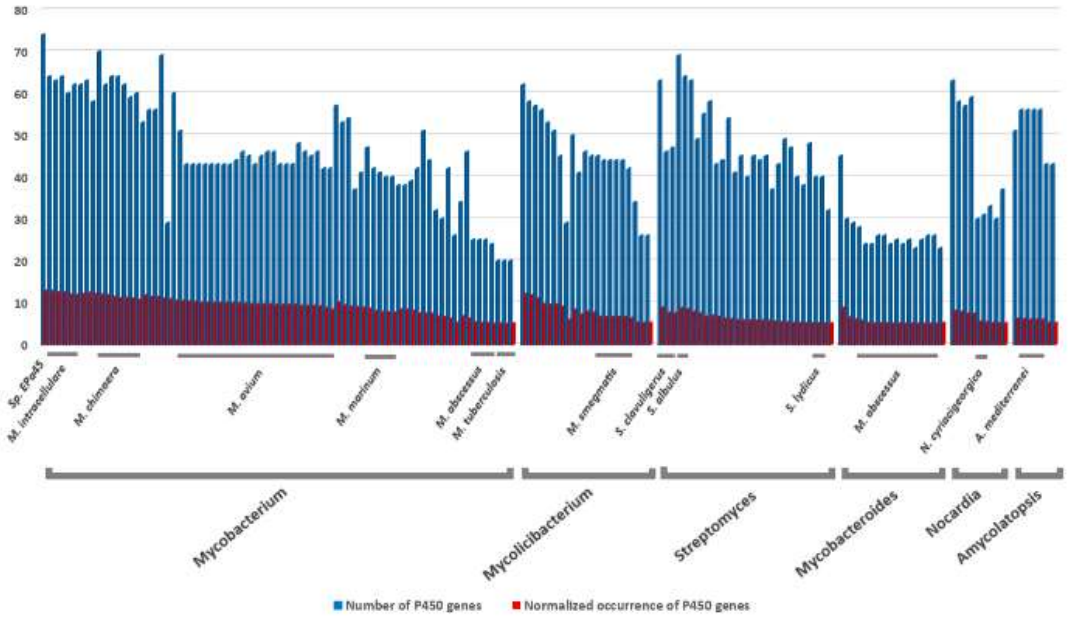

Figure 3: Cytochromes P450 rich genomes of different genera of bacteria
The bacterial genomes of different phyla which contain the highest number of cytochrome P450 genes in their genomes are shown. Here x-axis shows the genomes which belong to different phyla while the y-axis shows the total number of P450 genes in each genome and the number of P450 genes per thousand coding genes in each genome. Bars in blue color represent the total number of P450 genes in each genome, while the red color bar represents the number of P450 genes per thousand coding genes in each genome.

Sphingomonas, the correlation coefficient value between P450 and ferredoxin genes, is 0.773, between P450 and ferredoxin reductase genes, is 0.831 and between ferredoxin and ferredoxin reductase genes is 0.613. While P450 and ferredoxin genes show a good correlation in many other genera. And some genera show the negative correlation between these genes.

Intergenic distance analysis of cytochromesP450 class I redox system genes

To determine the distribution of P450 class I redox system genes, we analyzed the intergenic distance of these genes in complete bacterial genomes. For P450 class I redox system genes, we collected these genes locations on the genome as indicated by the genomic coordinates. Those gene sequence locations were used to calculate the intergenic distance between one gene to another gene in a genome. The intergenic distances between the P450 class I redox system genes are determined by calculating the difference between their genomic locations (Table 1).

Table 1: Ranges of intergenic distances between the closest cytochromes P450 class I redox system genes

Distance between genes	No. of bacterial genomes	Range of intergenic distances between closest genes (bp)
P450 class I redox system genes		
P450 and Ferredoxin	4476	-58 − 4,939,319
P450 and Ferredoxin Reductase	3913	-187 − 5,802,590
Ferredoxin and Ferredoxin Reductase	3892	-88 − 4,096,222

Note: Minus sign with numbers shows the overlapping region between two genes.

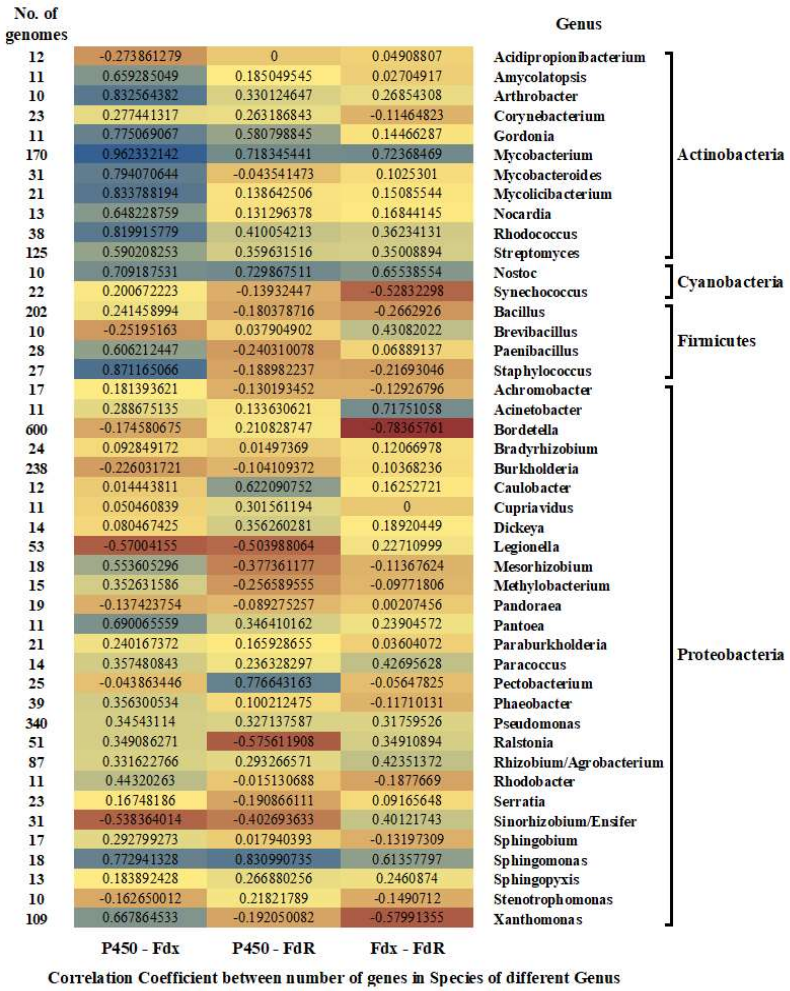

Figure 4: Correlation coefficient analysis between P450 class I redox system genes in genomes of different genera

It demonstrates the correlation coefficient analysis between P450 and ferredoxin, P450 and ferredoxin reductase genes, and ferredoxin and ferredoxin reductase genes, among the genomes of different bacteria genera, which contain three-member genes of P450 class I redox system genes. Here, Fdx represents the ferredoxin, while FdR represents the ferredoxin reductase.

The minus signs are the indication of the overlapping sequences. In bacteria, approximately one-third of genes are found in overlapping arrangement only by a few base pairs, and function in the regulation of gene expression [56].I n our results, we found mostly one nucleotide frameshift overlaps as compare to two nucleotide frameshift overlaps. P450/ferredoxin and ferredoxin/ ferredoxin reductase genes overlap more abundantly, while P450/ferredoxin reductase genes overlaps are found in very few numbers of genomes.

The density plot of distances shows that, in 1,552 out of 3,867 bacterial genomes, the distances between P450 and the closest ferredoxin genes are less than 100 kb, among them, the distances in 713 genomes are less than 3 kb (Figure 5A), most of which belong to phylum *Proteobacteria*. It shows that the ferredoxin genes are very close to P450 genes in many genomes, suggesting that their gene products may interact with each other. In contrast, the density plot of distances between P450 and ferredoxin reductase genes shows two dense regions. The distances between P450 and ferredoxin reductase in 1,425 genomes are less than100 kb, among them, the distance is less than 10 kb only in 363

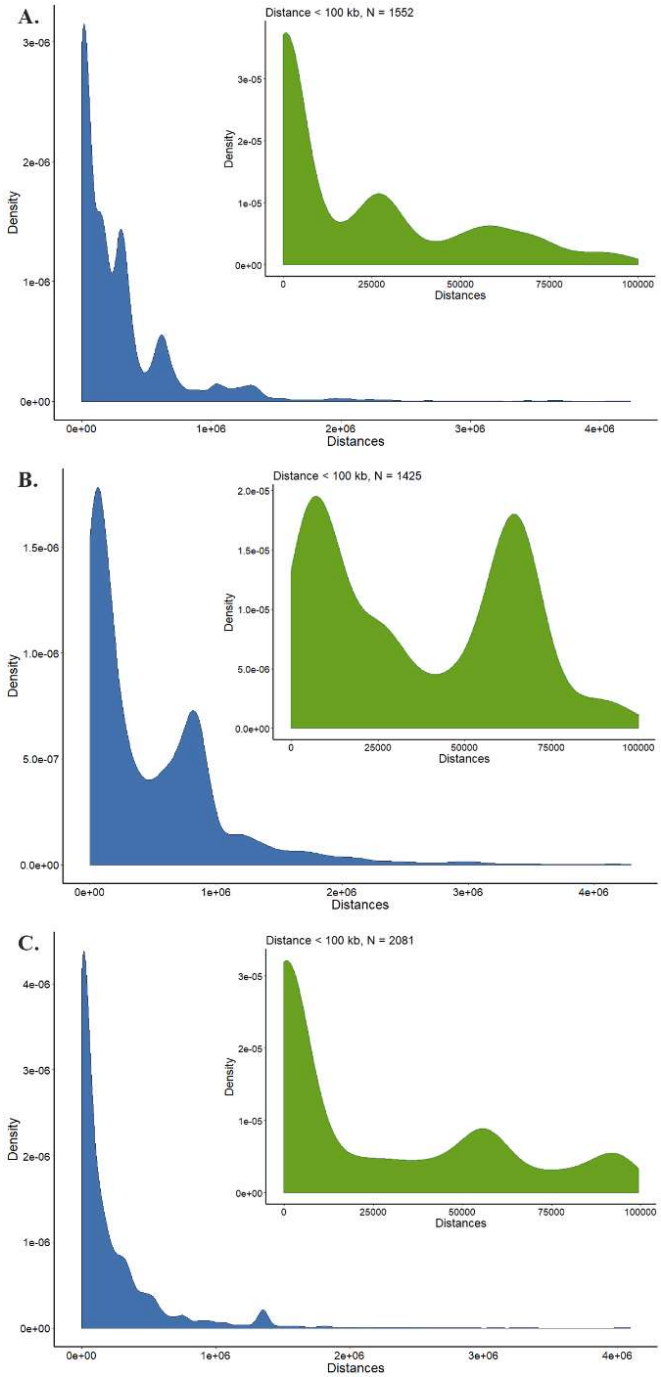

Figure 5: Density plots of intergenic distances between the closest genes of P450, ferredoxin and ferredoxin reductases

Density plots show the density of closest intergenic distances between P450, ferredoxin and ferredoxin reductase genes in 3,867 bacterial genomes, which contain all these three genes. **(A)** Intergenic distances plot of P450 and ferredoxin genes. In 713 genomes, the distances between these genes are less than 3 kb. **(B)** Intergenic distances plot of P450 and ferredoxin reductase genes showing two regions of high density. In 363 genomes, the distances between these genes are less than 10 kb. **(C)** Intergenic distances plot of ferredoxin and ferredoxin reductase genes. In 1,072 genomes, the distances between these genes are less than 10 kb. The values of the x-axis are the distances in base pairs.

genomes (Figure 5B). The density plot shows that the distances between the closest ferredoxin and ferredoxin reductase genes in 2,081 genomes are less than100kb, among them, the distance in 1,072 genomes are less than 10 kb (Figure 5C). In many genomes, ferredoxin reductase genes are not located very close to P450 genes. Since expression of ferredoxin reductase in some genomes may not be regulated along with the expression of the P450 or ferredoxin genes, it does not need to be located nearby. On the other hand, the ferredoxin reductase genes located close to the P450 or ferredoxin genes, as indicated by the peak centered at the zero, suggestingthat these genes are probably enclosed in the same operon or cluster. We analyzed the genomes in which the distances between P450, ferredoxin, and ferredoxin reductase genes are less than 10kb and found that mostly the P450 genes are located very close to ferredoxin and ferredoxin reductase genes in genomes of phylum *Actinobacteria*. While ferredoxin and ferredoxin reductase genes are mostly located close to each other in genomes of phylum *Proteobacteria*.

Moreover, to further investigate the distance between the genes, we selected the genomes of *Mycobacterium tuberculosis* bacterium strains from the genomes, which contain all three genes of P450 class I redox system. We analyzed the distance between the closest genes of P450 class I system in the genomes of the bacterium *Mycobacterium tuberculosis* (Table 2). We found that the distances between the closest P450 and ferredoxin genes in 81 genomes are thee base-pairs, in eight genomes is two bp while the distance in only one genome is 199bp. The distances between the closest P450 and ferredoxin reductase genes in71 genomes are in range 8,144 to 11,294 bp.

Gene cluster prediction analysis of cytochromes P450 class I redox system genes

We analyzed in the previous section that P450 genes are not located as close to all ferredoxin reductase genes as close to ferredoxin genes in most of the genomes. As it the well-known that P450s play an essential role in secondary metabolism pathways [57]. To further investigate the distribution of P450 class I redox system genes, we performed the gene cluster prediction analysis. The sequencing data of 3,867 genomes, which contains three-set P450 class I redox system genes, were analyzed for the gene clusters of P450 class I redox system genes. We examined the four types of gene clusters, Type I; the clusters containing P450, ferredoxin, and ferredoxin reductase genes, Type II; the clusters containing P450 and ferredoxin genes, Type III; the clusters containing P450 and ferredoxin reductase genes, and Type IV; the clusters containing ferredoxin and ferredoxin reductase genes (Table 3).

The Type I gene clusters were detected in 82 bacterial genomes. Among these, 70 genomes are classified as *Actinobacteria,* and only ten are classified as *Proteobacteria*. In all these genomes, the majority have only one P450 class I redox system gene cluster, and the genomes of *Streptomyces albulus* and *Amycolatopsis mediterranei* have three gene clusters with full P450 class I redox system genes (Figure 6). This type of gene clusters mostly found (22 genomes) in genus *Streptomyces*.

The Type II gene clusters were identified in 610 bacterial genomes, and 487 of them belong to phylum

Table 2: Ranges of intergenic distances between the closest cytochrome P450 class I redox system genes in *Mycobacterium tuberculosis*

Distance between genes	No. of bacterial genomes	Range of intergenic distances between closest genes (bp)
P450 class I redox system genes		
P450 and Ferredoxin	90	2 − 199
P450 and Ferredoxin Reductase	90	8144 – 109,757
Ferredoxin and Ferredoxin Reductase	90	67,312 − 427,385

Table 3: Cytochromes P450 class I redox system genes identified in different types of gene clusters

Gene cluster type	Genes in cluster	No. of genomes containing the clusters	Total no. of identified clusters
P450 class I genes clusters in 4,267 genomes dataset			
1	P450 + Ferredoxin + Ferredoxin Reductase	82	97
2	P450 + Ferredoxin	610	1570
3	P450 + Ferredoxin Reductase	68	74
4	Ferredoxin + Ferredoxin Reductase	105	107

Actinobacteria, and 110 belong to *Proteobacteria*. These identified genomes contain total 1,570 gene

The Type II gene clusters were identified in 610 bacterial genomes, and 487 of them belong to phylum *Actinobacteria,* and 110 belong to *Proteobacteria*. These identified genomes contain total 1,570 gene clusters. This type of gene cluster mostly found (163 genomes) in genus *Mycobacterium*.

The Type III gene clusters were detected in 68 bacterial genomes. Among them, 55 are classified as *Actinobacteria,* and only 13 are classified as

Proteobacteria. These genomes contain a total of 74 gene clusters and this type of gene clusters mostly found (13 genomes) in genus *Mycobacterium*.

The Type IV gene clusters were detected in 105 bacterial genomes. Among these genomes, 71 belong to phylum *Proteobacteria,* and 22 belong to phylum *Actinobacteria*. This type of gene cluster mostly found (17 genomes) in genus *Sinorhizobium*.

DISCUSSION

Cytochromes P450 (P450s) superfamily is a large and diverse group of enzymes which catalyze the diverse range of reactions with the help of their redox partner proteins. P450 redox protein system is very diverse in bacteria, but P450 mostly use the ferredoxin redox partner proteins system in bacteria. Distribution of cytochrome P450 genes and corresponding genes encoding ferredoxin redox system proteins and their gene interaction were studied in 13,565 complete bacterial genomes. Results showed that around 35.7% bacterial genomes comprise the P450 genes, while the ferredoxin genes were found in ~93% genomes and ~79% genomes contain ferredoxin reductase genes. These results suggest that many bacteria do not use cytochrome P450 systems. It is not unexpected as the *E. coli* does not have P450 systems [21, 22].

The results revealed that approximately 28.5% of the complete bacterial genomes contain the complete set of P450 class I redox system genes, which mostly belong to phylum *Proteobacteria*. Some genomes do not contain any ferredoxin redox system genes, they may use some other P450 systems like flavodoxin redox system, fusion protein P450s and/or self-sufficient P450s [58-60].

The bacterial genomes usually lack long-distance gene interaction, it is possible the P450 prefer pairing with the closest ferredoxin and ferredoxin reductase to perform their catalytic activities. In prokaryotes, many studies suggested that P450 genes reside close on the chromosome with the genes encoding electron transfer proteins [23, 25, 61-63], and mostly these electron transport proteins are ferredoxin and ferredoxin reductase. Many studies confirmed experimentally that ferredoxin and ferredoxin reductase proteins are work together as electron transfer proteins with P450 proteins [32, 64-68]. Therefore, we analyzed the bacterial genomes to determine the intergenic distance between P450, ferredoxin, and ferredoxin reductase genes in order to understand how products of these genes may interact with each other. Distance analysis results demonstrated that mostly P450 genes locate close to the

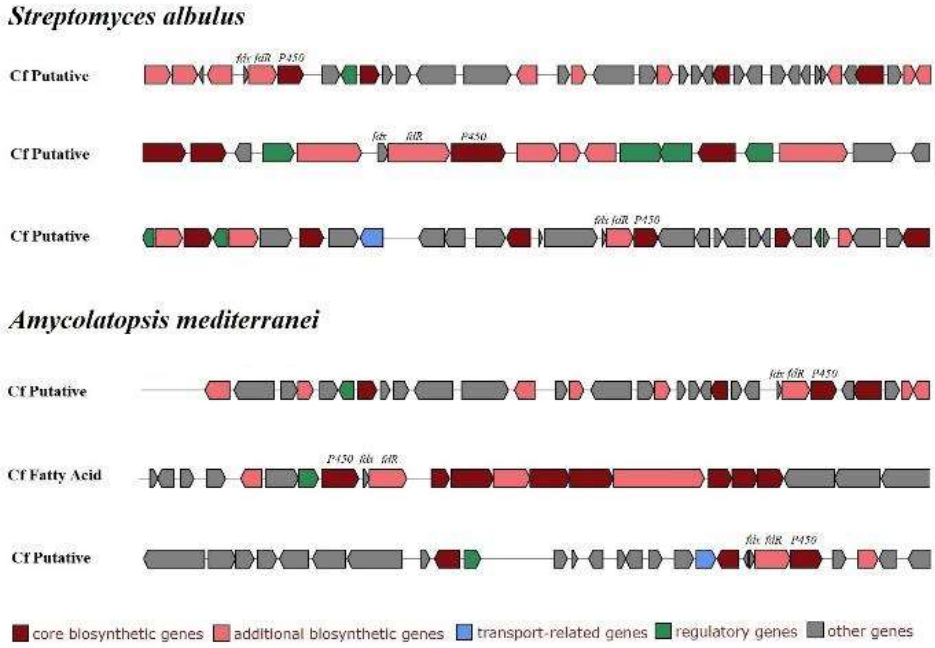

Figure 6: The representative gene clusters of P450 class I redox system genes in the genomes of *Streptomyces albulus* and *Amycolatopsis mediterranei*
The diagram shows the P450 class I system genes (P450, ferredoxin and ferredoxin reductase) found adjacent to each other in different gene clusters in genomes of *Streptomyces albulus* and *Amycolatopsis mediterranei*. Legends in the figure showing the different categories of genes in gene cluster.

ferredoxin genes as compare to the ferredoxin reductase genes, while ferredoxin reductase genes locate close to the ferredoxin genes as compare to P450 genes. It suggests that, in many genomes, ferredoxin genes locate between its closest P450 and the closest ferredoxin reductase genes in the genomes. As Chun *et al.* showed that the P450 gene could interact with distant electron transfer protein in its electron transfer pathway [69], it suggests that the ferredoxin reductase genes, which locate distantly from P450 genes, may interact with P450 genes as its electron transfer protein. Specifically, the analysis of *Mycobacterium tuberculosis* also showed that the P450 genes are located very close to ferredoxin genes as compare to ferredoxin reductase genes. The distantly located ferredoxin reductase suggests that these genes may interact with P450 and ferredoxin while it locates distant from them.

P450s are often linked with the secondary metabolism pathways [57], and the P450 genes sequences are also located within the respective secondary metabolite gene clusters in bacteria [57]. Correspondingly, many studies reported that P450 genes express in gene clusters or operons with ferredoxin and ferredoxin reductase genes [23, 25]. Gene cluster analysis suggests that P450s genes clustered with its ferredoxin redox proteins in functionally different type of gene clusters. It is reported that P450 genes clustered with ferredoxin and ferredoxin reductase genes [28, 31] and our results suggest that these three genes are mostly clustered in *Streptomyces*. In this study we found many gene clusters which are already published in previous studies. Lamb *et al.* [28] reported that, in *Streptomyces avermitilis*, two P450 genes express in operon/gene cluster with ferredoxin and ferredoxin genes and our results confirmed that these genes are found in gene clusters in *Streptomyces avermitilis* genome. Meanwhile, in many genomes, P450 genes clustered with ferredoxin genes without ferredoxin reductase genes (reported by Lamb *et al.* [27] and Li *et al.* [31]), mostly found in *Mycobacterium*. It was reported that P450 genes locate next to the ferredoxin genes without

ferredoxin reductase genes in *Streptomyces coelicolor*, *Streptomyces griseus*, and *Streptomyces lividans*, and we found the same gene arrangement in these species. Parajuli *et al.* [30] reported that P450 gene with ferredoxin reductase gene are found in gene cluster without ferredoxin genes, and our results propose that these type of gene clusters are found mostly in *Streptomyces*. Interestingly, we also found that ferredoxin genes are clustered with ferredoxin reductase genes without P450 genes. However, in many genomes, P450 genes express with ferredoxin genes in a gene cluster without ferredoxin reductase genes. It indicates that ferredoxin reductase genes may express separately from the P450 containing gene clusters and interact with many systems, including the P450 system. Most of the genomes, which have P450-ferredoxin system redox genes containing clusters, belong to phylum *Actinobacteria*, which is consistent with rich secondary metabolites found in this phylum as P450 is often used by the bacteria to synthesize complicated compounds.

CONCLUSIONS

In conclusion, this report investigated the distribution and interaction of cytochromes P450 class I redox system genes in bacteria. The results indicate that a good portion of bacterial genomes does not have the P450 redox systems. The analysis of P450 class I redox protein systems suggests that such systems are frequently found in *Proteobacteria*. Meanwhile, the absence of P450 in many bacteria, such as *E. coli*, indicates that this system is not required for primary metabolism. Analysis of intergenic distance between the genes in the P450 class I redox system shows that most of the ferredoxin genes locate close to P450 genes, but not all the ferredoxin reductase genes are found nearby. In 713 bacterial genomes, the distance between P450 and ferredoxin genes are less than 3kb, indicating that these genes may work within the same gene clusters. Through gene cluster analysis, we found that P450s are clustered with ferredoxin and/or ferredoxin reductase, mostly in *Actinobacteria* genomes. Overall, our study provides novel insights into the distribution and interaction of cytochromes P450 class I redox system genes in bacteria.

Competing interests
The authors declare that they have no competing interests.

Conflict of interest statement
The authors declare that they do not have any conflict of interest.

REFERENCES

1. Gotoh, O., *Evolution of cytochrome p450 genes from the viewpoint of genome informatics.* Biol Pharm Bull, 2012. **35**(6): p. 812-7.
2. Omura, T. and R. Sato, *THE CARBON MONOXIDE-BINDING PIGMENT OF LIVER MICROSOMES. II. SOLUBILIZATION, PURIFICATION, AND PROPERTIES.* J Biol Chem, 1964. **239**: p. 2379-85.
3. Gonzalez, F.J. and H.V. Gelboin, *Human cytochromes P450: evolution and cDNA-directed expression.* Environ Health Perspect, 1992. **98**: p. 81-5.
4. Peter Guengerich, F., *Cytochrome P450: what have we learned and what are the future issues?* Drug metabolism reviews, 2004. **36**(2): p. 159-197.
5. Nelson, D.R., *Cytochrome P450 diversity in the tree of life.* Biochim Biophys Acta Proteins Proteom, 2018. **1866**(1): p. 141-154.
6. Parvez, M., et al., *Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s.* Sci Rep, 2016. **6**: p. 33099.
7. Sono, M., et al., *Heme-Containing Oxygenases.* Chem Rev, 1996. **96**(7): p. 2841-2888.
8. Pikuleva, I. and M. Waterman, *Cytochromes P450 in synthesis of steroid hormones, bile acids, vitamin D3 and cholesterol.* Mol Aspects Med, 1999. **20**(1-2): p. 33-42, 43-37.
9. Werck-Reichhart, D. and R. Feyereisen, *Cytochromes P450: a success story.* Genome Biol, 2000. **1**(6): p. Reviews3003.
10. Xue, Y. and D.H. Sherman, *Biosynthesis and combinatorial biosynthesis of pikromycin-related macrolides in Streptomyces venezuelae.* Metab Eng, 2001. **3**(1): p. 15-26.
11. Guengerich, F.P., *Common and uncommon cytochrome P450 reactions related to metabolism*

and chemical toxicity. Chem Res Toxicol, 2001. **14**(6): p. 611-50.
12. Sligar, S.G., T.M. Makris, and I.G. Denisov, *Thirty years of microbial P450 monooxygenase research: peroxo-heme intermediates—the central bus station in heme oxygenase catalysis.* Biochemical and biophysical research communications, 2005. **338**(1): p. 346-354.
13. Rodriguez-Antona, C. and M. Ingelman-Sundberg, *Cytochrome P450 pharmacogenetics and cancer.* Oncogene, 2006. **25**(11): p. 1679-91.
14. Bernhardt, R., *Cytochromes P450 as versatile biocatalysts.* J Biotechnol, 2006. **124**(1): p. 128-45.
15. Guengerich, F.P. and A.W. Munro, *Unusual cytochrome p450 enzymes and reactions.* J Biol Chem, 2013. **288**(24): p. 17065-73.
16. de Montellano, P.R.O., *Substrate oxidation by cytochrome P450 enzymes*, in *Cytochrome P450*. 2015, Springer. p. 111-176.
17. Hannemann, F., et al., *Cytochrome P450 systems--biological variations of electron transport chains.* Biochim Biophys Acta, 2007. **1770**(3): p. 330-44.
18. Ewen, K.M., M. Kleser, and R. Bernhardt, *Adrenodoxin: the archetype of vertebrate-type [2Fe–2S] cluster ferredoxins.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2011. **1814**(1): p. 111-125.
19. Waskell, L. and J.-J.P. Kim, *Electron transfer partners of cytochrome p450*, in *Cytochrome P450*. 2015, Springer. p. 33-68.
20. McLean, K.J., et al., *Biological diversity of cytochrome P450 redox partner systems.* Adv Exp Med Biol, 2015. **851**: p. 299-317.
21. Ikeda, H., et al., *Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis.* Nature biotechnology, 2003. **21**(5): p. 526.
22. McLean, K.J., et al., *The preponderance of P450s in the Mycobacterium tuberculosis genome.* Trends in microbiology, 2006. **14**(5): p. 220-228.
23. Koga, H., et al., *Cloning and nucleotide sequences of NADH-putidaredoxin reductase gene (camA) and putidaredoxin gene (camB) involved in cytochrome P-450cam hydroxylase of Pseudomonas putida.* J Biochem, 1989. **106**(5): p. 831-6.
24. Peterson, J.A., M.C. Lorence, and B. Amarneh, *Putidaredoxin reductase and putidaredoxin. Cloning, sequence determination, and heterologous expression of the proteins.* J Biol Chem, 1990. **265**(11): p. 6066-73.
25. Peterson, J.A., et al., *Cytochrome P-450terp. Isolation and purification of the protein and cloning and sequencing of its operon.* J Biol Chem, 1992. **267**(20): p. 14193-203.
26. Tully, R.E., et al., *Identification and sequencing of a cytochrome P450 gene cluster from Bradyrhizobium japonicum.* Biochim Biophys Acta, 1998. **1398**(3): p. 243-55.
27. Lamb, D.C., et al., *The cytochrome P450 complement (CYPome) of Streptomyces coelicolor A3(2).* J Biol Chem, 2002. **277**(27): p. 24000-5.
28. Lamb, D.C., et al., *Cytochrome p450 complement (CYPome) of the avermectin-producer Streptomyces avermitilis and comparison to that of Streptomyces coelicolor A3(2).* Biochem Biophys Res Commun, 2003. **307**(3): p. 610-9.
29. Lei, L., et al., *Availability of specific reductases controls the temporal activity of the cytochrome P450 complement of Streptomyces coelicolor A3(2).* Proc Natl Acad Sci U S A, 2004. **101**(2): p. 494-9.
30. Parajuli, N., et al., *Genome analyses of Streptomyces peucetius ATCC 27952 for the identification and comparison of cytochrome P450 complement with other Streptomyces.* Arch Biochem Biophys, 2004. **425**(2): p. 233-41.
31. Li, Z.Z., et al., *Identification and functional analysis of cytochrome P450 complement in Streptomyces virginiae IBL14.* BMC Genomics, 2013. **14**: p. 130.
32. Sasaki, M., et al., *Purification of cytochrome P450 and ferredoxin, involved in bisphenol A degradation, from Sphingomonas sp. strain AO1.* Appl Environ Microbiol, 2005. **71**(12): p. 8024-30.
33. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nat Methods, 2015. **12**(1): p. 59-60.
34. Blin, K., et al., *antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification.* Nucleic Acids Res, 2017. **45**(W1): p. W36-w41.

35. Ventura, M., et al., *Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum.* Microbiology and molecular biology reviews : MMBR, 2007. **71**(3): p. 495-548.
36. Falkinham, J.O., 3rd, *Epidemiology of infection by nontuberculous mycobacteria.* Clinical microbiology reviews, 1996. **9**(2): p. 177-215.
37. Collins, C.H., J.M. Grange, and M.D. Yates, *Mycobacteria in water.* J Appl Bacteriol, 1984. **57**(2): p. 193-211.
38. Wolinsky, E. and T.K. Rynearson, *Mycobacteria in soil and their relation to disease-associated strains.* Am Rev Respir Dis, 1968. **97**(6): p. 1032-7.
39. Capyk, J.K., et al., *Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids.* J Biol Chem, 2009. **284**(51): p. 35534-42.
40. Ouellet, H., J.B. Johnston, and P.R. Ortiz de Montellano, *The Mycobacterium tuberculosis cytochrome P450 system.* Archives of Biochemistry and Biophysics, 2010. **493**(1): p. 82-95.
41. Hudson, Sean A., et al., *Mycobacterium tuberculosis cytochrome P450 enzymes: a cohort of novel TB drug targets.* Biochemical Society Transactions, 2012. **40**(3): p. 573-579.
42. Senate, L.M., et al., *Similarities, variations, and evolution of cytochrome P450s in Streptomyces versus Mycobacterium.* Sci Rep, 2019. **9**(1): p. 3962.
43. Syed, P.R., et al., *Cytochrome P450 Monooxygenase CYP139 Family Involved in the Synthesis of Secondary Metabolites in 824 Mycobacterial Species.* International journal of molecular sciences, 2019. **20**(11): p. 2690.
44. Gupta, R.S., B. Lo, and J. Son, *Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus Mycobacterium into an Emended Genus Mycobacterium and Four Novel Genera.* Frontiers in Microbiology, 2018. **9**(67).
45. Oren, A. and G. Garrity, *List of new names and new combinations previously effectively, but not validly, published.* International Journal of Systematic and Evolutionary Microbiology, 2018. **68**(5): p. 1411-1417.
46. McLean, K.J., et al., *Azole antifungals are potent inhibitors of cytochrome P450 mono-oxygenases and bacterial growth in mycobacteria and streptomycetes.* Microbiology, 2002. **148**(10): p. 2937-2949.
47. Jackson, C.J., et al., *Conservation and cloning of CYP51: a sterol 14α-demethylase from Mycobacterium smegmatis.* Biochemical and Biophysical Research Communications, 2003. **301**(2): p. 558-563.48. Frank, D.J., et al., *Cytochrome P450 125A4, the Third Cholesterol C-26 Hydroxylase from Mycobacterium smegmatis.* Biochemistry, 2015. **54**(46): p. 6909-6916.
49. Ortiz de Montellano, P.R., *Potential drug targets in the Mycobacterium tuberculosis cytochrome P450 system.* J Inorg Biochem, 2018. **180**: p. 235-245.
50. Procópio, R.E., et al., *Antibiotics produced by Streptomyces.* Braz J Infect Dis, 2012. **16**(5): p. 466-71.
51. Martín, J.F. and J.F. Aparicio, *chapter 10 Enzymology of the Polyenes Pimaricin and Candicidin Biosynthesis*, in *Methods in Enzymology*. 2009, Academic Press. p. 215-242.
52. Lamb, D.C., M.R. Waterman, and B. Zhao, *Streptomyces cytochromes P450: applications in drug metabolism.* Expert Opin Drug Metab Toxicol, 2013. **9**(10): p. 1279-94.
53. Sun, D., et al., *Connecting Metabolic Pathways: Sigma Factors in Streptomyces spp.* Frontiers in Microbiology, 2017. **8**(2546).
54. Rudolf, J.D., et al., *Cytochromes P450 for natural product biosynthesis in Streptomyces: sequence, structure, and function.* Natural product reports, 2017. **34**(9): p. 1141-1172.
55. Cho, M.-A., et al., *Streptomyces Cytochrome P450 Enzymes and Their Roles in the Biosynthesis of Macrolide Therapeutic Agents.* Biomolecules & therapeutics, 2019. **27**(2): p. 127-133.
56. Johnson, Z.I. and S.W. Chisholm, *Properties of overlapping genes are conserved across microbial genomes.* Genome research, 2004. **14**(11): p. 2268-2272.

57. Greule, A., et al., *Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism.* Nat Prod Rep, 2018.
58. De Mot, R. and A.H. Parret, *A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes.* Trends Microbiol, 2002. **10**(11): p. 502-8.
59. Roberts, G.A., et al., *Identification of a new class of cytochrome P450 from a Rhodococcus sp.* J Bacteriol, 2002. **184**(14): p. 3898-908.
60. Warman, A.J., et al., *Flavocytochrome P450 BM3: an update on structure and mechanism of a biotechnologically important enzyme.* Biochem Soc Trans, 2005. **33**(Pt 4): p. 747-53.
61. Zotchev, S.B. and C.R. Hutchinson, *Cloning and heterologous expression of the genes encoding nonspecific electron transport components for a cytochrome P450 system of Saccharopolyspora erythraea involved in erythromycin production.* Gene, 1995. **156**(1): p. 101-106.
62. Hawkes, D.B., et al., *Cytochrome P450cin (CYP176A), isolation, expression, and characterization.* Journal of Biological Chemistry, 2002. **277**(31): p. 27725-27732.
63. Pandey, B.P., et al., *Identification of the specific electron transfer proteins, ferredoxin, and ferredoxin reductase, for CYP105D7 in Streptomyces avermitilis MA4680.* Appl Microbiol Biotechnol, 2014. **98**(11): p. 5009-17.
64. Trower, M.K., M.H. Emptage, and F.S. Sariaslani, *Purification and characterization of a 7Fe ferredoxin from Streptomyces griseus.* Biochim Biophys Acta, 1990. **1037**(3): p. 281-9.
65. Hussain, H.A. and J.M. Ward, *Enhanced heterologous expression of two Streptomyces griseolus cytochrome P450s and Streptomyces coelicolor ferredoxin reductase as potentially efficient hydroxylation catalysts.* Appl Environ Microbiol, 2003. **69**(1): p. 373-82.
66. Kelly, S.L., et al., *The biodiversity of microbial cytochromes P450.* Adv Microb Physiol, 2003. **47**: p. 131-86.
67. Sawada, N., et al., *Conversion of vitamin D3 to 1alpha,25-dihydroxyvitamin D3 by Streptomyces griseolus cytochrome P450SU-1.* Biochem Biophys Res Commun, 2004. **320**(1): p. 156-64.
68. Wang, F., et al., *Involvement of the cytochrome P450 system EthBAD in the N-deethoxymethylation of acetochlor by Rhodococcus sp. strain T3-1.* Appl Environ Microbiol, 2015. **81**(6): p. 2182-8.
69. Chun, Y.J., et al., *Electron transport pathway for a Streptomyces cytochrome P450: cytochrome P450 105D5-catalyzed fatty acid hydroxylation in Streptomyces coelicolor A3(2).* J Biol Chem, 2007. **282**(24): p. 17486-500.

.