

RESEARCH ARTICLE

Comparative genome analyses of mycobacteria give better insights into their evolution

Wei Yee Wee¹, Avirup Dutta¹, Siew Woh Choo^{2*}

1 Genome Informatics Research Laboratory, Centre for Research in Biotechnology and Agriculture (CEBAR), High Impact Research Building, University of Malaya, Kuala Lumpur, Malaysia, **2** Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou Dushu Lake Science and Education Innovation District, Suzhou Industrial Park, Suzhou, P. R. China

* lawrence.choo@xjtlu.edu.cn, l.choo@genomesolutions.com.my



OPEN ACCESS

Citation: Wee WY, Dutta A, Choo SW (2017) Comparative genome analyses of mycobacteria give better insights into their evolution. PLoS ONE 12(3): e0172831. <https://doi.org/10.1371/journal.pone.0172831>

Editor: Seyed Ehtesham Hasnain, Indian Institute of Technology Delhi, INDIA

Received: May 22, 2016

Accepted: February 10, 2017

Published: March 14, 2017

Copyright: © 2017 Wee et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information file.

Funding: This project was supported by University of Malaya and Ministry of Education, Malaysia under the High Impact Research (HIR) grant UM.C/625/1/HIR/MOE/CHAN/08 and UMRG grant (grant number: RG541-13HTM) from University of Malaya and Ministry of Education, Malaysia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Mycobacteria a genus of Actinobacteria are widespread in nature ranging from soil-dwelling saprophytes to human and animal pathogens. The rate of growth has been a classifying factor for the *Mycobacterium* spp., dividing them into the rapid growers and the slow growers. Here we have performed a comparative genome study of mycobacterial species in order to get better understanding of their evolution, particularly to understand the distinction between the rapid and slow growers. Our study shows that the slow growers had generally gained and lost more genes compared to the rapid growers. The slow growers might have eventually lost genes (*LivFGMH* operon, *shaACDEFG* genes and MspA porin) that could contribute to the slow growth rate of the slow growers. The genes gained and lost in mycobacteria had eventually helped these bacteria to adapt to different environments and have led to the evolution of the present day rapid and slow growers. Our results also show high number of *Mycobacterium abscessus* specific genes (811 genes) and some of them are associated with the known bacterial quorum sensing genes that might be important for *Mycobacterium abscessus* to adapt and survive in variety of unfavorable environments. *Mycobacterium abscessus* also does not contain genes involved in the bacterial defense system and together with the quorum sensing genes may have contributed to the high gene gain rate of *Mycobacterium abscessus*.

Introduction

Mycobacteria are a genus of Actinobacteria of the *Mycobacteriaceae* family which are obligate aerobic, immobile, acid-fast Gram-positive bacteria with high G+C content, and are widespread in nature ranging from soil-dwelling saprophytes to pathogens of humans and animals [1,2]. *Mycobacterium* have been classified into two distinct categories depending on their rate of growth; the slow growers which take more than 7 days on subculture to form colonies, and then there are the rapid growers forming colonies within 3–7 days [3]. Most of the major pathogenic Mycobacteria such as *M. tuberculosis* and *M. leprae* responsible for causing serious

Competing interests: The authors have declared that no competing interests exist.

diseases in humans and animals, such as tuberculosis and leprosy respectively, are slow growers. With regard to pathogenicity, some of the non-tuberculous mycobacteria (NTM) have also been shown to cause disease in humans especially in immunocompromised individuals [4]. However not all pathogenic NTMs are slow growers, such as *M. abscessus* a commonly isolated rapidly growing NTM is the third most common cause of lung disease [5]. It is believed that the slow growers have evolved from the rapid growing mycobacteria [6,7]. Interestingly, studies have also shown that a distinction between the rapid and the slow growers is in the presence of ESX-5, which are exclusively present in the slow-growing mycobacteria [8].

Here we present a pan-genome analysis of 28 mycobacterial species (comprised of both rapid and slow growers). This study was performed with the objective to have a better understanding of the distinction between the rapid and the slow growers at the genomic level of the evolution of the mycobacterial genus.

Materials and methods

Pan-genome analysis of *Mycobacterium* genus

28 genomes from different *Mycobacterium* species were downloaded from the NCBI database for comparative and pan-genome analysis (Table 1). For consistency, the 28 genomes were re-annotated using the Rapid Annotation using Subsystem Technology (RAST) pipeline [9]. The predicted protein sequences of the 28 strains from RAST were used pan-genome analysis using PGAP (pan-genome analysis pipeline) [10]. Functional orthologs among the 28 strains were searched using the Gene Family (GF) method [10]. The protein sequences of each strain were mixed together and marked with the strain identifiers. BLASTALL was first performed among the protein sequences with the minimum score value of 50 and e-value of 1e-8 [11,12]. The filtered BLAST results were clustered by MCL algorithm [13]. In order to group the same genes into clusters, the threshold for the global match region was set to a minimum of 50% of the longer gene protein sequence and 50% sequence identity.

The relation between pan-genome size and genome number was determined using the method described by Zhao et al. [14]. According to Tettelin's review on pan-genome research [15], Heaps' Law model is employed to fit the pan-genome size of strains. Thus, pan-genome size model:

$$y = A_1 x^{B_1} + C_1$$

x denoting genome number, y denoting pan-genome size, A_1 , B_1 and C_1 the fitting parameters.

Core genome size model:

$$y = A_2 e^{B_2 x} + C_2$$

x denoting genome number, y denoting core genome size, A_2 , B_2 and C_2 the fitting parameters.

Phylogenetic inferences

The phylogenetic trees of *Mycobacterium* isolates were first constructed using the identified core genes following the approaches described by Rokas and colleagues [16]. The protein sequences of these core genes were concatenated to construct a phylogenetic tree using Maximum Parsimony and 1000 times bootstrap resampling approach. Each of the concatenated sequences was aligned with ClustalW [17].

Table 1. Details of 28 different *Mycobacterium* species used in the analysis.

<i>Mycobacterium</i> species	Strain Name	Genome size (Mbp)	GC %	Contig	ORFs
<i>Mycobacterium africanum</i>	GM041182	4.39	65.6	1	4,369
<i>Mycobacterium avium</i>	K10	4.83	69.3	1	4,643
<i>Mycobacterium chubuense</i>	NBB4	5.58	68.7	1	5,346
<i>Mycobacterium colombiense</i>	CECT3035	5.58	68.1	17	5,341
<i>Mycobacterium fortuitum</i>	DSM46621	6.35	66.2	82	6,135
<i>Mycobacterium gilvum</i>	Spyr1	5.55	67.9	1	5,393
<i>Mycobacterium hassiacum</i>	DSM44199	5.0	69.5	169	4,849
<i>Mycobacterium indicus pranii</i>	MTCC9560	5.59	68	1	5,326
<i>Mycobacterium intracellulare</i>	ATCC13950	5.4	68.1	1	5,148
<i>Mycobacterium kansasii</i>	ATCC12478	6.42	66.1	108	6,007
<i>Mycobacterium leprae</i>	TN	3.27	57.8	1	1,605
<i>Mycobacterium mageritense</i>	JR2009	6.5	66.4	1.031	6,313
<i>Mycobacterium marinum</i>	M	6.64	65.7	1	5,826
<i>Mycobacterium parascrofulaceum</i>	ATCC_BAA614	6.56	68.4	124	6,093
<i>Mycobacterium phlei</i>	RIVM601174	5.68	69.2	102	5,622
<i>Mycobacterium rhodesiae</i>	NBB3	6.42	65.5	1	6,426
<i>Mycobacterium smegmatis</i>	JS623	6.46	65.1	1	6,591
<i>Mycobacterium</i> sp.	JLS	6.05	68.4	1	5,933
<i>Mycobacterium</i> sp.	KMS	5.74	68.4	1	5,605
<i>Mycobacterium</i> sp.	MCS	5.71	68.5	1	5,558
<i>Mycobacterium brisbanense</i>	UM_WWY	7.69	66.4	130	7,590
<i>Mycobacterium thermoresistibile</i>	ATCC19527	4.87	69	56	4,683
<i>Mycobacterium tuberculosis</i>	CCDC5079	4.4	65.6	1	4,553
<i>Mycobacterium ulcerans</i>	Agy99	5.63	65.5	1	5,566
<i>Mycobacterium vaccae</i>	ATCC25954	6.25	68.6	33	5,966
<i>Mycobacterium vanbaalenii</i>	PYR1	6.49	67.8	1	6,291
<i>Mycobacterium abscessus</i>	ATCC19977	5.0	64.1	1	5,004
<i>Mycobacterium xenopi</i>	RIVM700367	4.43	66.1	117	4,427

<https://doi.org/10.1371/journal.pone.0172831.t001>

Gene gain and loss detection

The constructed tree was used as a map for the evolution of each isolate and the insertion and deletion of genes were mapped onto the phylogenies according to the Maximum Likelihood principle. The evolution of the *Mycobacterium* was modelled by the gain and death (GD) and gain and death with the free rates model (GD-FR-ML) where each branch contains separate gene gain and death turnover rate [18].

Identification of bacteria defense system genes

From the RAST genome annotation, bacterial defense system related genes such as the restriction modification (RM) system and the toxin-antitoxin (TA) system were identified in the *Mycobacterium abscessus* ATCC19977, *Mycobacterium tuberculosis* CCDC5079, *Mycobacterium africanum* GM041182 and *Salmonella* Paratyphi A.

The CRISPR in the mentioned bacteria was further predicted using CRISPRfinder [19].

Results and discussion

28 different mycobacterial species comprising 16 rapid growing and 12 slow growing mycobacterial species were used in this study. For consistency, all 28 mycobacterial genomes were

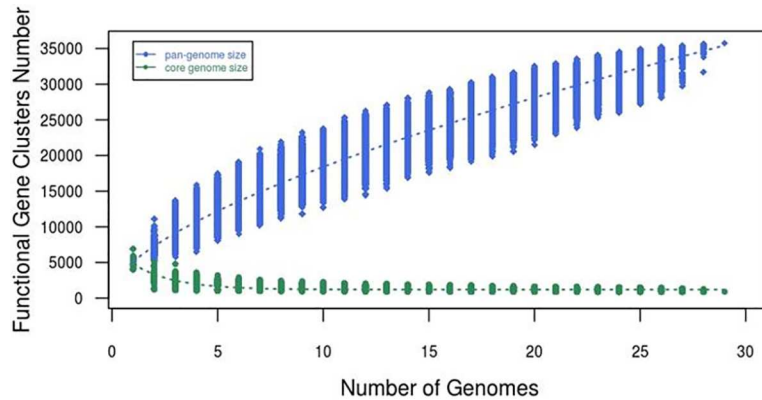


Fig 1. Prediction of *Mycobacterium* pan- and core- genomes size.

<https://doi.org/10.1371/journal.pone.0172831.g001>

annotated using the RAST pipeline. The annotations of each genome are summarized in [Table 1](#).

Overall, the genome size of the 28 mycobacterial strains ranged from 4Mbp to 6Mbp except for *Mycobacterium leprae* TN (around 3Mbp) which might due to genome degradation and *Mycobacterium brisbanense* UM_WWY which was the only mycobacterial genome having a genome size larger than 7Mbp.

Comparative genomic analysis among *Mycobacterium* species

***Mycobacterium* pan-genome analysis.** Using the predicted protein sequences of the 28 annotated *Mycobacterium* genomes as the input, the pan-genome analysis was conducted. Pan-genome analysis showed that the *Mycobacterium* genus had an open pan-genome suggesting that its pan-genome size of *Mycobacterium* could continue to increase when the number of genome added increased according to the following predicted functional model ([Fig 1](#)):

$$y = 3806.24083816781 * x^{0.651} + 1366.97772893166$$

y = Pan-genome size

x = Number of genome

In other words, this genus is expected to continue to gain genes and evolve in the future. This may be also explained by the high number of accessory gene clusters (95%) that we observed in this genus. A total of 35,751 unique gene families were identified from the 28 mycobacterial genomes which comprised of 1,829 (5%) core gene clusters and 33,922 (95%) accessory gene clusters. Of the 33,922 accessory genes families, 13,370 are species-specific genes which accounted around 40% of the accessory genes.

Mycobacterium core-genes derived phylogenetic tree

The 16S rRNA gene has been used for years as the candidate gene for phylogenetic studies for *Mycobacterium* [20]. However, it has been reported that this housekeeping gene might not produce a robust phylogenetic tree for mycobacteria due to the limited interspecies genetic variability in this genus [7]. Hence, the concept of concatenation of multiple housekeeping genes was proposed for generating a more robust phylogenetic tree [21]. To infer the phylogenetic relationship of *Mycobacterium* spp., we constructed a phylogenetic tree of 28 *Mycobacterium* spp. using the concatenated sequence of 531 core proteins with 1:1 orthologs found from the pan-genome analysis. The core genes were concatenated into a single 191,752 nucleotide

alignment (Fig 2). As anticipated, the tree was in concordance with the known relationships for *Mycobacterium* spp. [22].

Interestingly, the generated tree clearly separated the rapid-growing mycobacteria (RGM) and slow-growing mycobacteria (SGM). For instance, the known slow growers, the *Mycobacterium tuberculosis* complex (*Mycobacterium tuberculosis* and *Mycobacterium africanum*), *Mycobacterium leprae*, *Mycobacterium ulcerans*, *Mycobacterium marinum* and the *Mycobacterium avium* complex species (*Mycobacterium avium*, *Mycobacterium colombiense*, *Mycobacterium intracellulare*) were clustered together. For the known rapid-growers such as *Mycobacterium vaccae*, *Mycobacterium vanbaalenii*, *Mycobacterium gilvum*, *Mycobacterium smegmatis*, *Mycobacterium fortuitum* complex members (*Mycobacterium fortuitum*, *Mycobacterium mageritense*, and *Mycobacterium brisbanense*) were clustered into a large clade and clearly separated from other rapid growing mycobacteria, as well as the free-living mycobacterial strains such as *Mycobacterium* sp. JLS, *Mycobacterium* sp. KMS and *Mycobacterium* sp. MCS. The results were also in concordance with the genus tree from a previous study which also generated based on the concatenation of core genes [18].

Interestingly, most of the SGM with few exceptions, appeared to contain smaller genome size compared to the RGM. Most of the SGM genome size was lesser than 6Mbp except for

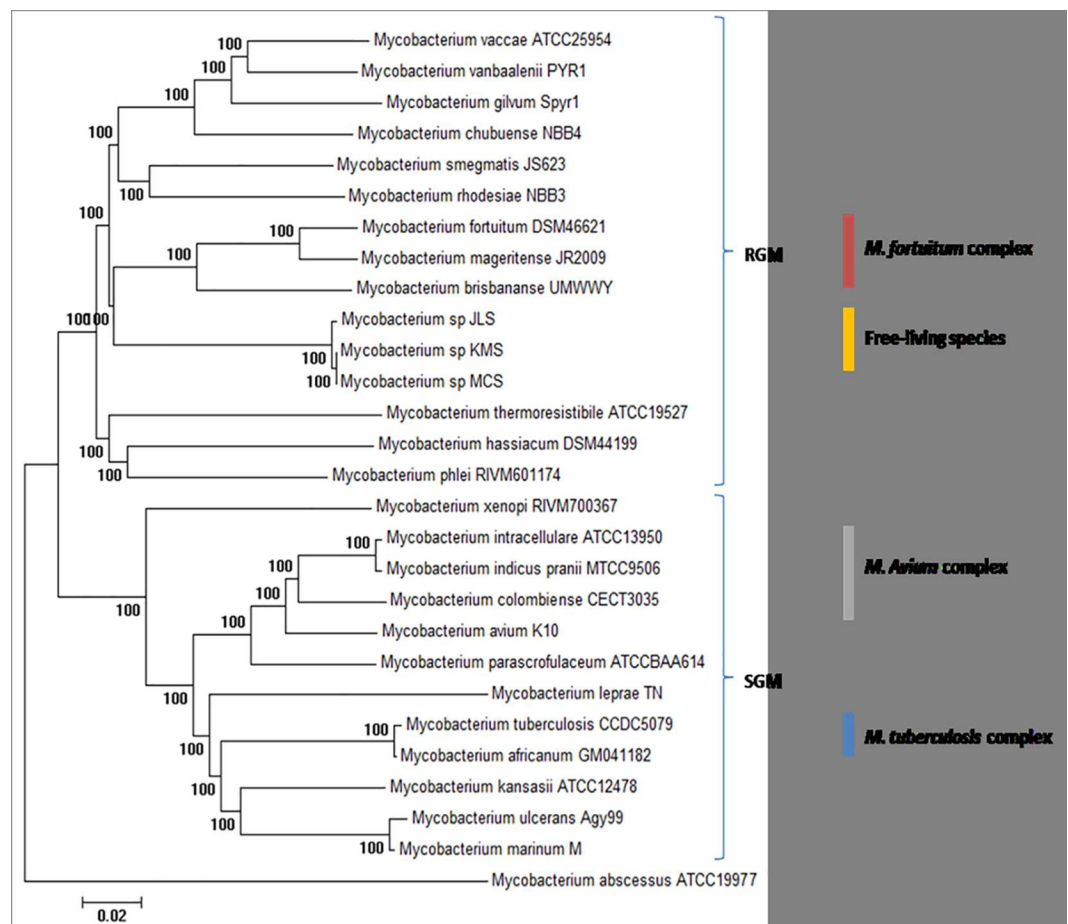


Fig 2. Phylogenetic tree of *Mycobacterium* species constructed using 531 core-concatenated genes. Bootstrap numbers were generated in 1,000 runs and nodes with bootstrap support values of 100 were indicated.

<https://doi.org/10.1371/journal.pone.0172831.g002>

Mycobacterium kansasii which has a genome size of 6.42Mbp. On the other hand, RGM has genome size larger than 5.6Mbp except for *Mycobacterium gilvum*, *Mycobacterium chubuense*, *Mycobacterium hassiacum*, *Mycobacterium thermoresistibile* and *Mycobacterium abscessus*.

Besides, we also noticed high number of species-specific genes in the *Mycobacterium* genus which accounted for around 40% of the total pan-genome. The high number of species-specific genes indicated that horizontal gene transfer had played a major part in the evolution of *Mycobacterium* genus as many non-mycobacterial genes were brought into the *Mycobacterium* genomes [16]. *Mycobacterium brisbananse* UM_WWY contained the highest species-specific genes followed by *Mycobacterium abscessus* and *Mycobacterium smegmatis* while *Mycobacterium* sp. KMS and *Mycobacterium* sp. MCS had the lowest species-specific gene clusters.

From the result, we also observed that *Mycobacterium brisbananse* UM_WWY contains the highest species-specific gene clusters (1,631), followed by *Mycobacterium abscessus* 1,423 genes, *Mycobacterium smegmatis* 923 genes and *Mycobacterium kansasii* 683 genes. While *Mycobacterium* sp. KMS (25 genes) and *Mycobacterium* sp. MCS (26 genes) which contains the lowest species specific gene clusters. The well-known *Mycobacterium* species—*Mycobacterium tuberculosis* also contains a low number of species-specific gene clusters of 93 genes. However, the result here may not imply the true number of species-specific genes as only one strain from each species was included in the analysis. Minor part of the genes may be strain-specific gene rather than the species-specific genes. The variation of bacteria genome size may underlie many important morphological, physiological, and behavioral differences between species and contribute much of the genetic and genomic diversity observed in nature [19,23–26].

The highest number of species-specific genes in *Mycobacterium brisbananse* UM_WWY was probably due to the high genome size of 7.69Mbp, the largest genome size among the *Mycobacterium* species. *Mycobacterium smegmatis* and *Mycobacterium kansasii* also contained genome sizes larger than 6Mbp. However, *Mycobacterium abscessus* having the second highest species-specific genes has a genome size of around 5Mbp only.

Mycobacterium complexes and their evolution

From phylogenetic analysis, we observed that a few known *Mycobacterium* complexes such as *Mycobacterium tuberculosis* complex, *Mycobacterium avium* complex and *Mycobacterium fortuitum* complex were clearly grouped in the *Mycobacterium* genus tree and each complex had its own characteristics. The recombination rate and number of genes gained and lost has been shown in Fig 3. To investigate the recombination rate at these complexes, we examined the gene loss and gain events at the branch 21 for *Mycobacterium fortuitum* complex, branch 37 for *Mycobacterium avium* complex and the branch 47 for the *Mycobacterium tuberculosis* complex.

During the evolution of the *Mycobacterium fortuitum* complex, 139 genes were gained in this complex. For the *Mycobacterium avium* complex, 70 genes were gained, whereas 496 genes were gained (as well as 16 genes were lost) in the *Mycobacterium tuberculosis* complex. Interestingly, of the 496 gene gains in the *Mycobacterium tuberculosis* complex, we have found multiple genes involved in toxin-antitoxin (TA) systems. For instance, there was a pair of YdcD and YdcE toxin-antitoxin, Doc and Phd toxin-antitoxin and 10 toxins VapC with PIN domain. TA system is able to facilitate the bacterial adjustment in response to the environmental changes such as less stable conditions which may cause growth inhibition and eventually cell death. When the conditions become more favorable and stable, reactivation of the TA system may facilitate the bacterial survival [27]. Furthermore, the TA systems are also involved in defense mechanism such as anti-addiction against similar plasmid-borne activation, stabilization of genomic region and defense against the phage infection. Therefore, the gain of these

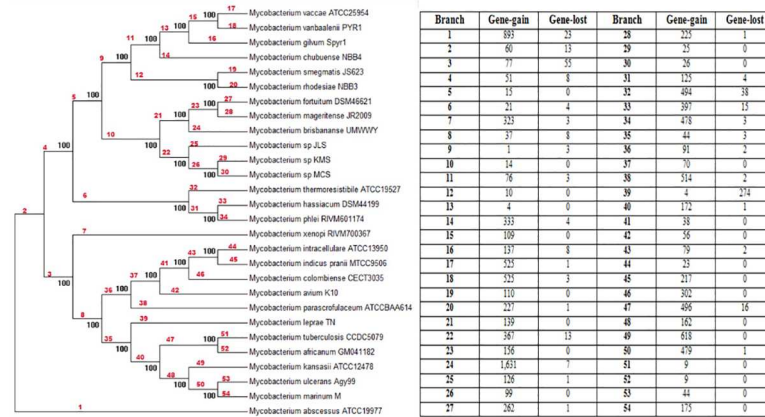


Fig 3. The recombination rate and number of gene gain and lost during the evolution of the *Mycobacterium* species.

<https://doi.org/10.1371/journal.pone.0172831.g003>

defense systems in the *Mycobacterium tuberculosis* complex may suggest that *Mycobacterium tuberculosis* complex members could have more conserved genomic structure and probably are more resistant to the invasion of foreign DNA, for example, through the horizontal gene transfer (HGT). This may be further supported by the observations that the low number of genes gained at the branches 51 and 52 (which are the genome recombination rate) for *Mycobacterium tuberculosis* and *Mycobacterium africanum*. These findings might also explain why the genome size of the *Mycobacterium tuberculosis* complex members are generally the smaller (except for *Mycobacterium leprae* that might have undergone genome degradation and *Mycobacterium vaccae* ATCC25954 which is not a complete genome) and all below 4.5Mbp compared to other known *Mycobacterium* spp. that are not in the complex.

Rapid and slow growing mycobacteria

From the generated core-genes phylogenetic tree, the mycobacterial species were clearly separated into two major clusters: rapid growers and slow growers except for *Mycobacterium abscessus* a rapid grower but emerged out as an outgroup. This was probably due to the high evolutionary rate in *Mycobacterium abscessus* which has already been shown to possess a high potential of gaining genes from the pan-genome analysis.

Mycobacterium genus may be differentiated into two classes at branch 3. Thus, branch 3 which is the recombination level for the slow growers may be responsible for the reason behind the differentiation of these 2 *Mycobacterium* classes. The results showed that during the mycobacterial evolution, all the slow growers had gained 77 genes and lost 55 genes. While on the other hand, the rapid growers gained 51 genes and lost 8 genes. According to Devulder, current mycobacterial species might have originated from ancestral rapid growing mycobacteria [7]. We believe that the slow-growers might have gained or losses some of the genes in order to evolve into the current SGM.

Interestingly, among the 55 genes lost in the slow growing mycobacteria, we found genes responsible for the access to extracellular nutrients which are related to the growth rate of mycobacteria. For instance, we found the slow-growers had lost a cluster of genes, the *LivFGMH* operon which encodes for the branched chain amino-acid transport ATP-binding LIV proteins. The LIV transport system has been reported to allow the transportation of leucine, isoleucine and valine into the bacteria which can enhance the growth of *Bacillus thuringiensis* [28]. The loss of these genes in the SGM might slow down the growth of these bacteria.

Besides, there was the cluster with 6 genes (*shaACDEFG*) found to be present in all the rapid growers but missing from the SGM species. They encode for a multi subunit transport system named as Sha, a Na^+/H^+ antiporter. The Sha system is a complex with cluster of 6 or 7 genes and is important for the homeostasis of Na^+ and H^+ under extreme condition [29]. Although, currently there are no evidences to indicate that the Sha system is responsible for the fast growth rate of the rapid growing *Mycobacterium*, but the Sha system allows fast growing *Mycobacterium* to adapt to different environmental changes and conditions which maybe the main reason that allowed rapid growers to have fast growth rate.

Besides that, the slow growers also lost a few genes which encode for the MspA porin protein. It has been reported that the low permeability of porin in slow growers may limit the efficiency of the hydrophilic drug in the bacteria and also growth rate of mycobacterium due to limited uptake of the polar nutrients [30]. Mailaender and colleagues have shown that the MspA porin is able to accelerate the growth of RGM [31]. The *mspA* gene has been extracted from *Mycobacterium smegmatis* and inserted into *Mycobacterium bovis* BCG and transformed with the expression vector. Their results showed that the *mspA*-expressing *Mycobacterium bovis* BCG strain grew significantly faster than the normal *Mycobacterium bovis* strain. The generation times were 25 hours for the *Mycobacterium bovis* BCG strain but 27 hours for the wild type strain [31]. Therefore, the loss of some genes in slow growers throughout the evolutionary time might be a main reason why these bacteria have become SGM.

Another interesting observation was, the slow growers showed the presence of components of Type VII secretion system which were absent in the rapid growers. All the slow growers had Type VII secretion AAA-ATPase EccA, Type VII secretion protein EccE and ESX-3 secretion system protein EccE3. It has been reported that type VII secretion systems have been used by *Mycobacteria* to secrete proteins across their complex cell envelope, while some of them have been shown to be essential for mycobacterial virulence and/or viability in pathogenic mycobacteria [6]. Recent studies have also shown that plasmids carrying ESX have played an important part serving as accelerators of adaptation and biodiversity with probable impact on the emergence of mycobacterial pathogenicity [32].

Mycobacterium abscessus-specific genes

As mentioned earlier, the high number of species-specific genes did not necessarily indicate a larger genome size, as in the case of *Mycobacterium abscessus* which had the second highest species-specific genes but with a genome size of around 5Mbp only. The high number of species-specific genes for *Mycobacterium abscessus* led us to further study the function of the *Mycobacterium abscessus*-species specific genes and its defense system which may provide some interesting unique features of *Mycobacterium abscessus*.

To study the genetic differences between *Mycobacterium abscessus* and other mycobacterial species, we further reduced the species-specific number by comparing the current set of 1,423 genes with the genes from other 41 *Mycobacterium abscessus* strains. By comparing more *Mycobacterium abscessus* genomes, we were able to filter out genes which may be strain-specific gene rather than the species-specific genes. Results showed that 811 genes families were present in *Mycobacterium abscessus*, but not in other mycobacterial species. To examine the functions of *Mycobacterium abscessus*-specific genes, we performed a homolog search on the COG database on all the 811 genes (Fig 4). The *Mycobacterium abscessus*-specific genes showed the highest distribution in the general function classes (95 genes) followed by transcription (74 genes) and function unknown (48 genes). The function of the 421 *Mycobacterium abscessus*-specific genes (52%) remains unknown as they had been annotated as hypothetical proteins. Thus, more studies are needed to investigate the functions of these genes. Under the

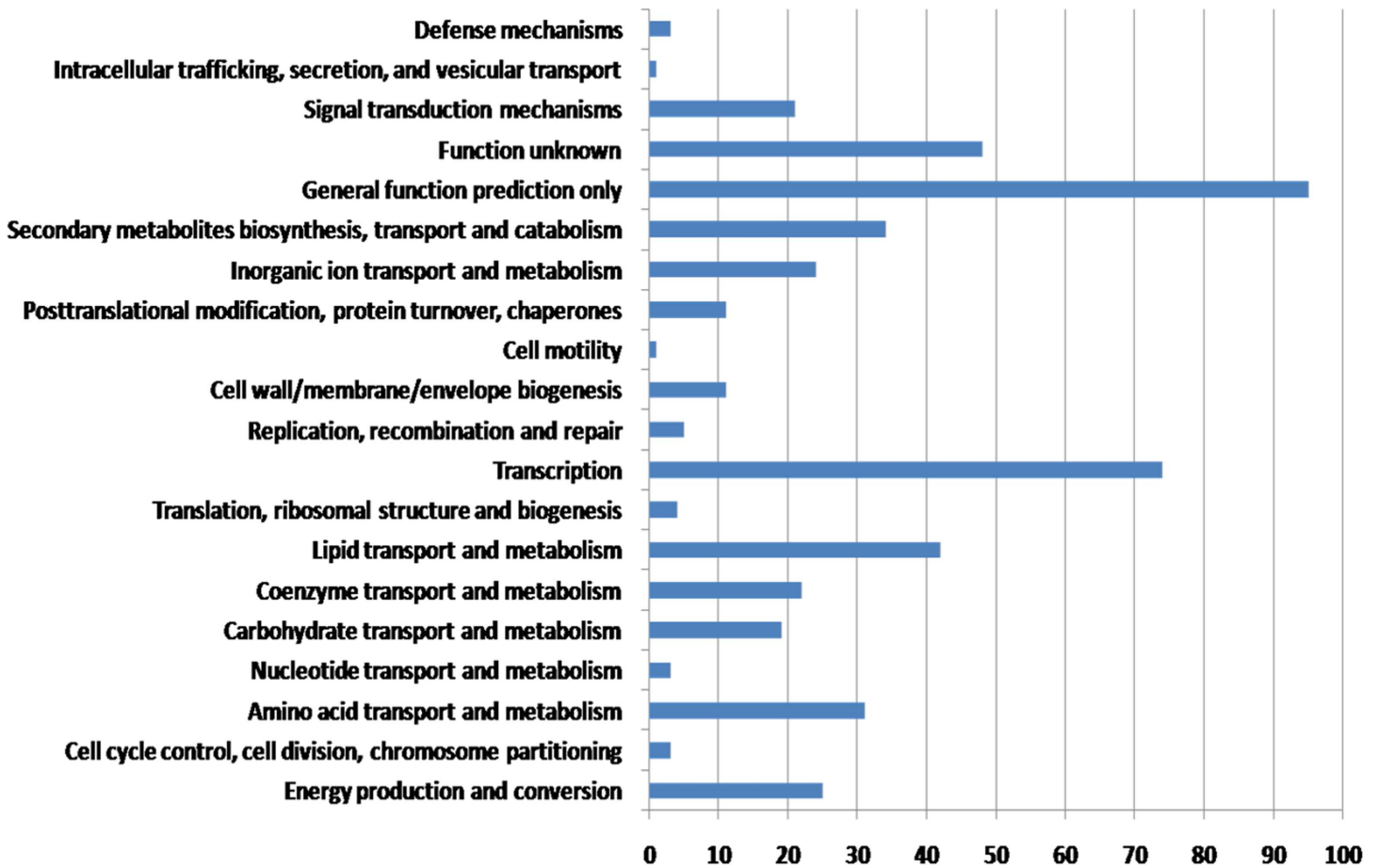


Fig 4. COG classification of *Mycobacterium abscessus*-specific genes.

<https://doi.org/10.1371/journal.pone.0172831.g004>

transcription category, there were 32 transcriptional regulator genes. The high number of transcriptional regulator genes specific to *Mycobacterium abscessus* may help in the survival of *Mycobacterium abscessus* in unstable environment as adaptive responses as mediated by transcriptional regulators [33].

Furthermore, out of the 811 gene families there were genes that were related to the bacterial quorum sensing. It has been reported that quorum sensing can cause virulence in bacteria by response to variety of signaling molecules called autoinducers [34]. This would allow bacteria to control their own gene expression and distinguish between low and high cell population density, which is important for them to adapt to environmental changes and also the changes in cell number [34]. Therefore, the acquisition of these *Mycobacterium abscessus*-specific quorum sensing genes/operons might be important to help *Mycobacterium abscessus* to survive in a wide range of adverse environmental conditions. One of the quorum sensing gene operons was the LuxI/LuxR signal-response system. It has been reported that this system controls the bioluminescence in *Vibrio fischeri* by monitoring the concentration acylated homoserine lactone (HSL), an autoinducer that regulates density-dependent light production in *Vibrio fischeri*. Moreover, researchers found that the homologs of LuxI and LuxR are also able to regulate the process of exoenzyme synthesis, conjugation, antibiotic production, luminescence and biofilm formation [35].

We also found a NTD operon consisting of *ntdA*, *ntdB* and *ntdC* (as well as a *glcP* gene) that were specific to *Mycobacterium abscessus*. The *glcP* located downstream of the *ntdABC* operon has been reported to co-express with the *ntdABC* operon [36]. The NTD operon is able to produce neutrehalosadiamine (NTD), amino acid antibiotics that act as an autoinducer. Besides that, the production of NTD can act as antibiotics which inhibit the growth of the other competing organisms [36].

Another *Mycobacterium abscessus*-specific quorum sensing operon was the phenazine operon. The secondary metabolites are usually produced by variety of bacteria especially pseudomonads. *Mycobacterium abscessus* possess 6 Phz homologues with three of the homologues (*phzC*, *phzD* and *phzE*) are linked and the organization of the genes are similar with *Pseudomonas aeruginosa*. Interestingly, it has been reported that the phz operon has a common ancestry and moved between species and inserted into *Mycobacterium abscessus* [37]. Phenazine is related to pathogenesis mainly due to their ability to generate reactive oxygen species (ROS) in other organisms and tissues. Bacterial virulence could be enhanced when phenazine production interfere the host cell functions. Thus, *Mycobacterium abscessus* might have gained some non-mycobacterial virulence genes and quorum sensing genes, further indicating that *Mycobacterium abscessus* has the capability to evolve and gain virulence genes.

Mycobacterium abscessus lacks of genes in bacteria defense system

According to Makarova, prokaryotes defense systems can be categorized into 2 groups. The first involve defense system based on the self-nonsel self discrimination principle and second is based on the programmed cell death induced by infection. The first group of the defense system includes restriction modification (RM) system and the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). RM system will attack non-self-invaders and CRISPR can memorize the foreign infectious agent and attack it afterwards. The second group of the defense system involves the toxin-antitoxin (TA) system [26,38]. With that in mind we searched for these bacterial defense system components in *Mycobacterium abscessus* and compared them to the other bacteria like *Mycobacterium tuberculosis*, *Mycobacterium africanum* and *Salmonella* Paratyphi A (Table 2).

Interestingly, we found no complete systems or operons encoding for the defense system in the *Mycobacterium abscessus* ATCC19977. There were no complete set of genes fully encoding the RM, TA systems in *Mycobacterium abscessus* genome. There were also no CRISPR in *Mycobacterium abscessus* from the prediction of the CRISPRfinder. On the other hand, *Mycobacterium tuberculosis* CCDC5079, *Mycobacterium africanum* GM041182 and *Salmonella* Paratyphi A ATCC9150 had a more conserved genomic structure with a number of genes involved in the defense system. *Mycobacterium tuberculosis* CCDC5079 had 1 complete RM system, 13 pairs of TA system and 8 predicted CRISPRs; *Mycobacterium africanum* GM041182 had 1 complete RM system, 12 pairs of TA and 7 predicted CRISPRs. Other than the *Mycobacterium* species, *Salmonella* Paratyphi A which has been shown contain conserved genome structure also contained these defense genes. Thus, the lack of genes involved in the defense

Table 2. Bacterial defense system for *Mycobacterium abscessus* compared to other bacterial genomes.

Strain	RM system	TA system	CRISPR
<i>Mycobacterium abscessus</i> ATCC19977	0	0	0
<i>Mycobacterium tuberculosis</i> CCDC5079	1	13	8
<i>Mycobacterium africanum</i> GM041182	1	12	7
<i>Salmonella</i> Paratyphi A ATCC9150	2	1	3

<https://doi.org/10.1371/journal.pone.0172831.t002>

system might cause *Mycobacterium abscessus* to be unable to recognize foreign DNA genome and allow genome insertion. Besides, quorum sensing genes enable *Mycobacterium abscessus* to have greater opportunity to get exposed to different harsh environments, which in turn may further enhance the possibility of insertion of the foreign materials.

Conclusion

Our analysis suggests that the mycobacterial genus has undergone a series of gene gain and gene loss events in the course of its evolution which likely enable them adapt to different environments and led to the evolution of the present day rapid and slow growers.

Supporting information

S1 Table. Gene annotation of core genes and specific genes for the 28 *Mycobacterium* species.
(XLSX)

Acknowledgments

We would like to thank all members from the Genome Research Informatics Group for their IT or bioinformatics assistance and supports.

Author Contributions

Conceptualization: WYW SWC.

Data curation: WYW.

Formal analysis: WYW AD.

Funding acquisition: SWC.

Investigation: WYW AD.

Methodology: WYW.

Project administration: WYW.

Resources: WYW SWC.

Software: WYW.

Supervision: WYW SWC AD.

Validation: WYW SWC AD.

Visualization: WYW SWC AD.

Writing – original draft: WYW SWC AD.

Writing – review & editing: WYW SWC AD.

References

1. Roberts ED, Koneman EW, Kim YK (1991) *Mycobacterium*. In: Balows A, Hausler J, W. J., Herrmann KL, Isenberg HD, Shadomy HJ, editors. *Manual of clinical microbiology*. 5 ed. Washington, D.C.: American Society for Microbiology. pp. 304–339.

2. Wayne LG, Kubica GP (1986) The mycobacteria. In: Sneath PHA, Mair NS, Sharpe ME, Holt JG, editors. *Bergey's manual of systematic bacteriology*. Baltimore: The Williams & Wilkins Co. pp. 1435–1457.
3. Kim CJ, Kim NH, Song KH, Choe PG, Kim ES, Park SW, et al. (2013) Differentiating rapid- and slow-growing mycobacteria by difference in time to growth detection in liquid media. *Diagn Microbiol Infect Dis* 75: 73–76. <https://doi.org/10.1016/j.diagmicrobio.2012.09.019> PMID: 23114094
4. Johnson MM, Odell JA (2014) Nontuberculous mycobacterial pulmonary infections. *J Thorac Dis* 6: 210–220. <https://doi.org/10.3978/j.issn.2072-1439.2013.12.24> PMID: 24624285
5. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al. (2007) An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med* 175: 367–416. <https://doi.org/10.1164/rccm.200604-571ST> PMID: 17277290
6. Houben EN, Korotkov KV, Bitter W (2014) Take five—Type VII secretion systems of *Mycobacteria*. *Biochim Biophys Acta* 1843: 1707–1716. <https://doi.org/10.1016/j.bbamcr.2013.11.003> PMID: 24263244
7. Devulder G, Perouse de Montclos M, Flandrois JP (2005) A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int J Syst Evol Microbiol* 55: 293–302. <https://doi.org/10.1099/ijs.0.63222-0> PMID: 15653890
8. Gey Van Pittius NC, Gamielien J, Hide W, Brown GD, Siezen RJ, Beyers AD (2001) The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol* 2: RESEARCH0044. PMID: 11597336
9. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9: 75. <https://doi.org/10.1186/1471-2164-9-75> PMID: 18261238
10. Liang W, Zhao Y, Chen C, Cui X, Yu J, Xiao J, et al. (2012) Pan-genomic analysis provides insights into the genomic variation and evolution of *Salmonella Paratyphi A*. *PLoS One* 7: e45346. <https://doi.org/10.1371/journal.pone.0045346> PMID: 23028950
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
12. Altschul SF, Lipman DJ (1990) Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A* 87: 5509–5513. PMID: 2196570
13. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575–1584. PMID: 11917018
14. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J (2012) PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28: 416–418. <https://doi.org/10.1093/bioinformatics/btr655> PMID: 22130594
15. Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* 11: 472–477. PMID: 19086349
16. Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804. <https://doi.org/10.1038/nature02053> PMID: 14574403
17. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. (2007) Clustal X and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
18. Librado P, Vieira FG, Sanchez-Gracia A, Kolokotronis SO, Rozas J (2014) Mycobacterial phylogenomics: an enhanced method for gene turnover analysis reveals uneven levels of gene gain and loss among species and gene families. *Genome Biol Evol* 6: 1454–1465. <https://doi.org/10.1093/gbe/evu117> PMID: 24904011
19. Grissa I, Vergnaud G, Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35: W52–57. <https://doi.org/10.1093/nar/gkm360> PMID: 17537822
20. Zakhm F, Belayachi L, Ussery D, Akrim M, Benjouad A, El Ouad R, et al. (2011) Mycobacterial species as case-study of comparative genome analysis. *Cell Mol Biol (Noisy-le-grand)* 57 Suppl: OL1462–1469.
21. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kampfer P, Maiden MC, et al. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* 52: 1043–1047. <https://doi.org/10.1099/00207713-52-3-1043> PMID: 12054223
22. Tortoli E (2012) Phylogeny of the genus *Mycobacterium*: many doubts, few certainties. *Infect Genet Evol* 12: 827–831. <https://doi.org/10.1016/j.meegid.2011.05.025> PMID: 21684354
23. Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW (2006) The evolution of mammalian gene families. *PLoS One* 1: e85. <https://doi.org/10.1371/journal.pone.0000085> PMID: 17183716

24. Zakham F, Aouane O, Ussery D, Benjouad A, Ennaji MM (2012) Computational genomics-proteomics and Phylogeny analysis of twenty one mycobacterial genomes (Tuberculosis & non Tuberculosis strains). *Microb Inform Exp* 2: 7. <https://doi.org/10.1186/2042-5783-2-7> PMID: 22929624
25. Rahman SA, Singh Y, Kohli S, Ahmad J, Ehtesham NZ, Tyagi AK, et al. (2014) Comparative analyses of nonpathogenic, opportunistic, and totally pathogenic mycobacteria reveal genomic and biochemical variabilities and highlight the survival attributes of *Mycobacterium tuberculosis*. *MBio* 5: e02020. <https://doi.org/10.1128/mBio.02020-14> PMID: 25370496
26. Makarova KS, Wolf YI, Koonin EV (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res* 41: 4360–4377. <https://doi.org/10.1093/nar/gkt157> PMID: 23470997
27. Sala A, Bordes P, Genevieux P (2014) Multiple toxin-antitoxin systems in *Mycobacterium tuberculosis*. *Toxins (Basel)* 6: 1002–1020.
28. Conner RM, Hansen PA (1967) Effect of valine, leucine, and isoleucine on the production of fly toxin by *Bacillus thuringiensis* and related organisms. *J Invertebr Pathol* 9: 114–125. PMID: 6073674
29. Kosono S, Haga K, Tomizawa R, Kajiyama Y, Hatano K, Takeda S, et al. (2005) Characterization of a multigene-encoded sodium/hydrogen antiporter (*sha*) from *Pseudomonas aeruginosa*: its involvement in pathogenesis. *J Bacteriol* 187: 5242–5248. <https://doi.org/10.1128/JB.187.15.5242-5248.2005> PMID: 16030218
30. Jarlier V, Nikaido H (1990) Permeability barrier to hydrophilic solutes in *Mycobacterium chelonae*. *J Bacteriol* 172: 1418–1423. PMID: 2307653
31. Mailaender C, Reiling N, Engelhardt H, Bossmann S, Ehlers S, Niederweis M (2004) The MspA porin promotes growth and increases antibiotic susceptibility of both *Mycobacterium bovis* BCG and *Mycobacterium tuberculosis*. *Microbiology* 150: 853–864. <https://doi.org/10.1099/mic.0.26902-0> PMID: 15073295
32. Dumas E, Christina Boritsch E, Vandenbogaert M, Rodriguez de la Vega RC, Thiberge JM, Caro V, et al. (2016) Mycobacterial Pan-Genome Analysis Suggests Important Role of Plasmids in the Radiation of Type VII Secretion Systems. *Genome Biol Evol* 8: 387–402. <https://doi.org/10.1093/gbe/evw001> PMID: 26748339
33. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY (2006) Prediction of amyloidogenic and disordered regions in protein chains. *PLoS Comput Biol* 2: e177. <https://doi.org/10.1371/journal.pcbi.0020177> PMID: 17196033
34. Rutherford ST, Bassler BL (2012) Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb Perspect Med* 2.
35. Oinuma K, Greenberg EP (2011) Acyl-homoserine lactone binding to and stability of the orphan *Pseudomonas aeruginosa* quorum-sensing signal receptor QscR. *J Bacteriol* 193: 421–428. <https://doi.org/10.1128/JB.01041-10> PMID: 21097632
36. Inaoka T, Ochi K (2007) Glucose uptake pathway-specific regulation of synthesis of neotrehalosadamine, a novel autoinducer produced in *Bacillus subtilis*. *J Bacteriol* 189: 65–75. <https://doi.org/10.1128/JB.01478-06> PMID: 17056753
37. Fitzpatrick DA (2009) Lines of evidence for horizontal gene transfer of a phenazine producing operon into multiple bacterial species. *J Mol Evol* 68: 171–185. <https://doi.org/10.1007/s00239-009-9198-5> PMID: 19189039
38. Librado P, Vieira FG, Rozas J (2012) BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* 28: 279–281. <https://doi.org/10.1093/bioinformatics/btr623> PMID: 22080468