

RESEARCH ARTICLE

# Whole-Genome Sequencing and Comparative Analysis of *Mycobacterium brisbanense* Reveals a Possible Soil Origin and Capability in Fertiliser Synthesis

Wei Yee Wee<sup>1,2</sup>, Tze King Tan<sup>1,2</sup>, Nicholas S. Jakubovics<sup>3,4</sup>, Siew Woh Choo<sup>1,2\*</sup>

**1** Genome Informatics Research Laboratory, High Impact Research Building (HIR) Building, University of Malaya, Kuala Lumpur, 50603, Malaysia, **2** Department of Oral Biology and Biomedical Sciences, Faculty of Dentistry, University of Malaya, Kuala Lumpur, 50603, Malaysia, **3** Center for Oral Health Research, School of Dental Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne, United Kingdom, **4** Genome Solutions Sdn Bhd, Suite 8, Innovation Incubator UM, Level 5, Research Management & Innovation Complex, University of Malaya, 50603, Kuala Lumpur, Malaysia

\* [lchoo@um.edu.my](mailto:lchoo@um.edu.my)



**OPEN ACCESS**

**Citation:** Wee WY, Tan TK, Jakubovics NS, Choo SW (2016) Whole-Genome Sequencing and Comparative Analysis of *Mycobacterium brisbanense* Reveals a Possible Soil Origin and Capability in Fertiliser Synthesis. PLoS ONE 11(3): e0152682. doi:10.1371/journal.pone.0152682

**Editor:** Riccardo Manganelli, University of Padova, Medical School, ITALY

**Received:** December 3, 2015

**Accepted:** March 17, 2016

**Published:** March 31, 2016

**Copyright:** © 2016 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 files.

**Funding:** This project was supported by University of Malaya and Ministry of Education, Malaysia under the High Impact Research (HIR) grant UM.C/625/HIR/MOHE/CHAN-08 and UMRG grant (grant number: RG541-13HTM) from University of Malaya and Ministry of Education, Malaysia.

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

## Abstract

*Mycobacterium brisbanense* is a member of *Mycobacterium fortuitum* third biovariant complex, which includes rapidly growing *Mycobacterium* spp. that normally inhabit soil, dust and water, and can sometimes cause respiratory tract infections in humans. We present the first whole-genome analysis of *M. brisbanense* UM\_WWY which was isolated from a 70-year-old Malaysian patient. Molecular phylogenetic analyses confirmed the identification of this strain as *M. brisbanense* and showed that it has an unusually large genome compared with related mycobacteria. The large genome size of *M. brisbanense* UM\_WWY (~7.7Mbp) is consistent with further findings that this strain has a highly variable genome structure that contains many putative horizontally transferred genomic islands and prophage. Comparative analysis showed that *M. brisbanense* UM\_WWY is the only *Mycobacterium* species that possesses a complete set of genes encoding enzymes involved in the urea cycle, suggesting that this soil bacterium is able to synthesize urea for use as plant fertilizers. It is likely that *M. brisbanense* UM\_WWY is adapted to live in soil as its primary habitat since the genome contains many genes associated with nitrogen metabolism. Nevertheless, a large number of predicted virulence genes were identified in *M. brisbanense* UM\_WWY that are mostly shared with well-studied mycobacterial pathogens such as *Mycobacterium tuberculosis* and *Mycobacterium abscessus*. These findings are consistent with the role of *M. brisbanense* as an opportunistic pathogen of humans. The whole-genome study of UM\_WWY has provided the basis for future work of *M. brisbanense*.

## Introduction

*Mycobacterium* is a genus within the Actinobacteria that includes well known human pathogens such as *Mycobacterium tuberculosis*, the causative agent of tuberculosis, and *Mycobacterium leprae*, the agent of leprosy. The nontuberculous mycobacteria (NTM), also known as environmental mycobacteria, are frequently isolated from environmental sources such as soil, water or animal materials [1,2]. Although these organisms tend to be found in the environment, many NTM are opportunistic pathogens, capable of causing human disease among immune-compromised individuals and even sometimes in immune-competent individuals [3].

The *M. fortuitum* complex includes numerous species, *M. fortuitum*, *Mycobacterium peregrinum*, *Mycobacterium mucogenicum*, *Mycobacterium mageritense*, *Mycobacterium farcinogenes*, *Mycobacterium septicum* [4] and an additional taxon with more than one species known as the third biovariant complex. This complex includes *Mycobacterium porcinum*, *Mycobacterium boenickei*, *Mycobacterium neworleansense*, *Mycobacterium houstonense*, *Mycobacterium conceptionense* and *M. brisbanense* [4,5]. *M. brisbanense* is a rapidly growing *Mycobacterium* which has been isolated from human infections and from soil [6–8]. Species of the *M. fortuitum* complex can cause a variety of human infections including respiratory disease and wound infections [9,10]. In this paper, we described the genome of *M. brisbanense* UM\_WWY (referred to as UM\_WWY after this) isolated from the sputum of a 70-year-old Malaysian male patient with a diagnosis of latent tuberculosis. The genome of UM\_WWY has been deposited at GenBank with the RefSeq ID of NZ\_AUWS00000001.

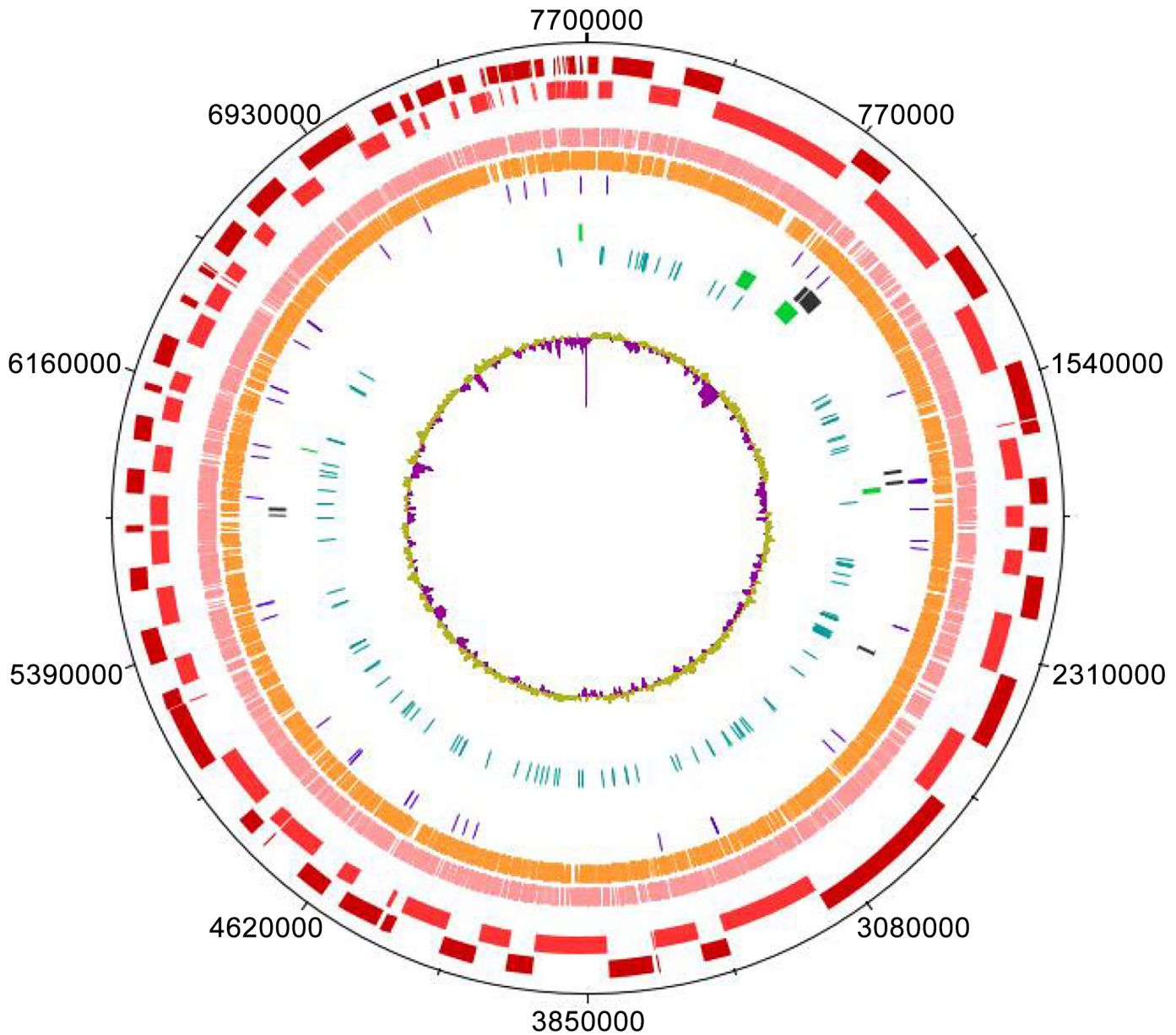
## Results

### Genome Overview

The genome of UM\_WWY was shot-gun sequenced using the Illumina HiSeq2000 platform, generating 63,311,594 raw sequencing reads. These reads were pre-processed to remove adaptor sequences and trimmed at a Phred quality score of 20 using CLC Genomic Workbench version 5.1 (CLC bio, Denmark). The assembly of 53,995,482 pre-processed reads resulted in 130 contigs. The assembly has a genomic size of 7,690,050bp with a G+C content of 66.4%. The sequenced genome of UM\_WWY has an N75 value of 74,225bp, N50 value of 112,506bp and N25 size of 198,038bp, indicating the high quality of this assembly. The UM\_WWY genome was BLAST-searched against NCBI database of known plasmids and no plasmid sequences were detected in the genome. An overview of the UM\_WWY genome is shown in Fig 1. Interestingly, a 980kbp intact prophage was predicted in the sequenced genome of UM\_WWY by the PHAST software. This large intact prophage observed within contig suggest it was recently inserted, contains 98 putative protein-coding genes and most of them have unknown functions (S1 Fig). Further studies on the functions of these hypothetical proteins may give better insights into the functions or impact of this prophage on the traits of UM\_WWY. There was evidence that this prophage is currently active as we observed the presence of reads aligned to a closed form of the origin of insertion (*attP* site), indicating that lytic phage were likely to have been present in the cultures (S2 Fig).

### Phylogenetic Inference

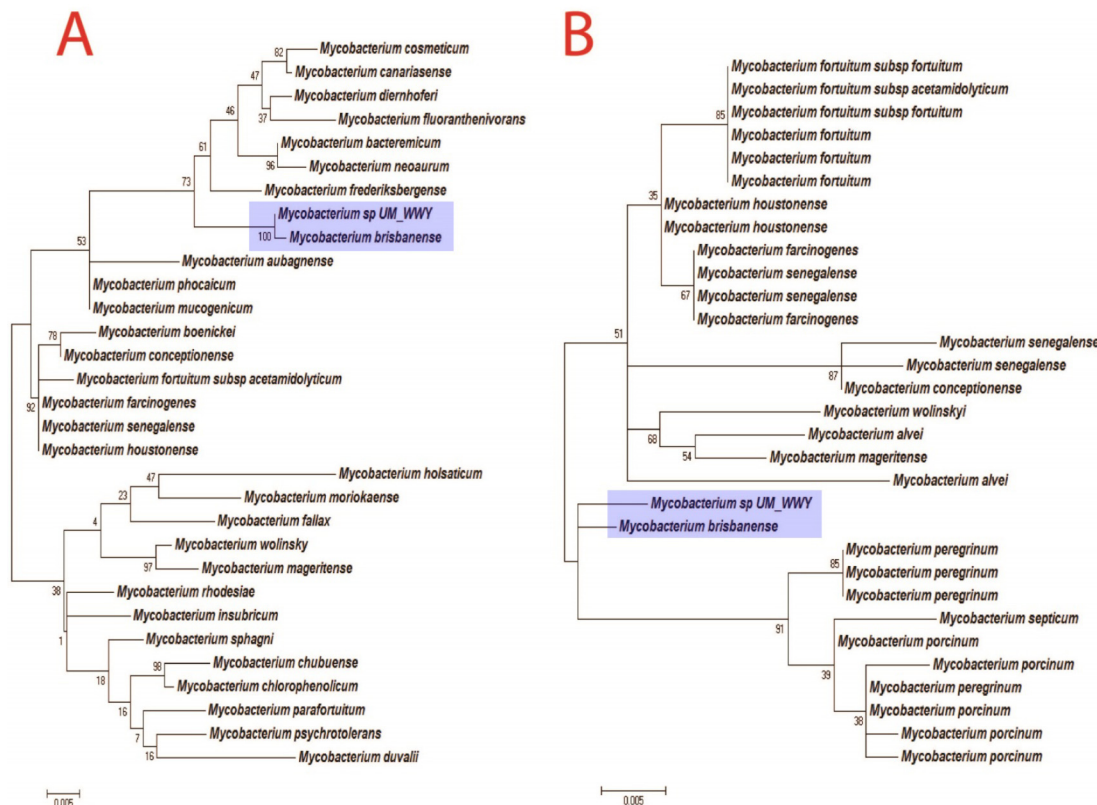
The taxonomic position of UM\_WWY was determined based on phylogenetic inferences of 16S *rRNA* and *hsp65* marker genes. The 16S *rRNA*-based phylogenetic tree was constructed using Hasegawa-Kishino-Yano DNA substitution model in MEGA5 [11]. The phylogenetic tree suggested that *M. brisbanense* is the closest neighbour of UM\_WWY with a strong bootstrapping value of 100% (Fig 2A). Most of the closely related species of UM\_WWY are



**Fig 1. Schematic circular diagram of the UM\_WWY genome.** The following color bars represent: Dark red and light red are contigs, pink and orange represent forward and reverse CDS respectively, purple is tRNA, dark blue is genomic island, green is prophage, grey is virulence gene and the last track is the percentage of GC content.

doi:10.1371/journal.pone.0152682.g001

members of the *M. fortuitum* complex, for example, *M. brisbanense*, *Mycobacterium cosmeticum*, *M. fortuitum*, *Mycobacterium housetonense*, *Mycobacterium senegalense*, *Mycobacterium boenickei* and *Mycobacterium mageritense*, suggesting that UM\_WWY is likely a new member of *M. brisbanense*. Additionally, UM\_WWY has the highest 16S rRNA gene sequence identity (99.8%) with *M. brisbanense* and second highest (98.53%) with *Mycobacterium cosmeticum*. The *hsp65*-based phylogenetic tree also showed that most of the closest neighbours of UM\_WWY are also the members of *M. fortuitum* complex (Fig 2B). Besides that, it also supported that *M. brisbanense* as the closest neighbour of UM\_WWY.



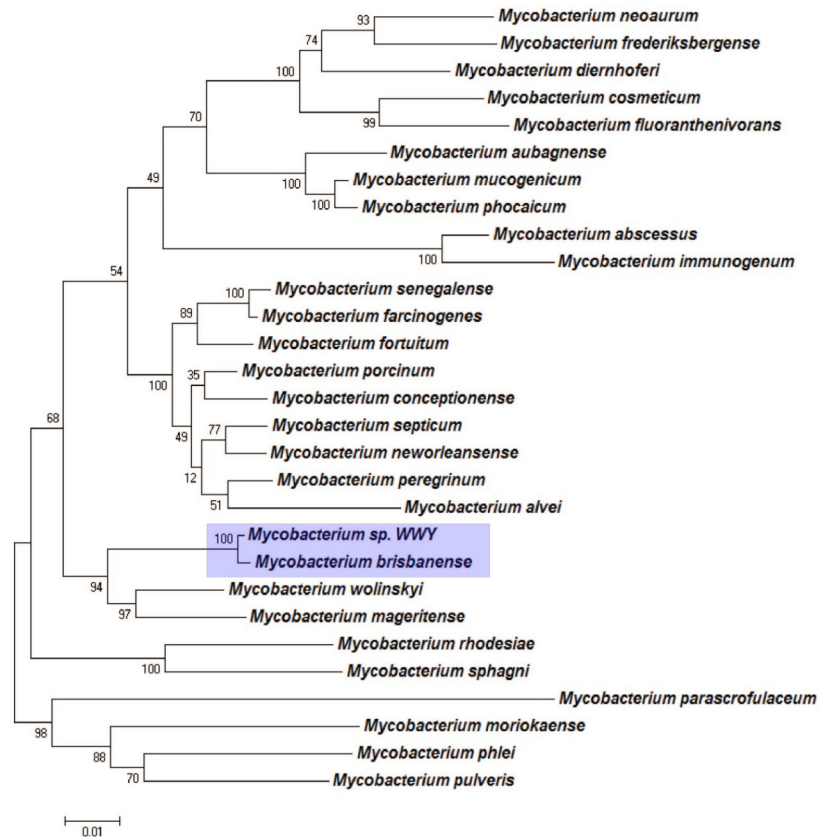
**Fig 2. Phylogenetic inference.** (A) 16S rRNA-based phylogenetic tree. (B) hsp65-based phylogenetic tree. Maximum likelihood phylogenetic trees were inferred to identify the UM\_WWY. Both results showed that *M. brisbanense* is the closest neighbour of UM\_WWY.

doi:10.1371/journal.pone.0152682.g002

To further validate the taxonomic position of UM\_WWY, we reconstructed a more robust tree using a multiple-genes approach based on the concatenated sequences of five housekeeping genes: 16S rRNA, *rpoB*, *tuf*, *sodA* and *hsp65* genes [12]. The concatenated sequences were used to construct a phylogenetic tree using the same criteria as the above single gene-based approaches. Again, the multi-gene tree analysis showed results that were consistent with the single gene-based trees. *M. brisbanense* is the closest neighbour of UM\_WWY with a sequence identity of 99% (Fig 3). Finally, a phylogenetic tree was built based on whole genome sequences of UM\_WWY and its relatives (S1 Fig). Although genome sequences are not available for many species of the *M. fortuitum* complex, UM\_WWY was found to cluster most closely to *M. fortuitum* complex species. Taken all together, our data support the view that UM\_WWY is a *M. brisbanense* species and grouped together with the members of *M. fortuitum* complex.

### Functional Analysis

The sequenced genome of UM\_WWY was annotated using the Rapid Subsystem Technology using Subsystem (RAST) pipeline [13]. We identified 7,504 protein-coding genes in the genome of UM\_WWY. Of these, RAST successfully assigned functions to 2,362 (32%) genes (S2 Fig). Many genes were distributed in categories such as amino acid and derivatives (731), carbohydrates (644), cofactors, vitamins, prosthetic group pigments (413), fatty acids, lipids and isoprenoids (331) which are primarily responsible for the basic biological functions of the UM\_WWY. Interestingly, we found 112 genes putatively involved in virulence, disease and defence. Of these, 59 genes were predicted to be involved in resistance to antibiotics and toxic



**Fig 3. Supermatrix phylogenetic tree by concatenating *rpoB*, *tuf*, *sodA*, *16S rRNA* and *hsp65* gene sequences.** Our results indicated that UM\_WWY is closely related to *M. brisbanense* and other members in the *M. fortuitum* complex.

doi:10.1371/journal.pone.0152682.g003

compounds, whereas the others (53 genes) were putatively involved in host cell invasion and intracellular resistance. The antimicrobial resistance genes include arsenic resistance genes, cobalt-zinc-cadmium resistance genes, fluoroquinolones resistance genes and a beta-lactamase gene. The presence of these resistance genes indicates that UM\_WWY may be resistant to certain heavy metals as well as fluoroquinolones and beta-lactam antibiotics (for example, penicillins, cephamycins and carbapenems) [14].

Three putative rRNAs and 73 tRNAs were identified in the genome of UM\_WWY. The number of tRNA genes in UM\_WWY is relatively high compared to other mycobacteria such as *Mycobacterium tuberculosis* (45) or *M. fortuitum* (55). Interestingly, further examination of these tRNA genes revealed that a large number of tRNA genes were clustered together in the genome of UM\_WWY, forming a tRNA island (or a genomic region with a high density of tRNA genes) of approximately 3,175 bp, comprising 23 putative tRNA genes. It is likely that the tRNA island has been inserted into the UM\_WWY genome through a horizontal gene transfer (HGT) event. The order of the tRNAs within this island is highly similar to the order of tRNA genes in tRNA islands in four *Mycobacterium abscessus* strains isolated from the United States that we reported previously (S2 Table and S3 Fig) [15]. Although they are highly similar, there is one missing tRNA (tRNA-Leu) gene in the fourth position and one extra tRNA (tRNA-Arg) gene at the 5' end of the tRNA island in the genome of UM\_WWY. Therefore, it appears likely that these islands in UM\_WWY and *Mycobacterium abscessus* have originated from the same

source, probably from a transposition event since there is a transposase located immediately upstream of the tRNA island.

## Horizontally Transferred Genomic Islands

Our data showed that the genome size of UM\_WWY strain is considerably larger than other *Mycobacterium* species that we analysed in this study. The genome of the clinically derived strain UM\_WWY is approximately 1Mbp larger than *Mycobacterium marinum*, which has the second largest genome of the *Mycobacterium* species that were included here. The large genome size of UM\_WWY may be due, at least in part, to HGT events involving the acquisition of foreign DNA over evolutionary time. Therefore, we were interested to search for possible horizontally transferred Genomic Islands (GIs) in the genome of UM\_WWY. Genomic Islands are clusters of genes that are inserted into a bacterial genome during a single HGT event. These islands often play important roles in microbial evolution, virulence, drug resistance and/or adaptation to different environments [16].

Using IslandViewer [17,18], we identified 35 putative GIs (S1 Table), covering approximately 0.36Mbp of the UM\_WWY genome. The large number of predicted GIs suggested that the HGT has played an important role in the evolution of UM\_WWY and has contributed to the genome size. GI34 corresponds to transposon Tn554, a high frequency, site-specific transposable element that encodes three genes associated with transposition, *tnpA*, *tnpB* and *tnpC* [19]. Transposases were also identified in GI23 and GI33, indicating that these elements may also have inserted into the genome through transposition events. Previous studies have shown that GIs can play a role in the evolution of pathogenic *Mycobacterium* species [20,21]. Interestingly, we observed a putative virulence related gene, *espL* in GI35. The *espL* gene is one the members of *ESX-1* gene cluster for the type VII secretion systems that are responsible for the export of virulence factors [22]. However, the whole *ESX-1* cluster is required for full virulence in *Mycobacterium tuberculosis* [23]. Although we observed *espL* in GI35, other genes in the *ESX-1* cluster were not present indicating that the *espL* gene may not confer virulence traits on UM\_WWY.

Additionally, we found a putative YefM toxin-antitoxin (TA) pathogenicity island (GI6) in the UM\_WWY genome. TA systems have been reported to be associated with persistence and adaptability in *Mycobacterium tuberculosis* during infection [24]. Thus the insertion of this island might contribute to the virulence of UM\_WWY.

## Comparative Genome Analysis

To retrieve more insights into the genome of UM\_WWY, we compared UM\_WWY with the *Mycobacterium* species that were found to be closely related according to our phylogenetic analyses above, namely *Mycobacterium mageritense* JR2009, *Mycobacterium septicum* DSM44393 and *M. fortuitum* DSM46621. Prior to the comparison of these genomes, we re-annotated all genomes using RAST server and the results were summarised in Table 1.

UM\_WWY has a genome size of approximately 7.69 Mbp, which is 0.79–1.35Mbp larger than the other three species. The difference was reflected in the high number of strain-specific gene clusters observed in the genome of UM\_WWY (Fig 4). For instance, UM\_WWY has the highest number of strain-specific genes (3,214 genes) compared to *Mycobacterium mageritense* JR2009 (931 genes), *Mycobacterium septicum* DSM44393 (1,143 genes) and *M. fortuitum* DSM46621 (753 genes). The high number of UM\_WWY-specific genes suggests that this strain is very divergent from the most closely related species.

To further increase the stringency of the strain-specific genes, the genome of UM\_WWY was further compared with genomes from 146 other mycobacterial strains. We clustered the

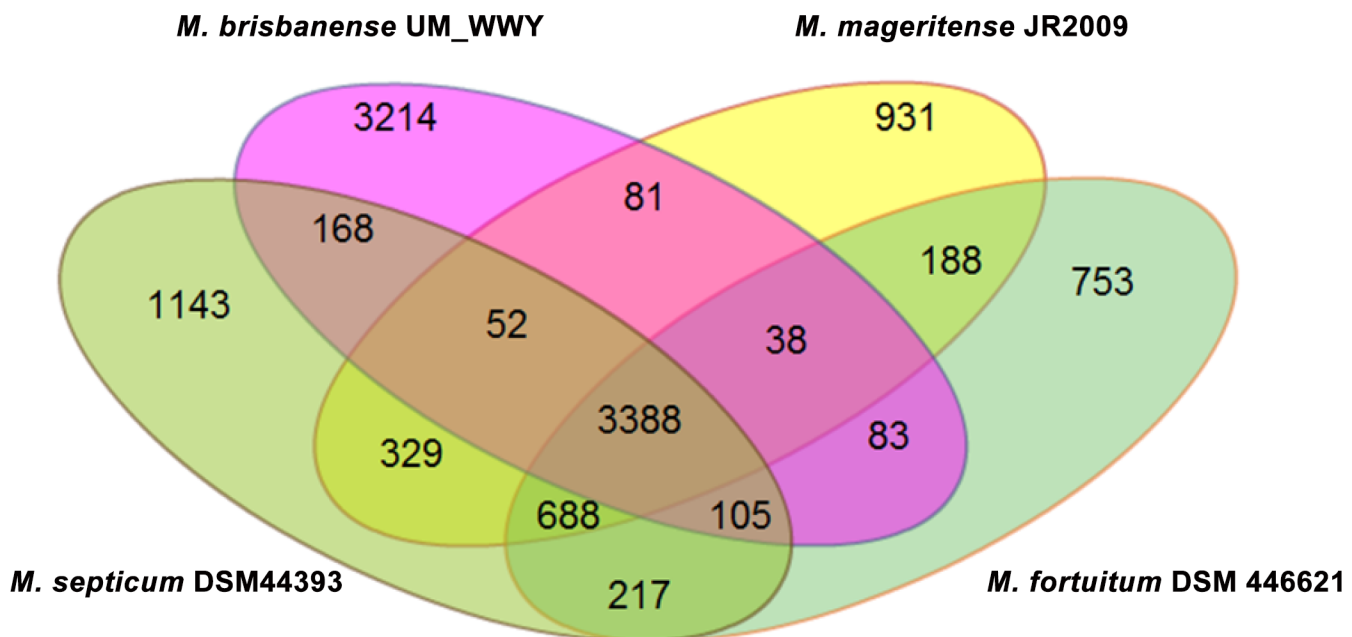
**Table 1. Genome statistics of UM\_WWY and three closely related species.**

Strains	<i>Mycobacterium brisbanense</i> UM_WWY	<i>Mycobacterium septicum</i> DSM44393	<i>Mycobacterium mageritense</i> JR2009	<i>Mycobacterium fortuitum</i> DSM46621
Genome Status	Draft	Draft	Draft	Draft
Genome Size (Mb)	7.69	6.90	6.49	6.34
Number of tRNAs	73	52	72	55
Number of rRNAs	3	3	3	3
Number of protein-coding genes	7,504	6,598	6,238	6,077

doi:10.1371/journal.pone.0152682.t001

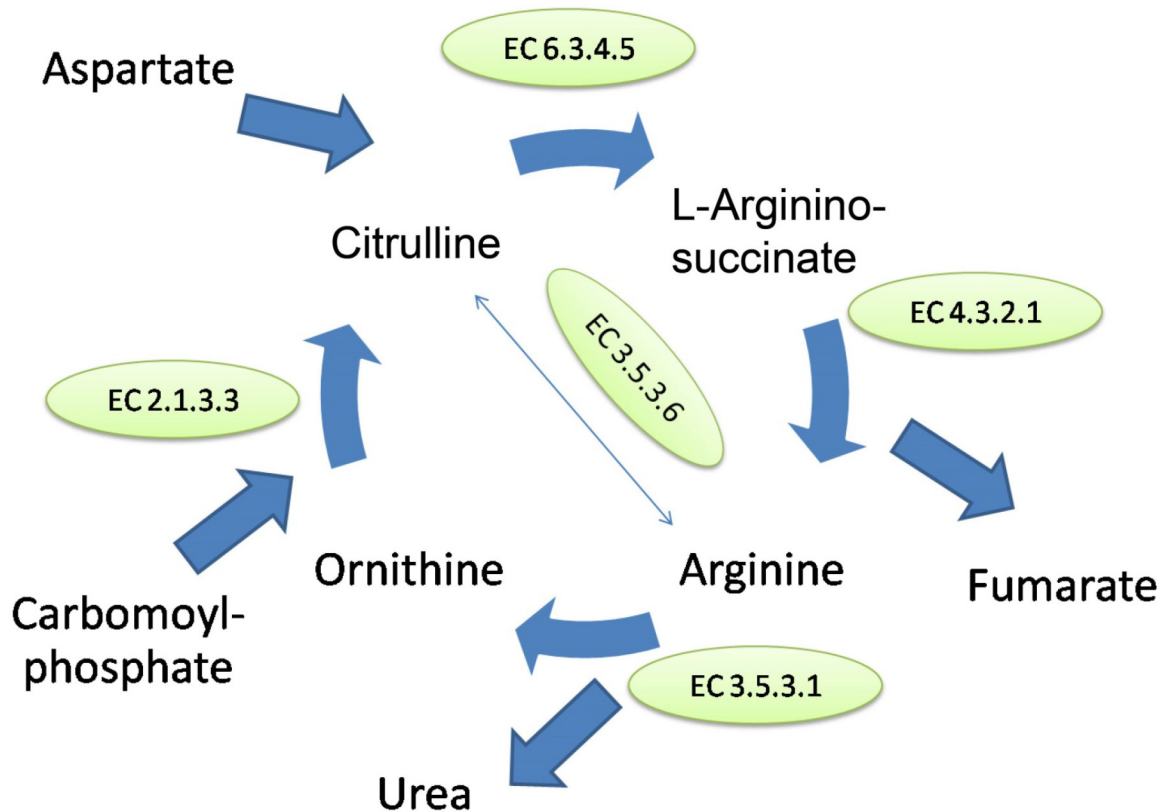
genes of UM\_WWY with the genes from the 146 strains composed of 27 different *Mycobacterium* species (S3 Table). Any UM\_WWY-specific genes that overlapped the genes present in 146 mycobacterial strains were discarded. This stringent filtering step successfully reduced the set of UM\_WWY-specific genes to 1,425. The presence of these 1,425 UM\_WWY-specific genes may contribute to the unique traits or phenotypes of UM\_WWY.

Interestingly, further examinations of these genes revealed a gene encoding arginase (EC3.5.3.1), which is an enzyme found in some eubacteria and eukaryotes [25]. Arginase is one of the key enzymes in the urea cycle (Fig 5). This enzyme plays an important role in the hydrolysis of L-arginine to ornithine and urea. We examined the urea cycle KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway and we found that the genome of UM\_WWY also encodes other enzymes needed to form a complete urea cycle such as argininosuccinate synthetase (EC 6.3.4.5), argininosuccinate lyase (EC 4.3.2.1) and ornithine transcarbamoylase (EC 2.1.3.3) (Fig 5). In addition, gene encoding urease present in UM\_WWY indicating that arginine or urea could be sources of nitrogen for this organism [26]. Overall, this analysis indicates



**Fig 4. Venn diagram showing the gene distribution of *M. fortuitum* complex members and UM\_WWY.** The four genomes shared 3,388 gene clusters and UM\_WWY contains the highest number of strain-specific genes.

doi:10.1371/journal.pone.0152682.g004

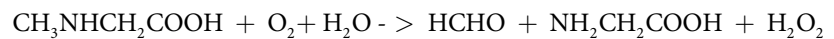


**Fig 5. The pathway for urea cycle derived from KEGG pathway analysis.** UM\_WWY possess a gene encoding for a putative enzyme arginase (EC 3.5.3.1) which hydrolyzes arginine into ornithine and urea. EC 6.3.4.5 = Argininosuccinate synthetase; EC 4.3.2.1 = Argininosuccinate lyase; EC 2.1.3.3 = Ornithine transcarbamoylase.

doi:10.1371/journal.pone.0152682.g005

that UM\_WWY is likely the first or the only known mycobacterial species which possesses the genetic basis for a complete urea cycle.

Apart from arginase, UM\_WWY also encodes a putative sarcosine oxidase (SOX) operon which is absent in other mycobacterial species. SOX is an inducible enzyme in soil bacteria [27]. Lawrence and colleagues showed that SOX is a heterotetrameric enzyme and can catalyze the reaction of sarcosine to glycine, formaldehyde and H<sub>2</sub>O<sub>2</sub>.



The SOX operon is composed of five closely packed genes which encode the four SOX subunit (SOX  $\alpha$ , SOX  $\beta$ , SOX  $\gamma$ , SOX  $\delta$ ) and serine hydroxymethyltransferase (*glyA*). These genes are arranged in the order of *glyA*, followed by SOX  $\beta$ , SOX  $\delta$ , SOX  $\alpha$  and SOX  $\gamma$  [28]. Surprisingly, a screen for these genes in the 146 mycobacterial genomes in MycoBase (manuscript under review), we found that UM\_WWY is the only mycobacterial genome that possesses the entire SOX operon described by Lawrence and colleagues. The presence of the SOX operon may enable UM\_WWY to grow on sarcosine as a sole source of carbon. Taken together, our data suggest that UM\_WWY possesses many specific genes that are predicted to be related to the nitrogen metabolism, suggesting that this bacterium may play an important role in the urea cycle and nitrogen cycling compared to other mycobacterial species.



## Comparative Virulence Gene Analysis

We found 122 putative non-redundant virulence genes in the genome of UM\_WWY. Almost all these virulence genes (121 genes) in the UM\_WWY genome are shared by other human pathogens such as *Mycobacterium tuberculosis*. The genome of UM\_WWY also contains genes such as *ahpC*, *katG*, *sodC* and *sodA* that encode enzymes involved in resistance to oxidative stress and might be important in combatting the oxidative burst during survival within macrophages [22]. Another virulence gene, *nuoG* encodes a protein that is one of the 14 subunits of type I NADH dehydrogenase and is a key virulence factor in *Mycobacterium tuberculosis* due to its ability to inhibit apoptosis of macrophages [29]. Thus, this gene allows the bacteria to escape from the natural cell death program by host cells.

Virulence genes were also identified in UM\_WWY that encode proteins responsible for the synthesis of bacterial cell envelope, including *fbpA*, *fbpB*, *fbpC*, *erp*, *hbhA* and *mce*. The protein products of these genes are able to adhere and transport solutes across the mycobacterial cell wall in order to construct complex envelope lipids and proteins. For example, genes encoding Mce proteins, are organized in operons, and allow mycobacteria to enter into mammalian cell and survive inside the macrophage [30]. The UM\_WWY genome has three *mce* operons, *mce1*, *mce3* and *mce4*. Moreover, UM\_WWY also contains genes that encode proteins responsible for metal transporter. Metals such as iron, magnesium, copper and zinc are essential for the survival of bacteria in the environment or within the cell. However, excess levels of these metals are toxic to bacteria and thus metals transporters are required to maintain an appropriate level of metals within bacterial cells. The genes *irtA*, *irtB* and *ideR* in UM\_WWY encode proteins that play a role in the acquisition of iron [22]. Another virulence gene, *mgtC*, encoding a protein responsible for transport of magnesium is also present in the genome of UM\_WWY. Magnesium uptake is critical for intracellular pathogens to survive within macrophages as the magnesium level is very low [22].

There are many other putative virulence genes in UM\_WWY such as four pairs of two component system genes (*devR-devS*; *mprA-mprB*; *phoP-phoR*; *prrA-prrB*), five sigma factors (*sigA*, *sigE*, *sigF*, *sigH*, *sigM*) that are involved in gene expression regulation, the *pknG* gene that acts as a protein kinase, and the metabolic genes (*mma4*, *mmpl8*, *pks2*, *papA1*, *papA2*, *stf0*) that are involved in lipid and fatty acid metabolism. The functions of other virulence-related genes remain unknown. The full list of the virulence genes in UM\_WWY is shown in [S4 Table](#).

After comparing with 30 other mycobacterial species ([S4 Table](#)), we found 25 virulence genes were highly conserved across all the *Mycobacterium* species. In comparison with *Mycobacterium tuberculosis*, UM\_WWY has fewer virulence genes involved in the biosynthesis of phthiocerol dimycocerosate (PDIM) and phenolic glycolipid (PGL). PDIM is essential in prevention of acidification of phagosomes which disable enzymic digestion of internalized bacteria [31], while PGL is important in modulating the host immune response to infection [32]. For instance, of 20 genes in *Mycobacterium tuberculosis* that are involved in the synthesis of PDIM and PGL, UM\_WWY has only three (*mas*, *fadD26* and *fadD28*). This suggests that the synthesis of PDIM and PGL might be disrupted in UM\_WWY as previously shown in other mycobacterial species [33]. In addition, UM\_WWY lacks of four genes encoding for phospholipase C-type enzymes which are involved in the virulence of *Mycobacterium tuberculosis* [34]. The absence of these genes suggests that UM\_WWY is likely to be a less virulent pathogen than *Mycobacterium tuberculosis*.

## Discussion

In this paper, we report the first whole-genome sequence and analysis of *M. brisbanense* UM\_WWY, which is a new member of *M. fortuitum* complex that was isolated from a

70-year-old Malaysian patient. Our molecular phylogenetic analyses using single marker gene and multiple gene approaches confirmed that UM\_WWY is *M. brisbanense*.

One of the interesting observations in our analyses is that UM\_WWY with a draft genome size of 7.69Mb (approximately 1MB bigger than the second largest *Mycobacterium* sp.) is the largest among the known genome size of the mycobacterial species that we used in this study. The large genome size of UM\_WWY may reflect *M. brisbanense*'s diverse genomic structure probably driven by active HGT events. The evidence of large number of horizontally transferred GIs and prophages found in the genome of UM\_WWY has further supported our view that the HGT has played important role in shaping the genome of UM\_WWY. Recently it has been shown that a linear plasmid could be transferred *in vitro* from *Mycobacterium avium* to *Mycobacterium kansasii* and to *Mycobacterium bovis*, demonstrating that mycobacteria can transfer DNA between species naturally [35]. It is likely that *M. brisbanense* is particularly adept at the uptake and incorporation of genes from foreign sources. Interestingly, we also found evidence for the presence of an active bacteriophage in the *M. brisbanense* culture. In future studies, it would be interesting to characterise *M. brisbanense* bacteriophages in more detail since these may be important sources of HGT.

Strikingly, we found a large number of UM\_WWY-specific genes compared to other mycobacteria, probably due to large genome of this bacterium. Further analysis of these specific genes has revealed that some of these genes are involved in pathway for urea synthesis. Interestingly, UM\_WWY is the only *Mycobacterium* sp. (among all the mycobacterial genomes that we analysed in this study) that has the complete set of genes involved in the urea cycle including an arginase that can potentially convert arginine to ornithine and urea. It has been reported that arginine can be used by macrophages to produce reactive nitrogen intermediates such as nitric oxide that serve as antimicrobial agents against invading pathogens [36]. It has also been shown in *Helicobacter pylori* that the presence of arginase is able to hydrolyze arginine and inhibit the production of these reactive nitrogen intermediates, allowing the survival of the bacteria in the host by inhibiting the host from producing nitric oxide [37]. Thus, it is possible that the arginase gene may play an important role in the virulence of UM\_WWY, for example, by inhibiting the production of reactive nitrogen intermediates.

According to a previous report from Amon and colleagues, *Mycobacterium smegmatis* contains the highest number of predicted nitrogen metabolism genes among the sequenced mycobacteria [38]. This is probably due to the fact that the *Mycobacterium smegmatis* naturally lives in the soil. Interestingly, we found that UM\_WWY also has all of these genes that are involved in the nitrogen metabolism, suggesting that UM\_WWY is also likely to exist primarily in the soil. Although UM\_WWY was isolated from a clinical infection, there have been reports of isolation of *M. brisbanense* from soil. The high number of putative virulence genes found in the genome of UM\_WWY might enable this bacterium to transition between its primary environmental habitats to growth in humans as an opportunistic pathogen. If HGT has played an important role in shaping the genome of UM\_WWY, as suggested based on our analyses such as the presence of GIs, it is possible that this bacterium may continue acquiring new genes and may become increasingly more virulent in the future.

In this study, we revealed a region with a high density of tRNA genes (also predicted as a GI) in the genome of UM\_WWY that is highly similar to the uncommon tRNAs islands found in some mycobacterial strains such as the highly antibiotic resistant strain *Mycobacterium boletii* M24 and some *Mycobacterium massiliense* strains from the United States [15]. Although this is an interesting observation, the biological significance of this tRNA island is not clear. It is possible that the large tRNA gene cluster may be associated with elevated levels of protein synthesis, or possibly tied in with pathogenicity. Further experimental works such as cell

invasion assays and genetic manipulation may help to give better insights into the pathogenic potential of *M. brisbanense* and the roles of specific genetic factors.

## Conclusion

Here we present the first genome sequence of *M. brisbanense*. Comparative analyses suggest that UM\_WWY is likely to be primarily a soil bacterium and may have industrial potential for the synthesis of urea for use in fertilizers. The presence of large numbers of virulence-associated genes and the apparent capacity for HGT may allow UM\_WWY to evolve into a more serious human pathogen. This whole-genome sequence will provide a model genome for studying *M. brisbanense* in the future.

## Methodology

### Library Construction and Whole-Genome Sequencing

UM\_WWY was isolated from the sputum of a 70-year-old male patient with a diagnosis of latent tuberculosis in Malaysia. The DNA of UM\_WWY was obtained from Professor Dr. Ngeow Yun Fong from the Faculty of Medicine, University of Malaya. The DNA sample was fragmented using Covaris S2 for 120 seconds at temperature of 5.5–6.0 degree Celsius. The quantity and quality of fragmented materials were examined using the Agilent BioAnalyzer 2100. The sample was size selected using Invitrogen 2% agarose E-gels. Fragments with adapter molecules at both ends further underwent 10 cycles of PCR for library construction. The Agilent BioAnalyzer 2100 was used to validate the constructed genomic library and the pool of 8pM was loaded onto Illumina HiSeq2000 flow cell v3 for 100bp paired-end sequencing at approximately 1,000X coverage.

### Raw Data Preprocessing and Genome Assembly

Exact duplicates and reverse complement duplicate reads were filtered using standalone PRINSEQ lite version 0.20 [39]. The final reads pre-processing based on Phred quality score of 20 and the genome assembly were performed using CLC Genomic Workbench version 5.1. The assembly of the reads was carried out at length fraction of 0.7 and similarity fraction of 0.9, accepting minimal size of contig of 500bp.

### Phylogenetic Inferences

The candidates used in the classification of UM\_WWY were identified using the updated version of Bioinformatics Bacteria Identification Tool—leBIBI V5 [40]. The 16S rRNA-based phylogenetic tree was constructed using Hasegawa-Kishino-Yano DNA substitution model in Molecular Evolutionary Genetics Analysis (MEGA5) [11] tool with bootstrap value of 500. Using the same criteria as the 16S rRNA phylogenetic study, *hsp65*-based maximum likelihood phylogenetic was inferred using Tamura-Nei DNA substitution model. Five selected bacterial classification markers, *16S*, *hsp65*, *rpoB*, *tuf*, *sodA* and *rRNA* from all the possible closest species were extracted and the complete gene sequences were concatenated for phylogenetic tree construction.

### Genome Annotation

The sequenced genome of UM\_WWY was submitted to the RAST server for annotation [13]. Genomic islands in UM\_WWY were predicted by IslandViewer which integrates several different approaches to predict the GIs in genomes [18]. These approaches include the sequence composition based SIGI-HMM [41] and IslandPath-DIMOB [42,43] that have been shown to

have specificity of 86% to 92% and accuracy of 86%, and the comparative genomics approach IslandPick [18,42,44]. The generated results from IslandViewer were further filtered by eliminating the islands situated within two contigs [44]. To predict virulence genes, the RAST-predicted protein sequences were BLAST searched against Virulence Factors Database (VFDB) [45–47] and results were filtered using an in-house Perl scripts to select orthologous genes that are at least 50% sequence identity and 50% sequence completeness.

## Comparative Analysis

The predicted protein sequences of UM\_WWY, *Mycobacterium mageritense* JR2009, *Mycobacterium septicum* DSM44393 and *M. fortuitum* DSM4662 were clustered into orthologs. In order to cluster orthologs into the same group, protein sequences must be at least 50% identity and 50% coverage to each other. To avoid false prediction, the orthologs that are not present in all the strains were further BLAST-searched against the whole-genome sequences to check the existence of the gene.

The same methods were used to identify the species-specific genes in UM\_WWY with the changing of reference sequences from three *M. fortuitum* complex strains to 146 *Mycobacterium* species strains covering 27 different mycobacterial species.

## Supporting Information

**S1 Fig. Structure of a predicted intact prophage.** Different colors of bar indicated different categories of genes in the prophage of UM\_WWY.  
(TIF)

**S2 Fig. Lytic phage screening.** The predicted intact prophage sequence was extracted from the genome of UM\_WWY. Both ends (broken Attachment site (AttP)) of the prophage were manually joined together into circular form (mimicking a lytic phage). All raw sequencing reads were mapped to the joint point of the broken Attachment site. We found 927 paired-end reads covered the joint point, indicating the prophage was likely excised.  
(TIF)

**S3 Fig. Functional classification of the predicted protein-coding genes in the genome of UM\_WWY.**  
(TIF)

**S4 Fig. The order of tRNA genes in the tRNA islands found in the genome of UM\_WWY and four *Mycobacterium abscessus* strains.** The order of tRNA genes are highly similar among the five strains except one tRNA gene in the 4<sup>th</sup> position was missing and one extra tRNA found in the UM\_WWY genome.  
(TIF)

**S1 File. Phylogenetic tree of core genome SNP.**  
(PDF)

**S1 Table. Predicted GIs in the genome of UM\_WWY.** Selected genes of GIs are shown in this table.  
(DOCX)

**S2 Table. Composition of the tRNAs island for UM\_WWY and four *M. abscessus* strains.**  
(DOCX)

**S3 Table. List of *Mycobacterium* genomes used in the clustering of the UM\_WWY genes.**  
(DOCX)

**S4 Table. Predicted virulence genes across the UM\_WWY and 30 other ycobacterial genomes.** “/” means presence of the predicted virulence gene. “-” means absence of the predicted virulence gene.

(DOCX)

## Acknowledgments

We would like to thank the members of Genome Informatics Research Laboratory, University of Malaya for their IT and bioinformatics assistance and inputs in this study. Special thanks to Professor Dr. Ngeow Yun Fong from Faculty of Medicine, University of Malaya for providing DNA material and inputs in this study.

## Author Contributions

Conceived and designed the experiments: SWC. Performed the experiments: WYW. Analyzed the data: WYW TKT. Contributed reagents/materials/analysis tools: WYW TKT. Wrote the paper: SWC WYW TKT NSJ.

## References

1. Falkinham JO (2009) Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. *Journal of Applied Microbiology* 107: 356–367. doi: [10.1111/j.1365-2672.2009.04161.x](https://doi.org/10.1111/j.1365-2672.2009.04161.x) PMID: [19228258](https://pubmed.ncbi.nlm.nih.gov/19228258/)
2. Kamala T, Paramasivan CN, Herbert D, Venkatesan P, Prabhakar R (1994) Evaluation of Procedures for Isolation of Nontuberculous Mycobacteria from Soil and Water. *Applied and Environmental Microbiology* 60: 1021–1024. PMID: [16349208](https://pubmed.ncbi.nlm.nih.gov/16349208/)
3. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, et al. (2007) An official ATS/IDSA statement: Diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *American Journal of Respiratory and Critical Care Medicine* 175: 367–416. PMID: [17277290](https://pubmed.ncbi.nlm.nih.gov/17277290/)
4. Talavlikar R, Carson J, Meatherill B, Desai S, Sharma M, et al. (2011) Mycobacterium senegalense tissue infection in a child after fish tank exposure. *Can J Infect Dis Med Microbiol* 22: 101–103. PMID: [22942887](https://pubmed.ncbi.nlm.nih.gov/22942887/)
5. Lamy B, Marchandin H, Hamitouche K, Laurent F (2008) Mycobacterium setense sp. nov., a Mycobacterium fortuitum-group organism isolated from a patient with soft tissue infection and osteitis. *International Journal of Systematic and Evolutionary Microbiology* 58: 486–490. doi: [10.1099/ijs.0.65222-0](https://doi.org/10.1099/ijs.0.65222-0) PMID: [18218954](https://pubmed.ncbi.nlm.nih.gov/18218954/)
6. Schinsky MF, Morey RE, Steigerwalt AG, Douglas MP, Wilson RW, et al. (2004) Taxonomic variation in the Mycobacterium fortuitum third biovariant complex: description of Mycobacterium boenickei sp. nov., Mycobacterium houstonense sp. nov., Mycobacterium neworleansense sp. nov. and Mycobacterium brisbanense sp. nov. and recognition of Mycobacterium porcinum from human clinical isolates. *Int J Syst Evol Microbiol* 54: 1653–1667. PMID: [15388725](https://pubmed.ncbi.nlm.nih.gov/15388725/)
7. Poh ME, Liam CK, Ng KP, Tan R (2014) Mycobacterium brisbanense species nova isolated from a patient with chronic cavitary lung infection. *Chest* 145: 858–860. doi: [10.1378/chest.13-1952](https://doi.org/10.1378/chest.13-1952) PMID: [24687707](https://pubmed.ncbi.nlm.nih.gov/24687707/)
8. Khurana JL, Jackson CJ, Scott C, Pandey G, Horne I, et al. (2009) Characterization of the phenylurea hydrolases A and B: founding members of a novel amidohydrolase subgroup. *Biochem J* 418: 431–441. doi: [10.1042/BJ20081488](https://doi.org/10.1042/BJ20081488) PMID: [19000034](https://pubmed.ncbi.nlm.nih.gov/19000034/)
9. Rhodes MW, Kator H, McNabb A, Deshayes C, Reyrat JM, et al. (2005) Mycobacterium pseudoshottsii sp. nov., a slowly growing chromogenic species isolated from Chesapeake Bay striped bass (*Morone saxatilis*). *Int J Syst Evol Microbiol* 55: 1139–1147. PMID: [15879246](https://pubmed.ncbi.nlm.nih.gov/15879246/)
10. Brown-Elliott BA, Wallace RJ Jr. (2002) Clinical and taxonomic status of pathogenic nonpigmented or late-pigmenting rapidly growing mycobacteria. *Clin Microbiol Rev* 15: 716–746. PMID: [12364376](https://pubmed.ncbi.nlm.nih.gov/12364376/)
11. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28: 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
12. Townsend TM, Alegre RE, Kelley ST, Wiens JJ, Reeder TW (2008) Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from squamate reptiles.

Molecular Phylogenetics and Evolution 47: 129–142. doi: [10.1016/j.ympev.2008.01.008](https://doi.org/10.1016/j.ympev.2008.01.008) PMID: [18313331](https://pubmed.ncbi.nlm.nih.gov/18313331/)

13. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST server: Rapid annotations using subsystems technology. *Bmc Genomics* 9.
14. Senda K, Arakawa Y, Ichiyama S, Nakashima K, Ito H, et al. (1996) PCR detection of metallo-beta-lactamase gene (blaIMP) in gram-negative rods resistant to broad-spectrum beta-lactams. *J Clin Microbiol* 34: 2909–2913. PMID: [8940421](https://pubmed.ncbi.nlm.nih.gov/8940421/)
15. Choo SW, Wee WY, Ngeow YF, Mitchell W, Tan JL, et al. (2014) Genomic reconnaissance of clinical isolates of emerging human pathogen *Mycobacterium abscessus* reveals high evolutionary potential. *Sci Rep* 4: 4061. doi: [10.1038/srep04061](https://doi.org/10.1038/srep04061) PMID: [24515248](https://pubmed.ncbi.nlm.nih.gov/24515248/)
16. Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annual Review of Microbiology* 54: 641–679. PMID: [11018140](https://pubmed.ncbi.nlm.nih.gov/11018140/)
17. Langille MGI, Hsiao WWL, Brinkman FSL (2010) Detecting genomic islands using bioinformatics approaches. *Nature Reviews Microbiology* 8: 372–382.
18. Langille MGI, Brinkman FSL (2009) IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25: 664–665. doi: [10.1093/bioinformatics/btp030](https://doi.org/10.1093/bioinformatics/btp030) PMID: [19151094](https://pubmed.ncbi.nlm.nih.gov/19151094/)
19. Bastos MC, Murphy E (1988) Transposon Tn554 encodes three products required for transposition. *EMBO J* 7: 2935–2941. PMID: [2846290](https://pubmed.ncbi.nlm.nih.gov/2846290/)
20. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, et al. (2005) Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog* 1: e5. PMID: [16201017](https://pubmed.ncbi.nlm.nih.gov/16201017/)
21. Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, et al. (2007) Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* 24: 1861–1871. PMID: [17545187](https://pubmed.ncbi.nlm.nih.gov/17545187/)
22. Forrellad MA, Klepp LI, Gioffre A, Sabio y Garcia J, Morbidoni HR, et al. (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4: 3–66. doi: [10.4161/viru.22329](https://doi.org/10.4161/viru.22329) PMID: [23076359](https://pubmed.ncbi.nlm.nih.gov/23076359/)
23. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537–+. PMID: [9634230](https://pubmed.ncbi.nlm.nih.gov/9634230/)
24. Sala A, Bordes P, Genevaux P (2014) Multiple Toxin-Antitoxin Systems in *Mycobacterium tuberculosis*. *Toxins* 6: 1002–1020. doi: [10.3390/toxins6031002](https://doi.org/10.3390/toxins6031002) PMID: [24662523](https://pubmed.ncbi.nlm.nih.gov/24662523/)
25. Viator RJ, Rest RF, Hildebrandt E, McGee DJ (2008) Characterization of *Bacillus anthracis* arginase: effects of pH, temperature, and cell viability on metal preference. *BMC Biochem* 9: 15. doi: [10.1186/1471-2091-9-15](https://doi.org/10.1186/1471-2091-9-15) PMID: [18522738](https://pubmed.ncbi.nlm.nih.gov/18522738/)
26. Lin W, Mathys V, Ang EL, Koh VH, Martinez Gomez JM, et al. (2012) Urease activity represents an alternative pathway for *Mycobacterium tuberculosis* nitrogen metabolism. *Infect Immun* 80: 2771–2779. doi: [10.1128/IAI.06195-11](https://doi.org/10.1128/IAI.06195-11) PMID: [22645285](https://pubmed.ncbi.nlm.nih.gov/22645285/)
27. Kvalnes-Krick K, Jorns MS (1986) Bacterial sarcosine oxidase: comparison of two multisubunit enzymes containing both covalent and noncovalent flavin. *Biochemistry* 25: 6061–6069. PMID: [3790506](https://pubmed.ncbi.nlm.nih.gov/3790506/)
28. Goyer A, Johnson TL, Olsen LJ, Collakova E, Shachar-Hill Y, et al. (2004) Characterization and metabolic function of a peroxisomal sarcosine and pipecolate oxidase from *Arabidopsis*. *Journal of Biological Chemistry* 279: 16947–16953. PMID: [14766747](https://pubmed.ncbi.nlm.nih.gov/14766747/)
29. Velmurugan K, Chen B, Miller JL, Azogue S, Gurses S, et al. (2007) *Mycobacterium tuberculosis* nuoG is a virulence gene that inhibits apoptosis of infected host cells. *Plos Pathogens* 3: 972–980.
30. Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley LW (1993) Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* 261: 1454–1457. PMID: [8367727](https://pubmed.ncbi.nlm.nih.gov/8367727/)
31. Astarie-Dequeker C, Le Guyader L, Malaga W, Seaphanh FK, Chalut C, et al. (2009) Phthiocerol dimycocerosates of *M. tuberculosis* participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. *PLoS Pathog* 5: e1000289. doi: [10.1371/journal.ppat.1000289](https://doi.org/10.1371/journal.ppat.1000289) PMID: [19197369](https://pubmed.ncbi.nlm.nih.gov/19197369/)
32. Sinsimer D, Huet G, Manca C, Tsenova L, Koo M-S, et al. (2008) The phenolic glycolipid of *Mycobacterium tuberculosis* differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infection and immunity* 76: 3027–3036. doi: [10.1128/IAI.01663-07](https://doi.org/10.1128/IAI.01663-07) PMID: [18443098](https://pubmed.ncbi.nlm.nih.gov/18443098/)
33. Azad AK, Sirakova TD, Fernandes ND, Kolattukudy PE (1997) Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J Biol Chem* 272: 16741–16745. PMID: [9201977](https://pubmed.ncbi.nlm.nih.gov/9201977/)
34. Raynaud C, Guilhot C, Rauzier J, Bordat Y, Pelicic V, et al. (2002) Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* 45: 203–217. PMID: [12100560](https://pubmed.ncbi.nlm.nih.gov/12100560/)

35. da Silva Rabello MC, Matsumoto CK, de Almeida P, Gonzaga L, Carmen Menendez M, et al. (2012) First description of natural and experimental conjugation between Mycobacteria mediated by a linear plasmid. *PloS one* 7.
36. Nathan C, Shiloh MU (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. *Proceedings of the National Academy of Sciences of the United States of America* 97: 8841–8848. PMID: [10922044](#)
37. Gobert AP, McGee DJ, Akhtar M, Mendz GL, Newton JC, et al. (2001) Helicobacter pylori arginase inhibits nitric oxide production by eukaryotic cells: A strategy for bacterial survival. *Proceedings of the National Academy of Sciences of the United States of America* 98: 13844–13849. PMID: [11717441](#)
38. Amon J, Titgemeyer F, Burkovski A (2009) A Genomic View on Nitrogen Metabolism and Nitrogen Control in Mycobacteria. *Journal of Molecular Microbiology and Biotechnology* 17: 20–29. doi: [10.1159/000159195](#) PMID: [18824837](#)
39. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864. doi: [10.1093/bioinformatics/btr026](#) PMID: [21278185](#)
40. Devulder G, Perriere G, Baty F, Flandrois JP (2003) BIBI, a bioinformatics bacterial identification tool. *Journal of Clinical Microbiology* 41: 1785–1787. PMID: [12682188](#)
41. Waack S, Keller O, Asper R, Brodag T, Damm C, et al. (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *Bmc Bioinformatics* 7.
42. Langille MGI, Hsiao WWL, Brinkman FSL (2008) Evaluation of genomic island predictors using a comparative genomics approach. *Bmc Bioinformatics* 9.
43. Hsiao WWL, Ung K, Aeschliman D, Bryan J, Finlay BB, et al. (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. *Plos Genetics* 1: 540–550.
44. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, et al. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *Fems Microbiology Reviews* 33: 376–393. doi: [10.1111/j.1574-6976.2008.00136.x](#) PMID: [19178566](#)
45. Chen LH, Xiong ZH, Sun LL, Yang J, Jin Q (2012) VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research* 40: D641–D645. doi: [10.1093/nar/gkr989](#) PMID: [22067448](#)
46. Yang J, Chen LH, Sun LL, Yu J, Jin Q (2008) VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. *Nucleic Acids Research* 36: D539–D542. PMID: [17984080](#)
47. Chen LH, Yang J, Yu J, Ya ZJ, Sun LL, et al. (2005) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* 33: D325–D328. PMID: [15608208](#)