ARTICLE

Understanding the nature and dynamics of *Mycobacterium ulcerans* cytochrome P450 monooxygenases (CYPs) – a bioinformatics approach

Saubashya Sur

Postgraduate Department of Botany, Life Sciences Block, Ramananda College, Bishnupur-722122, West Bengal, India

ABSTRACT		Cytochrome P450 monooxygenases (CYPs or P450s) are catalytically versatile hemoproteins, associated with drug metabolism, substrate utilization and pathogenesis. *Mycobacterium ulcerans* is a human pathogen causing Buruli ulcer. The study intended to investigate frequency and diversity of CYPs from *M. ulcerans* strains, understand the pan-CYPome clustering patterns and interconnection of CYPs using bioinformatics tools. *M. ulcerans* strains demonstrated the presence of 261 CYPs categorized into 35 families and 38 subfamilies. CYP138, CYP140, CYP189 and CYP125 were the flourishing families. Around, 20 CYP families and 20 subfamilies were conserved. Flourishing and conserved CYP families/subfamilies were associated with lipid metabolism, substrate utilization etc. CYP140 had a role in pathogenesis. CYP279 was the least dominant family. CYP135, CYP183, CYP190, CYP271 and CYP276 were diagnostic markers for *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741. The pan-CYPome specified that *M. ulcerans* is evolving by gaining CYPs. CYP centric clustering revealed diversity and resemblances among *M. ulcerans* strains. More diverse nature of the *M. ulcerans* strain Harvey could be attributed to its larger size and geographical location. Co-occurrence network demonstrated mutual associations amongst substantial number of CYP families/subfamilies. This work provided comprehensive understanding of previously unexplored CYPs from *M. ulcerans*.		**Acta Biol Szeged 65(1):93-103 (2021)**

Introduction

Hemoproteins like cytochrome P450 monooxygenases (CYPs) are universally present among all the living kingdoms (Kweon et al. 2020). The accessibility of large number of plants, animal and bacterial genomes in the public databases during last 20 years have resulted in the documentation of numerous cytochrome P450 monooxygenases. These hemo-thiolate enzymes first discovered in 1962, have been known to be vital for the evolution of different organisms (Nelson 2013). Phylogenetic studies on cytochrome P450 monooxygenases have highlighted that CYP51 is an ancient P450 which gave rise to the contemporary CYPs (Yoshida et al. 2000; Nelson 1999). CYP51 remained conserved across prokaryotes and eukaryotes (Parvez et al. 2016). Cytochrome P450 monooxygenases are associated with several biochemical reactions (Coon 2005) and form an indispensable part of an organism's primary and secondary metabolism (Parvez et al. 2016). Researches on the CYPs from bacteria have revealed their intricate structure function relationships (Parvez et al.

2016). Multiple lines of evidence have demonstrated the role of bacterial CYPs in drug metabolism and pathogenesis (Kweon et al. 2020; Furge and Guengerich 2006; Brezna et al. 2006). Bacterial CYPs are also known to be biotechnologically important (McLean et al. 2015).

Non-tuberculous mycobacteria dwell in varied environments and are responsible for nosocomial infections in immunocompromised and vulnerable persons (Ahmed et al. 2020). *Mycobacterium ulcerans* is a non-tuberculous mycobacterium responsible for a severe skin disease called Buruli ulcer (BU) (Ohtsuka et al. 2013). It has been a common infectious disease in Ivory Coast, Ghana, Benin, Democratic Republic of Congo and Uganda (van der Werf et al. 1999). However, a number of infections has been reported from aquatic environments in Japan (Luo et al. 2015), Australia (O' Brien et al. 2019), China, certain regions of Central and South America (Merritt et al. 2010). Anthropogenic activities have been one of the reasons behind the emergence *M. ulcerans* (Zingue et al. 2018). It initiates inflammation by attacking the skin, subcutaneous fat cells (Merritt et al. 2010) and secretes mycolactone toxin (Liu et al. 2019) during Buruli ulcer.

Table 1. Summary of the analysis of MU-7 dataset using CYPminer

| Dataset | Total No. of CYPs | CYP categorization | | Pan-CYPome | | | | Principal CYP families | Principal CYP subfamilies |
| | | | | Pan | | Core | | | |
		Family	Subfamily	Family	Subfamily	Family	Subfamily		
MU-7	261	35	38	35	38	20	20	CYP138	CYP164A3P
								CYP140	CYP140A
								CYP189	CYP125A6P
								CYP125	CYP138A

Consequently, a necrotizing and painless infection occur which damages the skin, tissues of upper and lower limbs as well as bones (de Souza et al. 2012). Delay in treatment often result in disabilities (Yotsu et al. 2015). Till date no vaccine is available, and surgery seems to be an option for treating Buruli ulcer (Tai et al. 2018). Combination of a couple of antibiotics along with surgery has also been successful (Yotsu et al. 2018). Still, none had been more effective compared to the other (Yotsu et al. 2018) and increased pathogenicity has been linked to stable evolution of *M. ulcerans* (Tai et al. 2018).

There have been limited studies on the bacterial CYPs (Parvez et al. 2016; Mthetwa et al. 2018; Senate et al. 2019; Lau et al. 2019; Khumalo et al. 2020) compared to eukaryotes (Parvez et al. 2016). Nevertheless, observations on the role of CYPs in bacterial pathogenesis (Kweon et al. 2020), use of CYP141 for diagnosis of *Mycobacterium tuberculosis* (Darban-Sarokhalil et al. 2011) and indispensability of CYP121 for the existence of tuberculosis pathogen (McLean et al. 2008) demonstrated its importance in infectious bacteria. An evolutionary analysis of CYPs from 60 mycobacterial species offered insights into their coverage, conservation and structural dynamics (Parvez et al. 2016). Others focused on genome level analysis, especially identification and classification of CYPs from some mycobacterial species along with eukaryotes (Kweon et al. 2020). Nevertheless, there is a dearth of work on CYPs from *M. ulcerans*. There is scarcity from the perspective of identification, classification and comparative analysis of the total CYP complement from *M. ulcerans*. Given the significance of *M. ulcerans* as an emerging pathogen with steady evolutionary rate (Tai et al. 2018), there is a need to recognize the nature and diversity of CYPs from the total CYP complement. Availability of the complete genomes of different strains of *M. ulcerans*, isolated from different geographical locations has unlocked the prospect of studying the whole CYP complement.

The objectives of this study were to a) identify and categorize the CYPs from *M. ulcerans* strains, b) investigate the nature of the pan-CYPome, c) survey CYP centric clustering patterns, d) understand interconnection of CYPs via co-occurrence network and e) examine the CYP cloud to recognize the frequency and diversity of CYPs. Given the wide-ranging functionalities of CYPs, this sort of investigation will generate significant insights into the characteristic features, patterns of dynamics and diversity of *M. ulcerans* CYPs. Furthermore, it will throw light on the nature of the physiology and lifestyle of *M. ulcerans.*

Materials and methods

Retrieval of sequences

Protein sequences from the genomes of 7 strains of human pathogenic *M. ulcerans* (publicly available as of 01/01/2021), viz., *M. ulcerans* strain SGL03 (Genome ID 1809.14), *M. ulcerans* strain S4018 (Genome ID 1809.16), latest annotated version of *M. ulcerans* strain P7741 (Genome ID 1809.18), *M. ulcerans* strain CSURQ0185 (Genome ID 1809.19), *M. ulcerans* strain Agy99 (Genome ID 362242.7) along with its plasmid, latest annotated version of *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 (Genome ID 1124626.4) along with its plasmid and *M. ulcerans* strain Harvey (Genome ID 1299332.3) were retrieved from the PATRIC database version 3.6.8. This complete dataset comprising 7 strains of *M. ulcerans* will henceforth be referred to as MU-7.

Software and databases

A Python based software CYPminer (https://github.com/Okweon/CYPminer), clustering algorithm USEARCH (Edgar 2010) and RPSBLAST software (Marchler-Bauer et al. 2009) were downloaded. Additionally, P450 and RPS databases necessary for running CYPminer were also retrieved (https://github.com/Okweon/CYPminer).

Data pre-processing and orthologous clustering

Protein sequences from the MU-7 dataset were used as input for CYPminer. Next, USEARCH (Edgar 2010) was utilized by CYPminer for the purpose of clustering with a sequence identity cut-off of 55%. This was done to accomplish CYP specific evaluation. A FASTA file with representative protein sequences and an UCLUST

Table 2. Frequency data of CYPs at the family and sub family in *M. ulcerans* strains and plasmids from the MU-7 dataset

Organism with No. of coding sequences (CDS)	CYP frequency	
	Family	Subfamily
M. ulcerans strain SGL03 (PATRIC CDS=5611)	34	34
M. ulcerans strain S4018 (PATRIC CDS=5926)	36	36
M. ulcerans strain P7741 (PATRIC CDS=5894)	37	37
M. ulcerans strain CSURQ0185 (PATRIC CDS=5945)	34	34
M. ulcerans strain Agy99 (PATRIC CDS=5844)	38	38
M. ulcerans subsp. shinshuense strain ATCC 33728 (PATRIC CDS=5689)	37	37
M. ulcerans strain Harvey (PATRIC CDS=9118)	45	45

formatted file harboring cluster information was generated as output.

Identification and categorization of CYPs

In order to identify sequences with CYPs, CYPminer executed the RPSBLAST software by utilizing FASTA file of the representative protein sequences. Subsequently, these representative protein sequences were compared with the localized RPS database housing CYP domains. A cut-off was set at 0.00001. Next, the protein sequences containing conserved CYP domains were used for USEARCH BLAST against the localized P450 database (Kweon et al. 2020). Then, CYPminer categorized CYP proteins having ≥40% and ≥55% identity to corresponding P450 families and subfamilies, while those having <40% identity was designated as novel. This categorization was based on international nomenclature (Nebert et al. 1987).

Determination of pan-CYPome and clustering

Here, CYPminer utilized the UCLUST file housing cluster information and the result of CYP categorization for a number of Genome-CYP Matrix (GCM) analysis. Investigation of the pan-CYPome and clustering are integrated components of the GCM analysis. CYPminer yielded a pan-CYPome comprising of total complement of CYPs (Kweon et al. 2020) obtained from the studied strains of *M. ulcerans*. Moreover, it also yielded the core-CYPome. Additionally, hierarchically clustered heatmaps of CYP family and subfamilies were created by utilizing the python heatmap cluster library.

Determination of co-occurrence network and CYP cloud

The CYPminer workflow was subjected to GCM analysis for constructing the CYP co-occurrence network and CYP cloud. For this purpose, the Python pyvis and word cloud libraries were employed. While CYP co-occurrence network signified relationships between CYPs, CYP cloud examined CYP frequency. The former considered CYPs present within a genome to be related resulting in co-occurrence, while the latter generated a word cloud

containing an impression of the rate of CYP incidence. The CYP co-occurrence networks were generated for both CYP families and subfamilies.

Results

Mycobacterium ulcerans CYPs and their categorization

Data mining of the MU-7 dataset using CYPminer revealed the occurrence of 261 CYPs (Table 1). These CYPs were categorized into 35 families and 38 subfamilies. There were no new families or subfamilies. Closer inspection of CYP families indicated that CYP138, CYP140, CYP189 and CYP125 were the principal ones having 19, 13, 13 and 12 members respectively. About 20 CYP families viz. CYP51, CYP105, CYP123-126, CYP130, CYP136-139, CYP142-143, CYP150, CYP164, CYP185, CYP189, CYP191, CYP226 and CYP278 were present among all the 7 strains of *M. ulcerans* indicating their conserved nature. In contrast, the CYP279 family was only present in *M. ulcerans* strain P7741. CYP families viz. CYP135, CYP147, CYP183, CYP190, CYP271 and CYP276 were only found in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741. Likewise, CYP140A, CYP2164A3P, CYP125A6P and CYP138A were the principal subfamilies housing 13, 13, 12 and 12 members. About 20 CYP subfamilies viz. CYP51B, CYP105Q, CYP123-124A, CYP125A6, CYP126A, CYP130A, CYP136-138A, CYP139A3P, CYP142-143A, CYP150A, CYP164A3P, CYP185A, CYP189A, CYP191A, CYP226B and CYP278A were extant among all the 7 strains of *M. ulcerans*. While CYP138B occurred only in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728, CYP279A was only found in *M. ulcerans* strain P7741. Moreover, CYP subfamilies viz. CYP135B, CYP147G, CYP183B, CYP190A, CYP271A and CYP276A were noticed only in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741.

Table 1. showed the occurrence of 35 and 38 pan CYP families and subfamilies. Table 2. highlighted that *M. ul-*
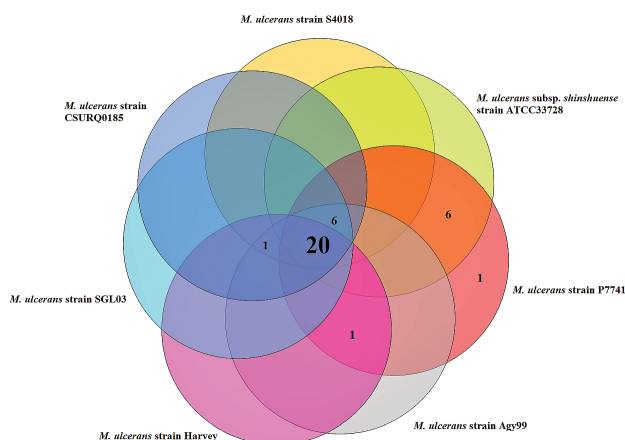
Figure 1. Pan-CYPome family profile from the MU-7 dataset. The Venn diagram intended to demonstrate the Pan and Core CYP families. Circles symbolized genomes and the overlapping regions signified CYP families shared by respective genomes. Numeral figures represented the number of CYP families found therein. The number of Core families (in the centre) in the Pan-CYPome were 20. The Pan-CYPome consisted of 35 families in total.

Figure 2. Pan-CYPome subfamily profile from the MU-7 dataset. The Venn diagram intended to demonstrate the Pan and Core CYP subfamilies. Circles symbolized genomes and the overlapping regions signified CYP subfamilies shared by respective genomes. Numeral figures represented the number of CYP subfamilies found therein. The number of Core subfamilies (in the centre) in the Pan-CYPome were 20. The Pan-CYPome consisted of 38 subfamilies in total.

cerans strain Harvey housed the highest number of CYPs (45 families and subfamilies) followed by *M. ulcerans* strain Agy99 (38 families and subfamilies). *M. ulcerans* strain SGL03 and *M. ulcerans* strain CSURQ0185 had the lowest CYP frequency (34 families and subfamilies each) amongst the 7 strains. The frequency of CYPs corresponded with the number of coding sequences in some of the strains.

Pan-CYPome of Mycobacterium ulcerans

Fig. 1. demonstrated the pan-CYPome family profile from MU-7 dataset. It highlighted the full complement of CYP families incorporating the studied strains of *M. ulcerans*. CYPMiner showed a total of 35 families in the pan-CYPome. This consisted of 20 core families that were common to all the seven strains. Thus, a sizeable proportion of the CYP families belong to the core. Six CYP families were shared among six strains. *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741 also shared 6 CYP families. The number of CYP family exclusively shared by five strains was 1. Three strains viz., *M. ulcerans* strain Harvey, *M. ulcerans* strain Agy99 and *M. ulcerans* strain P7741 too shared 1 CYP family. Additionally, the presence of one unique strain specific CYP family in *M. ulcerans* strain P7741 supported earlier observations.

Fig. 2. portrayed the pan-CYPome subfamily profile from MU-7, dataset signifying the full complement of CYP subfamilies. A total of 38 subfamilies were present in the pan-CYPome. A few parallels were noticed between the pan-CYPome family and subfamily profiles. The pan-CYPome subfamily profile comprised of a high
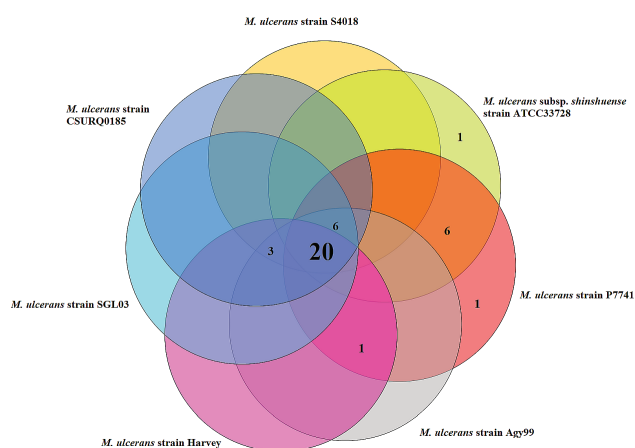
share of 20 core CYP subfamilies present in all seven trains. Similar to the pan-CYPome family profile, six CYP subfamilies were also shared between six strains. *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741 too shared 6 CYP subfamilies. Three CYP subfamilies were exclusively shared by five strains. *M. ulcerans* strain Harvey, *M. ulcerans* strain Agy99 and *M. ulcerans* strain P7741 too shared 1 CYP subfamily. Unique strain specific subfamilies were observed only in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741 respectively.

CYP centric clustering analysis

Fig. 3. portrayed the CYP centric family heatmap cluster with dendrograms. It was observed that the 7 *M. ulcerans* genomes were grouped into two major clusters consisting of smaller clusters. It was noticed from Fig. 3 that *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741 had a comparable CYP family profile and were distinctly clustered together at some distance from the rest. This was the first major cluster. In this cluster, the frequency of 19 CYP families like CYP51, CYP105, CYP124-126, CYP130, CYP136-137, CYP139, CYP142, CYP147, CYP183, CYP185, CYP189-190, CYP226, CYP271, CYP276 and CYP278 was 1 in both the organisms. Similarly, the frequency of CYP135 and CYP150 were 2 in both. Furthermore, the frequency of CYP144 were 0 in both the organisms. The frequency of other CYP families varied between these microorganisms.

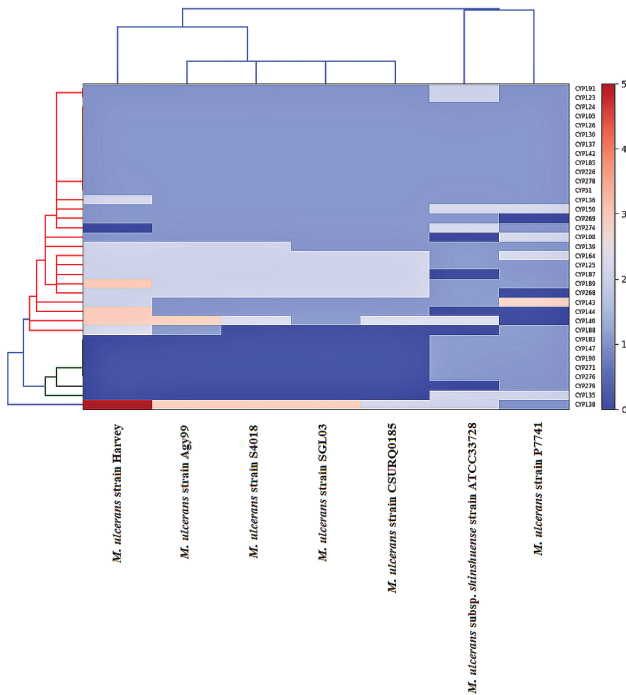Others viz. *M. ulcerans* strain Harvey, *M. ulcerans* strain

Figure 3. CYP centric family heat map cluster from *M. ulcerans*. Right side of the vertical axis denoted the frequency of incidence of individual CYP families while left side denoted relationship among CYP subfamilies. Lower side of the horizontal axis represented the studied strains of *M. ulcerans*. Upper side of the horizontal axis represented the CYP centric phylogenetic clusters. Colours in the scale of 0-5 indicated frequency.

Figure 4. CYP centric subfamily heat map cluster from *M. ulcerans*. Right side of the vertical axis denote the frequency of incidence of individual CYP subfamilies and left side relationships among CYP subfamilies. Lower side of the horizontal axis represented the studied strains of *M. ulcerans*. Upper side of the horizontal axis represented CYP centric phylogenetic clusters. Colours in the scale of 0-3 indicated frequency.

CSURQ0185, *M. ulcerans* strain SGL03, *M. ulcerans* strain S4018 and *M. ulcerans* strain Agy99, formed the second major cluster. They were divided into smaller clusters which were close to each other. A closer investigation of these 5 genomes revealed that, *M. ulcerans* strain S4018, *M. ulcerans* strain Agy99, *M. ulcerans* strain SGL03 and *M. ulcerans* strain CSURQ0185 were assembled together owing to some resemblances in the CYP family profile. However, *M. ulcerans* strain Harvey showed some variations regarding frequencies of CYP136, CYP138, CYP140, CYP143-144, CYP188-189 and CYP274. As far as the parallels are concerned, the frequency of CYP135, CYP147, CYP183, CYP190, CYP271, CYP276 and CYP279 were 0 in all these 5 genomes. Likewise, the frequency of CYP125, CYP164, CYP187 and CYP268 were 2 in these. A frequency of 1 was observed in 14 CYP families viz. CYP51, CYP105, CYP123-124, CYP126, CYP130, CYP137, CYP142, CYP150, CYP185, CYP191, CYP226, CYP269 and CYP278 within these 5 genomes. Overall, all 7 genomes showed resemblance in having a frequency of 1 for 10 CYP families viz. CYP51, CYP105, CYP124, CYP126, CYP130, CYP137, CYP142, CYP185, CYP226
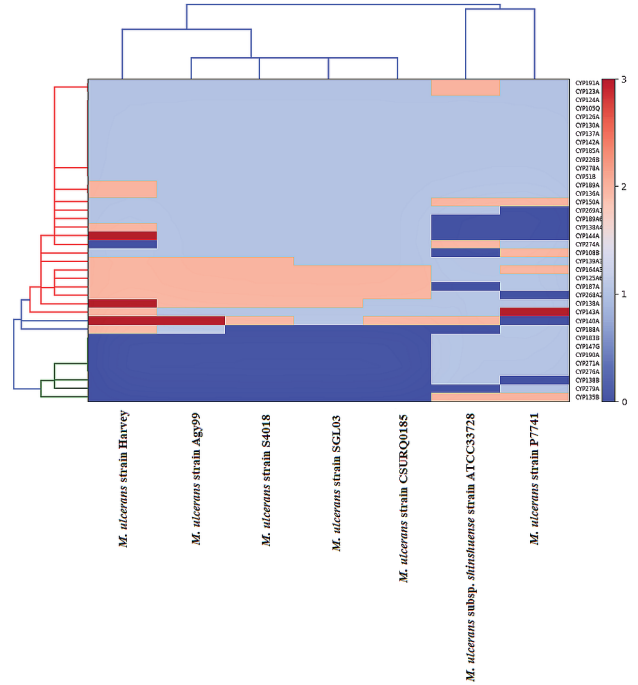
and CYP278.

The CYP centric subfamily heatmap cluster with dendrograms (Fig. 4) which showed some similarity with the pattern visible in the family heatmap. Here too the 7 *M. ulcerans* genomes were grouped into two main clusters. Fig. 4. demonstrated that *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741 had good degree of similarity in the CYP subfamily profile. They were grouped together forming one of the two major clusters at a distance from others. In this cluster, the frequency of 20 CYP subfamilies viz. CYP51B, CYP105Q, CYP124A, CYP125A6P, CYP126A, CYP130A, CYP136-138A, CYP139A3P, CYP142A, CYP147G, CYP183B, CYP185A, CYP189-190A, CYP226B, CYP271A, CYP274A, CYP276A and CYP278A were 1 in both. Likewise, the frequency was 2 for CYP135B and CYP150A. Additionally, the frequency of CYP138A4P, CYP144A and CYP189A6P was 0 in both. *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741 showed differences in rest of the subfamilies.

M. ulcerans strain Harvey, *M. ulcerans* strain Agy99, *M. ulcerans* strain S4018, *M. ulcerans* strain SGL03 and *M. ulcerans* strain CSURQ0185 shaped the second major
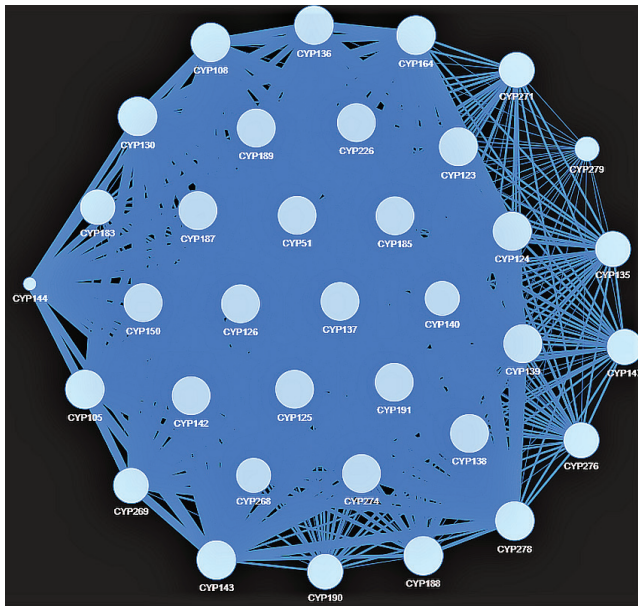
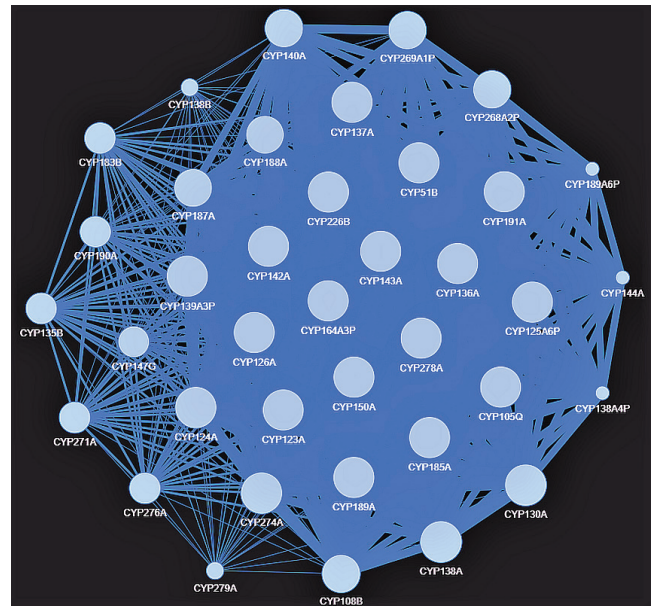**Figure 5.** CYP family co-occurrence network from MU-7 dataset.

Figure 6. CYP subfamily co-occurrence network from MU-7 dataset.

cluster. It was grouped into smaller clusters. Analysis of these 5 genomes revealed that, *M. ulcerans* strain Agy99, *M. ulcerans* strain S4018, *M. ulcerans* strain SGL03 and *M. ulcerans* strain CSURQ0185 were clustered together due to some similarities in the frequency of CYP subfamilies. *M. ulcerans* strain Harvey differed from other 4 genomes within the second major cluster in the frequency of 8 CYP subfamilies viz. CYP36A, CYP138A, CYP138A4P, CYP143-144A, CYP188-189A and CYP274A. As far as similarities are concerned, it was noticed that the frequency of CYP135B, CYP138B, CYP147G, CYP183B, CYP190A, CYP271A CYP276A and CYP279A were 0 in all the 5 genomes within this cluster. Additionally, a frequency of 1 was detected in 16 CYP subfamilies viz. CYP51B, CYP105Q, CYP108B, CYP123-124A, CYP126A, CYP130A, CYP137A, CYP142A, CYP150A, CYP185A, CYP189A6P, CYP191A, CYP226B, CYP269A1P and CYP278A in each of these genomes. Furthermore, each of them had a frequency of 2 in 4 CYP subfamilies viz. CYP125A6P, CYP164A3P, CYP187A and CYP268A2P. Examination of the 7 genomes from two major clusters revealed similarity in 10 CYP subfamilies. These subfamilies viz., CYP51B, CYP105Q, CYP124A, CYP126A, CYP130A, CYP137A, CYP142A, CYP185A, CYP226B and CYP278A had a frequency of 1.

CYP co-occurrence network in Mycobacterium ulcerans

Figs. 5 and 6 portrayed the CYP family and subfamily co-occurrence networks from the MU-7 dataset. Based on their corresponding occurrence in *M. ulcerans* genomes, the mutual connectivity of the identified CYPs were

demonstrated by these co-occurrence networks. It was observed from Fig. 5. that a considerable number CYP families formed big nodes and were strongly clustered together. They had a high connection degree (greater number of edges connected to the nodes) with thick width. The degree of the nodes was reflected in the size. These included the principal CYP families, CYP138, CYP140, CYP189, CYP125, conserved CYP51 (Parvez et al. 2016) in addition to a host of others. On the other hand, less dominant CYP families viz. CYP135, CYP147, CYP183, CYP190, CYP271, CYP276 and CYP279 showed reduced compactness in the connection with thin width. The co-occurrence network of the CYP subfamilies (Fig. 6) also displayed a condensed cluster of a number of CYP subfamilies with big nodes, thick width and high connection degree. This cluster housed key CYP subfamilies viz. CYP125A6P, CYP164A3P, conserved CYP51B as well as others. However, the least dominant subfamilies displayed reduced compactness, connectivity and thin width. These included CYP135B, CYP138B, CYP183B, CYP190A, CYP271A, CYP276A and CYP279A.

CYP cloud of Mycobacterium ulcerans

Fig. 7(a-b) graphically represented the CYP cloud displaying CYP frequency at the family and subfamily level from the MU-7 dataset. The principal families and subfamilies depicted in bigger font have been highlighted earlier (Table 1). Moderate and least dominant families and subfamilies were displayed in medium and smaller fonts respectively. A deeper probe revealed that *M. ulcerans* strain Harvey had the highest concentration of the principal CYP families

CYP138, CYP140 and CYP189 with a copy number of 5, 3 and 3. A copy number of 2 was observed in CYP125 and CYP268 for *M. ulcerans* strain Harvey, *M. ulcerans* strain CSURQ0185, *M. ulcerans* strain SGL03, *M. ulcerans* strain S4018 and *M. ulcerans* strain Agy99. CYP279 was the least dominant family having a solitary copy number in *M. ulcerans* strain P7741. Other less dominant families, CYP147, CYP183, CYP190, CYP271 and CYP276 with a frequency of 1 were present only in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741.

M. ulcerans strain Harvey and *M. ulcerans* strain Agy99 had a copy number of 3 in the principal subfamilies, CYP140A and CYP138A. Other principal subfamilies, CYP125A6P and CYP164A3P had a copy number of 2 in *M. ulcerans* strain Harvey, *M. ulcerans* strain CSURQ0185, *M. ulcerans* strain SGL03, *M. ulcerans* strain S4018, *M. ulcerans* strain Agy99 and *M. ulcerans* strain P7741. The least dominant subfamilies, CYP138B and CYP279A with solitary counts were detected in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741. Additionally, other less dominant subfamilies, CYP183B, CYP190A, CYP271A and CYP276A with an incidence of 1 were found in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741.

Discussion

In this work, cytochrome P450 monooxygenases (CYPs) were identified and categorized from various *M. ulcerans* strains. Using this information, a number of downstream investigations at the family and subfamily level concerning the pan-CYPome, CYP centric clustering, interrelationship networks, frequency and variations of CYPs were carried out. The results were mutually supportive. The outcome of these analyses had underlined several important characteristics of *M. ulcerans* CYPs that could pave the way for further studies.

M. ulcerans is a human pathogen. Examination of individual *M. ulcerans* strains revealed that they had low concentration (ranging from 34-45) of CYP families. The outcome agrees with the fact that mycobacterial human pathogens tend to possess lower CYPs owing to their adaptation to human pathogenic lifestyle where they must survive by fighting the host's immunity and exploit carbon sources (Parvez et al. 2016; Senate et al. 2019). *M. ulcerans* strains showed difference between them regarding the number of CYP families and subfamilies. This had something to do with the size of the genomes, especially in *M. ulcerans* strain Harvey. Here greater number of coding sequences corresponded to increased CYPs while smaller number coding sequences in *M. ulcerans* strain SGL03 resulted in lower CYPs. This is in line with the observations from other mycobacterial species (Parvez et al. 2016). Again, *M. ulcerans* strain Harvey was isolated in USA (https://patricbrc.org/view/Taxonomy/2#view_tab=genomes&filter=and(keyword(Mycobacterium),keyword(ulcerans)) and its' difference from other strains regarding number of CYPs, clustering of families and subfamilies in the CYP heatmaps (Figs. 3-4) may be attributed to geographical location. Previous works have specified the role of geographical location in variation among bacterial strains (Sen et al. 2008; Sur et al. 2008).

The incidence of comparatively higher number of members in CYP families, CYP138, CYP140, CYP189 and CYP125 designated that they were flourishing. This phenomenon has also been reported in other organisms (Syed et al. 2014; Parvez et al. 2016). These flourishing families played a crucial part in *M. ulcerans* physiology. CYP138, CYP140, CYP189 and CYP125 are known to be associated with heme binding (https://www.uniprot.org/uniprot/P9WPM3), mycolactone synthesis (Mve-Obiang et al. 2005), monooxygenase activity (Parvez et al. 2016) and steroid hydroxylation (Mc Lean et al. 2009). Given
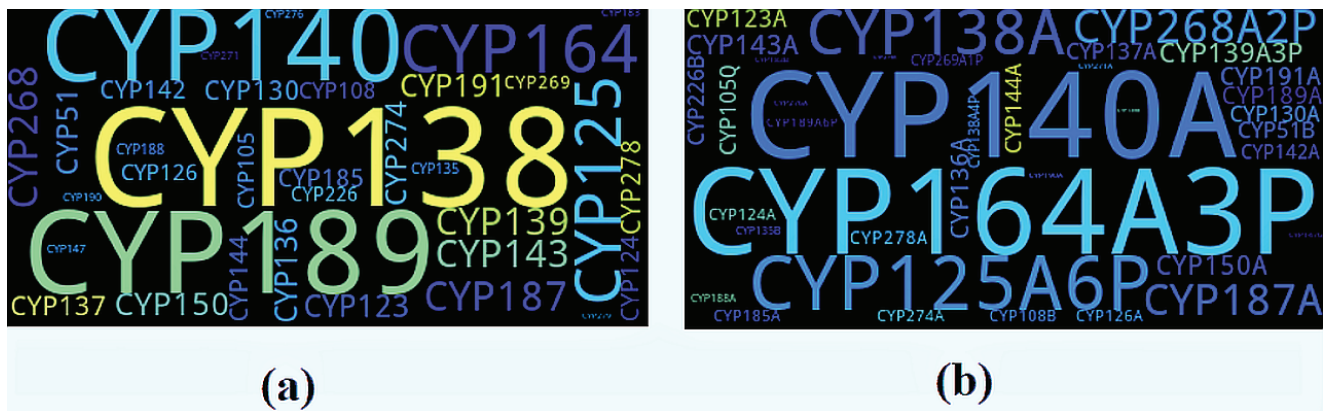

Figure 7. CYP clouds of the (a) family and (b) subfamily from MU-7 dataset.

that *M. ulcerans* secrete mycolactone toxin during Buruli ulcer, the dominance of CYP140 point to its influence on the disease. This is strengthened by its presence in almost all the *M. ulcerans* strains. Occurrence of 20 CYP families among all the 7 *M. ulcerans* strains, portrayed that they were conserved within these strains and suggested their importance. A number of them like CYP51, CYP105, CYP124, CYP125, CYP142, CYP150, CYP185 function in lipid metabolism, steroid metabolism (Kelly and Kelly 2013), diterpenoid oxidation (Moody and Loveridge 2014), steroids, fatty acid hydroxylation (Johnston et al. 2009; Johnston et al. 2010; Mc Lean et al. 2009) and polycyclic aromatic hydrocarbon hydroxylation (Brezna et al. 2006). Earlier works have specified that these families assisted *Mycobacterium* in assimilating host compounds (Senate et al. 2019). Presence of CYP279 family only in *M. ulcerans* strain P7741 indicated that this can act as a diagnostic marker for this strain. Nevertheless, incidence of CYP135, CYP147, CYP183, CYP190, CYP271 and CYP276 only in *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 and *M. ulcerans* strain P7741 revealed that these families may be improved diagnostic markers than CYP279 owing to occurrence in two strains. Their existence stated their importance in these *M. ulcerans* strains. Most of the least dominant CYP families limited to few *M. ulcerans* strains were orphans (Parvez et al. 2016). This is line with earlier observations on P450s from other mycobacteria (Parvez et al. 2016). The outcome from CYP subfamilies were consistent with that of the families. Analysis of the CYP families and subfamilies from *M. ulcerans* indicated that they were designed for utilization of lipids.

The pan-CYPome family and subfamily profile from *M. ulcerans* indicated accumulation of new CYPs with each newly added genome, and the ability to continuously extent their total set of CYPs. It is probable that by acquisition of CYPs and diversification, *M. ulcerans* is evolving. This may have consequences in terms of the diversity and lifestyle of *M. ulcerans* strains. Considerable proportion of core CYP families and subfamilies in the pan-CYPome, is probably an outcome of a number of strains, family and subfamily-specific CYPs, reflecting the dynamics of *M. ulcerans*. Analysis of the core CYP families and subfamilies provided clues on the role of common CYPs essential for the lifestyle of *M. ulcerans*. As mentioned earlier most of them are flourishing and are involved in lipid, steroid metabolism, heme binding etc. These are vital for the survival of the species.

CYP centric clustering pattern of the families and subfamilies revealed that despite isolation from different geographical locations, *M. ulcerans* subsp. *shinshuense* strain ATCC 33728 (isolated from Japan) and *M. ulcerans* strain P7741 (isolated from French Guyana) (https://www.patricbrc.org/view/GenomeList/?and(keyword(M

ycobacterium),keyword(ulcerans))#view_tab=genomes), were grouped together in the first major cluster owing to similarities in the CYP frequency of 22 families and 25 subfamilies. The CYP centric clustering pattern in the second major cluster indicated that, *M. ulcerans* strain Harvey (isolated from USA), *M. ulcerans* Agy99 (isolated from Ghana), *M. ulcerans* strain S4018 (isolated from Benin), *M. ulcerans* strain SGL03 (isolated from Democratic Republic of Congo) and *M. ulcerans* strain CSURQ0185 (isolated from Ivory Coast) (https://www.patricbrc.org/view/GenomeList/?and(keyword(Mycobacterium),keyword(ulcerans))#view_tab=genomes), had resemblances in the CYP frequency of 25 families and 28 subfamilies. While the four African strains showed some parallels with each other, *M. ulcerans* strain Harvey (isolated from USA) differed from these in the CYP frequency of 8 families and subfamilies. Interestingly, this included the flourishing CYP families and subfamilies like CYP140, CYP138, CYP138A respectively. This difference exhibited by *M. ulcerans* strain Harvey was due to higher number of CYPs in the families and subfamilies compared to others. It is possible that higher number offered advantage for better adaptation to the host's environment and needs to be corroborated by future studies. Again, occurrence of 10 CYP families and subfamilies with a frequency of 1 among seven *M. ulcerans* strains highlighted the resemblances among them.

Findings from CYP family and subfamily co-occurrence network were in agreement with their frequencies. Sizeable number of CYP families including CYP138, CYP140, CYP189, CYP125 with big nodes, robust clustering, high connection degree and thick width were indicative of their high frequency and co-occurrence in *M. ulcerans* genomes. Again, strong clustering of majority of CYP families specified their mutual associations regarding CYP dynamics in *M. ulcerans* genomes. In contrast, low connection degree and thin width of CYP135, CYP147, CYP183, CYP190, CYP271, CYP276 and CYP279 signified less frequency and dissociation. The result of CYP subfamily co-occurrence network were supportive of the CYP family counterpart.

The observations from CYP family and subfamily clouds were in line with their frequency and variation in *M. ulcerans* strains. The incidence of high copy number of the flourishing families (CYP138, CYP140, CYP189, CYP125, CYP268) and subfamilies (CYP141A, CYP138A, CYP125A6P, CYP164A3P) indicated that they played a vital role in the metabolism and lifestyle of *M. ulcerans* strain Harvey, *M. ulcerans* strain CSURQ0185, *M. ulcerans* strain SGL03, *M. ulcerans* strain S4018 and *M. ulcerans* Agy99. Comparatively higher concentration of these in *M. ulcerans* strain Harvey indicated that this strain showed variation from the rest. This further strengthened the

hypothesis that comparatively larger genome manifested by higher number of coding sequences (CDS), and different geographical location of *M. ulcerans* strain Harvey may be the reason behind such variation.

Conclusions

This study using CYPminer pipeline provided relative understandings of CYPs from *Mycobacterium ulcerans* strains at the genome level. An attempt was made to get a picture of the CYPs from *M. ulcerans* strains since they are known to be involved in myriad metabolic processes. A total of 261 CYPs were identified. While some families and subfamilies flourished others had solitary presence. Several them were conserved across strains. A substantial number of them had strong mutually supportive relationships. Low number of CYP families coupled with the role of flourishing, conserved families and subfamilies in mycolactone synthesis, steroid hydroxylation, lipid metabolism, etc. indicated their physiological role in the lifestyle of *M. ulcerans*. The *M. ulcerans* strains showed difference and similarities between them in the context of incidence, frequency, copy number and clustering of CYP families and subfamilies. The outcome of this work enhanced our knowledge regarding the CYP dynamics in *M. ulcerans*. Future work should focus on exploring the role of *M. ulcerans* CYPs in secondary metabolite synthesis followed by assessment of their biosynthetic gene clusters since they may hold potential for the biopharmaceutical industry.

Acknowledgements

References

Ahmed I, Tiberi S, Farooqi J, Jabeen K, Yeboah-Manu D, Migliori GB, Hasan R (2020) Non-tuberculous mycobacterial infections-A neglected and emerging problem. IJID 92S:S46-S50.

Brezna B, Kweon O, Stingley RL, Freeman JP, Khan AA, Polek B, Jones RC, Cerniglia CE (2006) Molecular characterization of cytochrome P450 genes in the polycyclic aromatic hydrocarbon degrading *Mycobacterium vanbaalenii* PYR-1. Appl Microbiol Biotechnol 71:522-532.

Coon MJ (2005) Cytochrome P450: nature's most versatile biological catalyst. Annu Rev Pharmacol Toxicol 45:1-25.

Darban-Sarokhalil D, Fooladi AA, Bameri Z, Nasiri MJ, Feizabadi MM (2011) Cytochrome CYP141: A new target for direct detection of *Mycobacterium tuberculosis* from clinical specimens. Acta Microbiol Immunol Hung 58:211-217.

de Souza DK, Quaye C, Mosi L, Addo P, Boakye D (2012) A quick and cost effective method for the diagnosis of *Mycobacterium ulcerans* infection. BMC Infect Dis 12:8.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26:2460-2461.

Furge LL, Guengerich FP (2006) Cytochrome P450 enzymes in drug metabolism and chemical toxicology: an introduction. Biochem Mol Biol Educ 34:66-74.

Johnston JB, Kells PM, Podust LM, Ortiz de Montellano PR (2009) Biochemical and structural characterization of CYP124: a methyl-branched lipid ω-hydroxylase from *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA 106:20687-20692.

Johnston JB, Ouellet H, Ortiz de Montellano PR (2010) Functional redundancy of steroid C-26-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. J Biol Chem 285:36352-36360.

Kelly SL, Kelly DE (2013) Microbial cytochrome P450: Biodiversity and biotechnology, where do cytochrome P450 come from, what do they do and what can they do for us? Phil Trans R Soc B Biol Sci 368:20120476.

Khumalo JM, Nzuza N, Padayachee T, Chen W, Yu JH, Nelson DR, Syed K (2020) Comprehensive analyses of cytochrome P450 monooxygenases and secondary metabolite biosynthetic gene clusters in cyanobacteria. Int J Mol Sci 21:656.

Kweon K, Kim SJ, Kim JH, Nho SW, Bae D, Chon J, Hart M, Baek DH, Kim YC, Wang W, Kim SK, Sutherland JB, Cerniglia CE (2020) CYPminer: an automated cytochrome P450 identification, classification, and data analysis tool for genome data sets across kingdoms. BMC Bioinformatics 21:160.

Lau CKI, Feyereisen R, Nelson DR, Bell SG (2019) Analysis and preliminary characterisation of the cytochrome P450 monooxygenases from *Frankia* sp. EuI1c (*Frankia inefficax* sp.). Arch Biochem Biophys 669:11-21.

Liu Y, Gao Y, Liu J, Tan Y, Liu Z, Chhotaray C, Jiang H, Lu Z, Chiwala G, Wang S, Makafe G, Islam MM, Hameed HMA, Cai X, Wang C, Li X, Tan S, Zhang T (2019) The compound TB47 is highly bactericidal against *Mycobacterium ulcerans* in a Buruli ulcer mouse model. Nat Commun 10:524.

Luo Y, Degang Y, Ohtsuka M, Ishido Y, Ishii N, Suzuki K (2015) Detection of *Mycobacterium ulcerans* subsp. *shinshuense* DNA from a water channel in familial Buruli ulcer cases in Japan. Future Microbiol 10:461-469.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales

NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH (2009) CDD: specific functional annotation with the conserved domain database. Nucleic Acids Res 37:D205-210.

McLean KJ, Carroll P, Lewis DG, Dunford AJ, Seward HE, Neeli R, Cheesman NR, Marsollier L, Douglas P, Smith WE, Rosenkrands I, Cole ST, Leys D, Parish T, Munro AW (2008) Characterization of active site structure in CYP121. A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J Biol Chem 283:33406-33416.

McLean KJ, Lafite P, Levy C, Cheesman MR, Mast N, Pikuleva IA, Leys D, Munro AW (2009) The structure of *Mycobacterium tuberculosis* CYP125: molecular basis for cholesterol binding in a P450 needed for host infection. J Biol Chem 284:35524-35533.

McLean KJ, Leys D, Munro AW (2015) Microbial cytochrome P450s. In Oritz de Montellano PR, Ed., Cytochrome P450: Structure, Mechanism, and Biochemistry. 4th Edn. Springer International Publishing, 261-407.

Merritt RW, Walker ED, Small PLC, Johnson PDR, Benbow ME, Boakye DE (2010) Ecology and transmission of Buruli ulcer disease: a systematic review. PLoS Negl Trop Dis 4:e911.

Moody SC, Loveridge EJ (2014) CYP105-diverse structures, functions and roles in an intriguing family of enzymes in *Streptomyces*. J Appl Microbiol 117:1549-1563.

Mve-Obiang A, Lee RE, Umstot ES, Trott KA, Grammer TC, Parker JM, Ranger BS, Grainger R, Mahrous EA, Small PLC (2005) A newly discovered mycobacterial pathogen isolated from laboratory colonies of *Xenopus* species with lethal infections produces a novel form of mycolactone, the *Mycobacterium ulcerans* macrolide toxin. Infect Immun 73:3307-3312.

Mthetwa BC, Chen W, Ngwenya ML, Kappo AP, Syed PR, Kapoormath R, Yu JH, Nelson DR, Syed K (2018) Comparative analyses of cytochrome P450s and those associated with secondary metabolism in *Bacillus* species. Int J Mol Sci 19:3623.

Nebert DW, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Kemper B, Levin W, Phillips IR, Sato R, Waterman MR (1987) The P450 gene superfamily: recommended nomenclature. DNA 6:1-11.

Nelson DR (1999) Cytochrome P450 and the individuality of species. Arch Biochem Biophys 369:1-10.

Nelson DR (2013) A world of cytochrome P450s. Phil Trans R Soc B Biol Sci 368:20120430.

O'Brien DP, Murrie A, Meggesey P, Priestley J, Rajcoomar A, Athan E (2019) Spontaneous healing of *Mycobacterium ulcerans* disease in Australian patients. PLoS Negl Trop Dis 13:e0007178.

Ohtsuka M, Kikuchi N, Yamamoto T, Suzutani T, Nakanaga K, Suzuki K, Ishii N (2013) Buruli ulcer caused by *Mycobacterium ulcerans* subsp *shinshuense*: a rare case of familial concurrent occurrence and detection of insertion sequence 2404 in Japan. JAMA Dermatol 150:64-67.

Parvez M, Qhanya L, Mthakathi N, Kgoseimang IKR, Bamal HD, Pagadala NK, Xie T, Yang H, Chen H, Theron CW, Monyaki R, Raslemane SC, Salawe V, Mongale BL, Matowane RG, Abdalla SMH, Booi WI, van Wyk M, Olivier D, Boucher CE, Nelsen DR, Tuszynski JA, Blackburn JM, Yu JH, Mashele SS, Chen W, Syed K (2016) Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. Sci Rep 6:33099.

Sen A, Sur S, Bothra AK, Benson DR, Normand P, Tisa LS (2008) The implication of life style on codon usage patterns and predicted highly expressed genes for three *Frankia* genomes. Antonie van Leeuwenhoek 93:335-346.

Senate LM, Tjatji MP, Pillay K, Chen W, Zondo NM, Syed PR, Mnguni FC, Chiliza ZE, Bamal HD, Karpoormath R, Khoza T, Mashele SS, Blackburn JM, Yu JH, Nelson DR, Syed K (2019) Similarities, variations, and evolution of cytochrome P450s in *Streptomyces* versus *Mycobacterium*. Sci Rep 9:3962.

Sur S, Bothra AK, Bajwa M, Tisa LS, Sen A (2008) *In silico* analysis of *Chlorobium* genomes divulge insights into the lifestyle of the bacteria. Res J Microbiol 3:600-613.

Syed K, Shale K, Pagadala NS, Tuszynski J (2014) Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. PLoS ONE 9:e86683.

Tai AYC, Athan E, Friedman ND, Hughes A, Walton A, O'Brien DP (2018) Increased severity and spread of *Mycobacterium ulcerans*, Southeastern Australia. Emerg Infect Dis 24:58-64.

van der Werf TS, van der Graaf WT, Tappero JW, Asiedu K (1999) *Mycobacterium ulcerans* infection. Lancet 354:1013-1018.

Yoshida M, Nakanaga K, Ogura Y, Toyoda A, Ooka T, Kazumi Y, Mitarai S, Ishii N, Hayashi T, Hoshino Y (2016) Complete genome sequence of *Mycobacterium ulcerans* subsp. *shinshuense*. Genome Announc 4:e01050-16.

Yoshida Y, Aoyama Y, Noshiro M, Gotoh O (2000) Sterol 14-demethylase P450 (CYP51) provides a breakthrough for the discussion on the evolution of cytochrome P450 gene superfamily. Biochem Biophys Res Commun 273:799-804.

Yotsu RR, Murase C, Sugawara M, Suzuki K, Nakanaga K, Ishii N, Asiedu K (2015) Revisiting Buruli ulcer. J Dermatol 42:1033-1041.

Zingue D, Bouam A, Tian RBD, Drancourt M (2018) Bu-

ruli ulcer, a prototype for ecosystem-related infection, caused by *Mycobacterium ulcerans*. Clin Microbiol Rev 31:e00045-17.