PLOS ONE

RESEARCH ARTICLE

Comparative genome sequencing and analyses of *Mycobacterium cosmeticum* reveal potential for biodesulfization of gasoline

Wei Yee Wee[1], **Avirup Dutta**[2], **Jayasyaliny Jayaraj**[1], **Siew Woh Choo**[3,4]*

1 Monash University Malaysia, School of Science, Bandar Sunway, Malaysia, **2** The Novo Nordisk Foundation Center for Basic Metabolic Research, Human Genomics and Metagenomics in Metabolism, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, **3** Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou Dushu Lake Science and Education Innovation District, Suzhou Industrial Park, Suzhou, P. R. China, **4** Suzhou Genome Centre (SGC), Health Technologies University Research Centre (HT-URC), Xi'an Jiaotong-Liverpool University, Suzhou Dushu Lake Science and Education Innovation District, Suzhou Industrial Park, Suzhou, P. R. China

* csw1978@hotmail.com

Citation: Wee WY, Dutta A, Jayaraj J, Choo SW (2019) Comparative genome sequencing and analyses of *Mycobacterium cosmeticum* reveal potential for biodesulfization of gasoline. PLoS ONE 14(4): e0214663. https://doi.org/10.1371/journal.pone.0214663

Editor: Hasnain Seyed Ehtesham, Jamia Hamdard, INDIA

Received: September 25, 2018

Accepted: March 18, 2019

Published: April 9, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This project was under CSW and 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. There was no additional external funding received for this study.

Abstract

Mycobacterium cosmeticum is a nontuberculous *Mycobacterium* recovered from different water sources including household potable water and water collected at nail salon. Individual cases of this bacterium have been reported to be associated with gastrointestinal tract infections. Here we present the first whole-genome study and comparative analysis of two new clinically-derived *Mycobacterium* sp. UM_RHS (referred as UM_RHS after this) and *Mycobacterium* sp. UM_NYF (referred as UM_NYF after this) isolated from patients in Indonesia and Malaysia respectively to have a better understanding of the biological characteristic of these isolates. Both strains are likely *Mycobacterium cosmeticum* as supported by the evidence from molecular phylogenetic, comparative genomic and Average Nucleotide Identity (ANI) analyses. We found the presence of a considerably large number of putative virulence genes in the genomes of UM_RHS and UM_NYF. Interestingly, we also found a horizontally transferred genomic island carrying a putative *dsz* operon proposing that they may have potential to perform biodesulfization of dibenzothiophene (DBT) that may be effective in cost reduction and air pollution during fuel combustion. This comparative study may provide new insights into *M. cosmeticum* and serve as an important reference for future functional studies of this bacterial species.

Introduction

Mycobacterium is a genus under the actinobacteria phylum classified together with other well-known human pathogens like *M. tuberculosis* (causing tuberculosis) and *M. leprae* (causing leprosy) [1–3]. This genus consists of another group of mycobacteria known as the nontuberculous mycobacteria (NTM). NTM has been associated with human diseases and was first reported in pathological human secretions in 1884 [4]. It represents a diverse group of

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

environmentally opportunistic human pathogens widely found at peat-rich potting soil, drinking water in buildings and households, on animals and also in food [5–7]. The NTM can cause human infections mainly occurring under environmental exposures, sternal wound infections, plastic surgery wound infections, or post-injection abscesses [8–10].

M. cosmeticum is usually recovered from water including household potable water [11] and water collected at the nail salon [12] and activated sludge from wastewater treatment [13]. A case where *M. cosmeticum* has been implicated as a gastrointestinal tract pathogen causing ascites in a 63-year-old woman [14] has been reported. Moreover, it has also been reported that this bacterium has induced severe diffuse granulomatous colitis in a non-immunocompromised 32-year-old Turkish patient [15].

In this paper, we sequenced two new clinically-derived *Mycobacterium* sp. UM_RHS (referred as UM_RHS after this) and *Mycobacterium* sp. UM_NYF (referred as UM_NYF after this). We have also performed bioinformatics analyses particularly, comparative analyses to further understand the genomics, phylogeny and biology of this bacterial species. The genome sequences of UM_RHS and UM_NYF have been deposited at GenBank with the accession numbers of GCA_000455185.1 and GCA_000987455.1, respectively.

Results and discussion

Genome sequencing and assembly

The genome of UM_RHS sequenced using Illumina HiSeq 2000 sequencing technology yielded 51,391,676 paired-end (PE) reads. 50,813,644 usable reads were obtained after quality based trimming with Phred score of 20 and removal of exact duplicates and reverse complement duplicate reads using PRINSEQ lite version 0.20 [16]. The *de novo* assembly of these reads generated 167 contigs with a total genomic length of 6,775,899bp and G+C content of 67.9%. This UM_RHS assembly has a N50 value of 95,298bp with minimum contig length of 510bp and maximum contig size of 242,034bp, suggesting considerably high quality of this assembly for downstream analyses.

Similar to the UM_RHS, the genome of UM_NYF sequenced with the same sequencing platform yielded 39,868,088 raw PE reads. After the filtering steps, 39,529,963 preprocessed reads were used for assembly. The assembly of UM_NYF genome resulted in 332 contigs with a total genomic length of 6,809,253bp and a G+C content of 67.9% and N50 value of 61,947bp.

Recognition of Isolated Species

To determine the taxonomic positions of the UM_RHS and UM_NYF, we first constructed phylogenetic trees using housekeeping genes. We constructed a 16S rRNA-based phylogenetic tree using UM_RHS, UM_NYF and other mycobacterial species (Fig 1). Our data suggested that our strains were closely related to *M. cosmeticum* with four mismatches and 99% sequence similarities to the reference *M. cosmeticum* DSM44829. The *16S rRNA* gene-based tree has been used to separate between the rapid and slow growing mycobacteria [17]. As anticipated, our data clearly separated the two distinct groups and suggested that both UM_RHS and UM_NYF are rapid growing mycobacteria (Fig 1).

The possibility of our strains being *M. cosmeticum* is further supported by a supermatrix tree constructed using multiple genes: *hsp65*, *rpoB*, *tuf*, *sodA* and *16S rRNA*. This approach would produce more robust tree compared to the single gene approach (Fig 2) [18]. Our supermatrix tree showed that both UM_RHS and UM_NYF shared highest genome similarities (98.0%) with the reference *M. cosmeticum* strain, again acting as an evidence that UM_RHS and UM_NYF are likely *M. cosmeticum*.

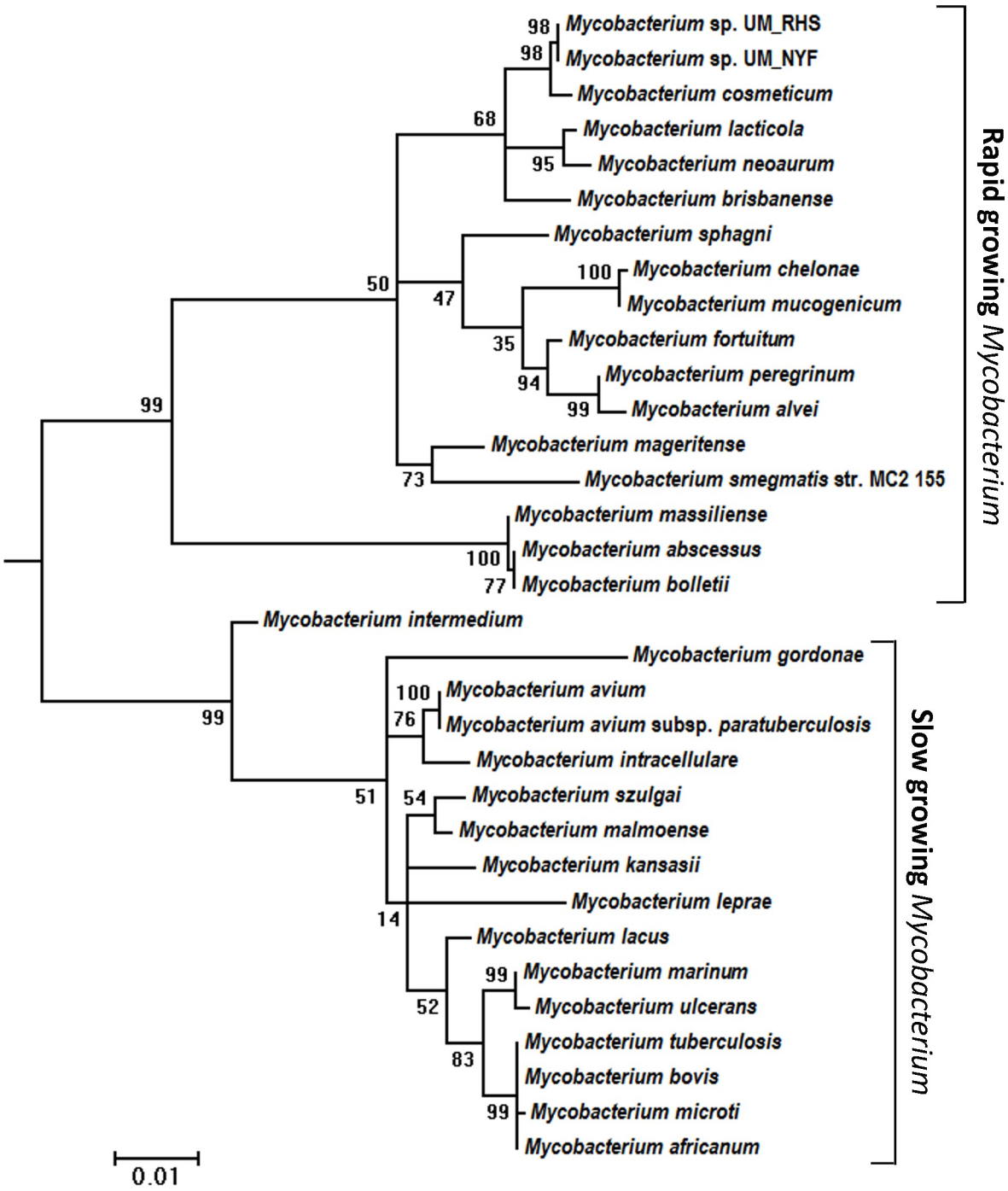

Fig 1. 16S rRNA based phylogeny analysis of 32 Mycobacteria strains belonging to different species: Differentiation into rapid and lowly growing Mycobacteria. The *16S rRNA* gene-based phylogenetic tree clearly distinguishes *Mycobacterium* grouping based on the growth rates. UM_RHS and UM_ NYF were clustered in the rapid growing mycobacterial group.

https://doi.org/10.1371/journal.pone.0214663.g001

To further confirm the identity of UM_RHS and UM_NYF, we also performed ANI analysis using whole-genome data. The ANI is one of the most robust measurements of genomic relatedness between bacterial strains [19], and has a great potential in the taxonomy classification of bacteria as a substitute for the traditional labor intensive DNA-DNA hybridization

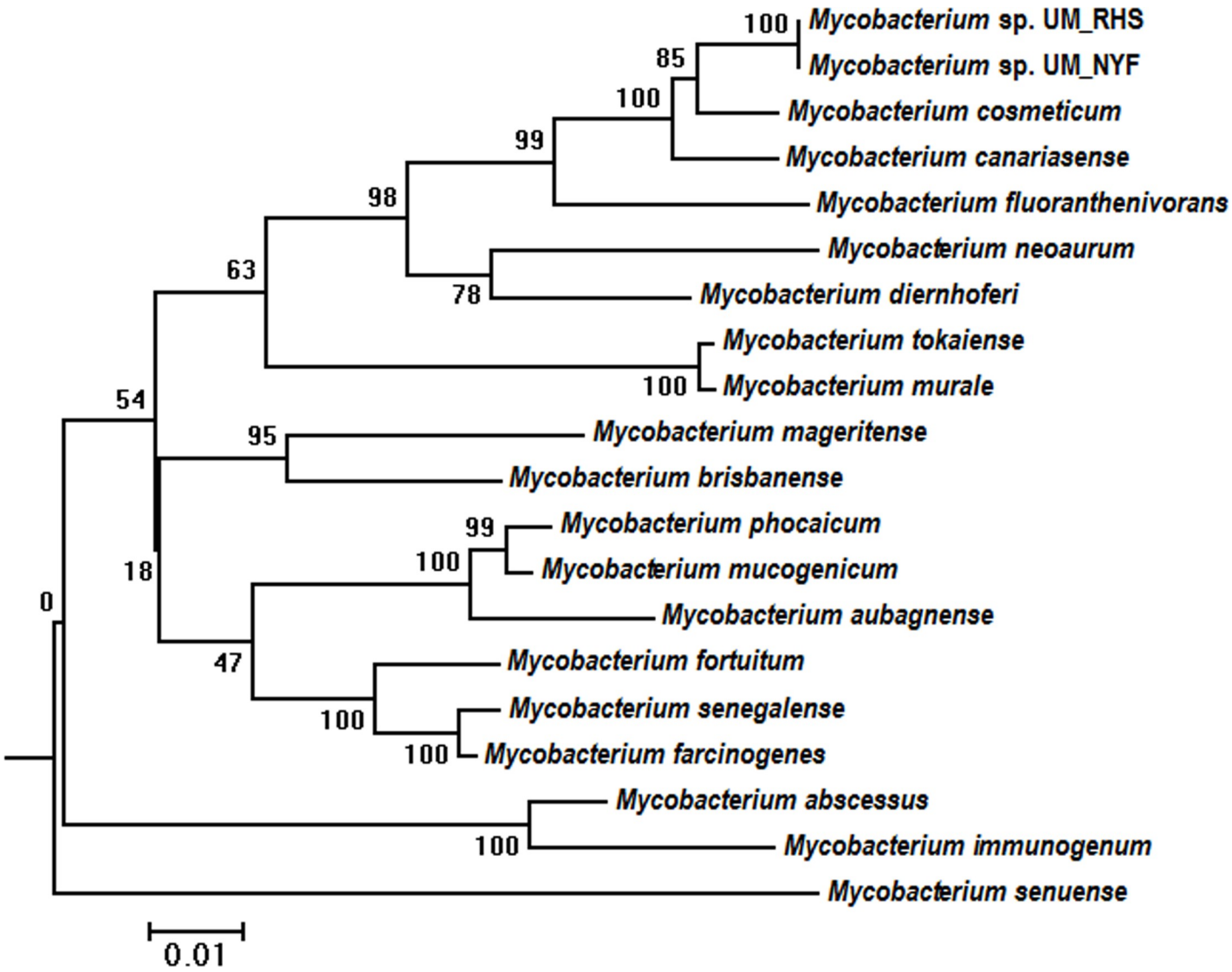

Fig 2. Supermatrix tree of five markers (*hsp65*, *rpoB*, *tuf*, *sodA* and *16S rRNA*). UM_RHS and UM_NYF are closest to *M. cosmeticum*, supported by a high bootstrap value of 85%.

https://doi.org/10.1371/journal.pone.0214663.g002

(DDH) technique. The algorithm designed is based on the calculation of average percentage of whole-genome sequence similarity between a pair of bacterial genome. An ANI threshold of 95% determined for species demarcation has previously been suggested based on intensive comparative investigations [19].

To calculate the ANI values (in percentage), we compared the UM_RHS and UM_NYF separately to other 35 *Mycobacterium* species (representative strains for known *Mycobacterium* species) that have genome sequences available in the National Center for Biotechnology Information (NCBI) GenBank depository.

The ANI values of each pairwise genome comparison varied from 70% to 99% of sequence identity (Fig 3). The ANI values of UM_RHS and UM_NYF against other 34 strains (excluded the reference strain of *M. cosmeticum* DSM44829) ranged from 70% to 78% (below the cut-off to define a species), suggesting that our strains do not belong to these known species. However,

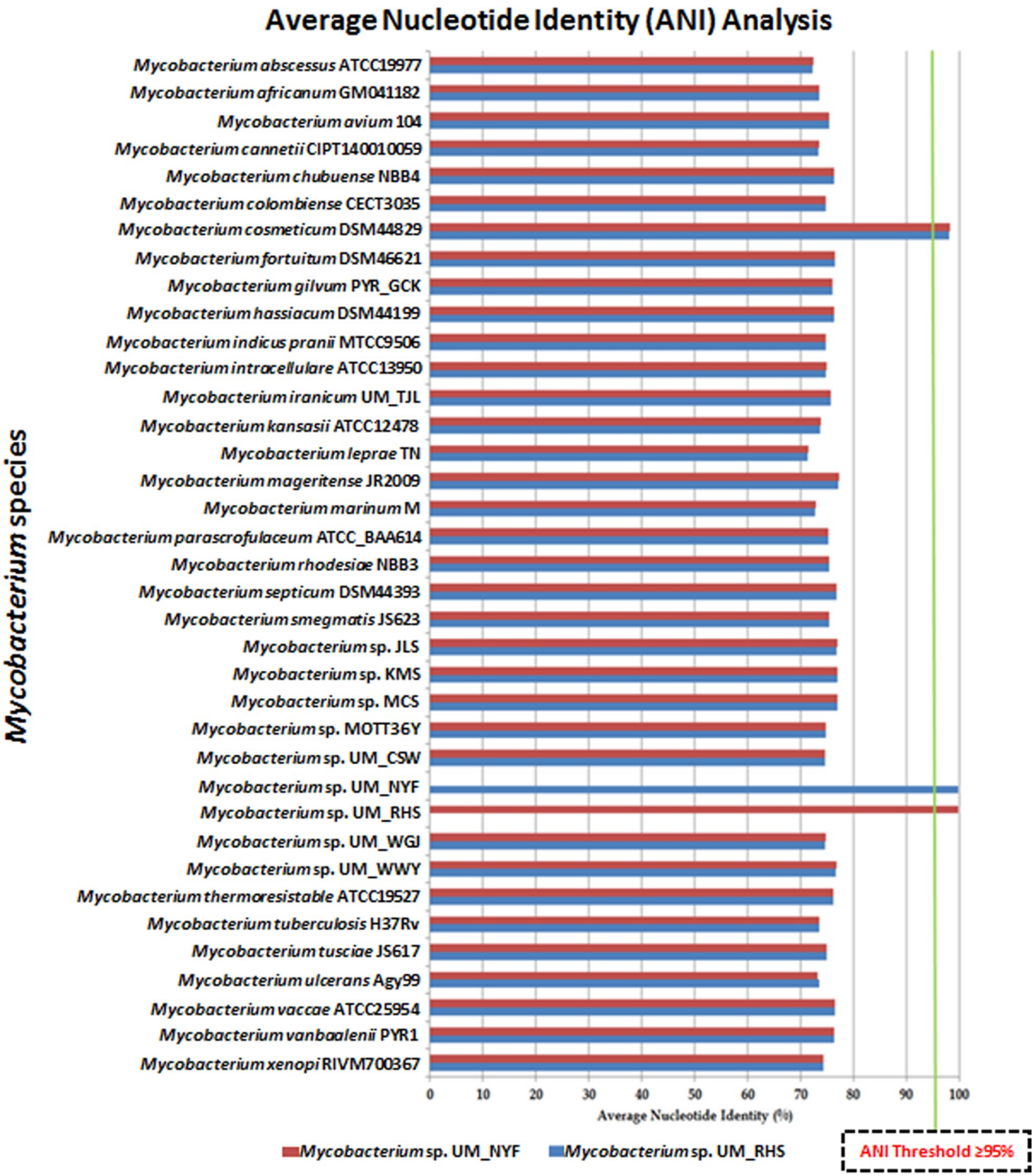

Fig 3. Average nucleotide analysis for 37 *Mycobacterium* species. The ANI values of UM_NYF and UM_RHS against *M. cosmeticum* DSM44829 are above 95%, supporting that these strains belong to the species *M. cosmeticum*.

https://doi.org/10.1371/journal.pone.0214663.g003

when comparing against the reference *M. cosmeticum* DSM44829, the UM_RHS and UM_NYF have ANI values of 98.14% and 98.16% respectively, again supporting our view that both UM_RHS and UM_NYF are likely *M. cosmeticum*.

Here we have successfully sequenced and analyzed the genomes of probable two new members of *M. cosmeticum*, UM_RHS and UM_NYF, which were isolated from patients in Indonesia and Malaysia respectively. The taxonomic position of the two UM_RHS and UM_NYF has been supported by evidence from phylogenetic and ANI analysis.

Gene prediction and annotation

As anticipated, both UM_RHS and UM_NYF genomes generally share similar genomic features such as the genome size, number of protein-coding genes and RNA as predicted by the Rapid Annotation using Subsystem Technology (RAST) pipeline [20]. For instance, the UM_RHS has a genome size of 6,780,714bp with 6,608 protein-coding genes and 49 RNA genes, whereas UM_NYF has a genome size of 6,809,253bp with 6,579 protein-coding genes and 50 RNA genes (Table 1). Both genomes have single copy of rRNA operons. The summary of functional assignments of the RAST-predicted protein-coding genes for the UM_RHS, UM_NYF and *M. cosmeticum* DSM44829 are shown in Fig 4. As anticipated, all the 3 genomes generally share very similar functional distributions since they are most probably belonging to the same species.

Functional annotation showed that most of the genes were involved in functional categories such as amino acid and derivatives, carbohydrates, cofactors, vitamins, fatty acids, lipids and isoprenoids, which are responsible for basic functions of bacteria. No plasmids were predicted in either of the genomes. RAST predicted the presence of a number of genes encoding integrase, transposase like proteins, phage like proteins and mobile element proteins in both genomes (Table 2). There are about 136 genes categorized under the virulence, disease and defence.

However, there are subtle differences between the UM_RHS and UM_NYF in the categories of RNA metabolism, and cell wall and capsule. For instance, UM_NYF genome has 107 genes involved in the RNA metabolism, which is relatively higher compared to the UM_RHS genome (72 genes) (S1 Table and S2 Table). Interestingly, further examination of these genes revealed that UM_NYF possessed many genes in two extra sub-categories: tRNA modification bacteria (26 genes) and *16S rRNA* modification within the P site of its ribosome (6 genes), which were absent in UM_RHS. Out of these six genes categorized under the *16S rRNA* modification within the P site of its ribosome, we identified two methyltransferases, *RsmH* and *RsmI* responsible for the N^4-methylation and 2'-*O*-methylation in the 16S rRNA [21]. These two genes can stabilize the local structure and interaction of the ribosome P-site to accommodate the codon-anticodon helix[21]. Kimura and Suzuki showed that deletions of *rsmH* or *rsmI* can affect the efficiency of non-AUG initiation and the fidelity of translation. Thus, the absence of these two genes in UM_RHS could affect its decoding fidelity [21].

Comparative genome analysis of *M. cosmeticum*

To better understand the genomic structure of *M. cosmeticum*, we compared UM_RHS and UM_NYF to a reference genome, *M. cosmeticum* DSM 44829 at genome and gene levels. At the genome level, we aligned and reordered the genome sequences of the three strains using Mauve software with *M. cosmeticum* DSM 44829 as a reference [22]. We found that the three *M. cosmeticum* genomes were generally similar or conserved as most of the genome regions for the three strains were nicely aligned to each other (Fig 5). The UM_RHS and UM_NYF mapped approximately 92% of the reference genome with a high sequence identity of 98%. However, the genome size of UM_RHS and UM_NYF was comparatively larger (~0.4Mbp) than the reference genome which has a genome size of about 6.4Mbp. Both UM_RHS and UM_NYF have larger genome size probably due to the presence of considerably large number of horizontally transferred genomic islands found in both genomes which we will discussed below.

Table 1. RNAs identified by RAST in the genomes of UM_NYF and UM_RHS.

	UM_NYF	RAST Feature ID	UM_RHS	RAST Feature ID
1	tRNA-Leu-TAA	fig|6666666.28483.rna.1	tRNA-Pro-GGG	fig|6666666.28480.rna.1
2	tRNA-Ile-GAT	fig|6666666.28483.rna.2	tRNA-Asn-GTT	fig|6666666.28480.rna.2
3	tRNA-Ala-TGC	fig|6666666.28483.rna.3	tRNA-Lys-CTT	fig|6666666.28480.rna.3
4	tRNA-Leu-CAG	fig|6666666.28483.rna.4	tRNA-Arg-CCG	fig|6666666.28480.rna.4
5	tRNA-Pro-TGG	fig|6666666.28483.rna.5	tRNA-Thr-CGT	fig|6666666.28480.rna.5
6	tRNA-Gly-TCC	fig|6666666.28483.rna.6	tRNA-Tyr-GTA	fig|6666666.28480.rna.6
7	tRNA-Trp-CCA	fig|6666666.28483.rna.7	tRNA-Met-CAT	fig|6666666.28480.rna.7
8	tRNA-Met-CAT	fig|6666666.28483.rna.8	tRNA-Met-CAT	fig|6666666.28480.rna.8
9	tRNA-Thr-GGT	fig|6666666.28483.rna.9	tRNA-Ala-GGC	fig|6666666.28480.rna.9
10	tRNA-Arg-ACG	fig|6666666.28483.rna.10	tRNA-Val-GAC	fig|6666666.28480.rna.10
11	tRNA-Pseudo-GCT	fig|6666666.28483.rna.11	tRNA-Cys-GCA	fig|6666666.28480.rna.11
12	tRNA-Ser-TGA	fig|6666666.28483.rna.12	tRNA-Gly-GCC	fig|6666666.28480.rna.12
13	tRNA-His-GTG	fig|6666666.28483.rna.13	tRNA-Val-CAC	fig|6666666.28480.rna.13
14	tRNA-Tyr-GTA	fig|6666666.28483.rna.14	5S RNA	fig|6666666.28480.rna.14
15	tRNA-Gln-CTG	fig|6666666.28483.rna.15	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	fig|6666666.28480.rna.15
16	tRNA-Glu-CTC	fig|6666666.28483.rna.16	Small Subunit Ribosomal RNA; ssuRNA; SSU rRNA	fig|6666666.28480.rna.16
17	tRNA-Ala-CGC	fig|6666666.28483.rna.17	tRNA-Cys-GCA	fig|6666666.28480.rna.17
18	tRNA-Met-CAT	fig|6666666.28483.rna.18	tRNA-Leu-CAA	fig|6666666.28480.rna.18
19	tRNA-Gln-TTG	fig|6666666.28483.rna.19	tRNA-Leu-TAG	fig|6666666.28480.rna.19
20	tRNA-Val-TAC	fig|6666666.28483.rna.20	tRNA-Ser-GGA	fig|6666666.28480.rna.20
21	tRNA-Lys-CTT	fig|6666666.28483.rna.21	tRNA-Ser-CGA	fig|6666666.28480.rna.21
22	tRNA-Met-CAT	fig|6666666.28483.rna.22	tRNA-Pro-TGG	fig|6666666.28480.rna.22
23	5S RNA	fig|6666666.28483.rna.23	tRNA-Gly-TCC	fig|6666666.28480.rna.23
24	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	fig|6666666.28483.rna.24	tRNA-Arg-TCT	fig|6666666.28480.rna.24
25	Small Subunit Ribosomal RNA; ssuRNA; SSU rRNA	fig|6666666.28483.rna.25	tRNA-Ile-GAT	fig|6666666.28480.rna.25
26	tRNA-Leu-GAG	fig|6666666.28483.rna.26	tRNA-Ala-TGC	fig|6666666.28480.rna.26
27	tRNA-Leu-CAA	fig|6666666.28483.rna.27	tRNA-Leu-TAA	fig|6666666.28480.rna.27
28	tRNA-Thr-TGT	fig|6666666.28483.rna.28	tRNA-Phe-GAA	fig|6666666.28480.rna.28
29	tRNA-Ala-GGC	fig|6666666.28483.rna.29	tRNA-Asp-GTC	fig|6666666.28480.rna.29
30	tRNA-Asn-GTT	fig|6666666.28483.rna.30	tRNA-Glu-TTC	fig|6666666.28480.rna.30
31	tRNA-Thr-CGT	fig|6666666.28483.rna.31	tRNA-Lys-TTT	fig|6666666.28480.rna.31
32	tRNA-Ser-CGA	fig|6666666.28483.rna.32	tRNA-Arg-CCT	fig|6666666.28480.rna.32
33	tRNA-Ser-GGA	fig|6666666.28483.rna.33	tRNA-His-GTG	fig|6666666.28480.rna.33
34	tRNA-Arg-CCT	fig|6666666.28483.rna.34	tRNA-Leu-GAG	fig|6666666.28480.rna.34
35	tRNA-Leu-CAA	fig|6666666.28483.rna.35	tRNA-Thr-TGT	fig|6666666.28480.rna.35
36	tRNA-Cys-GCA	fig|6666666.28483.rna.36	tRNA-Ser-TGA	fig|6666666.28480.rna.36
37	tRNA-Gly-CCC	fig|6666666.28483.rna.37	tRNA-Pseudo-GCT	fig|6666666.28480.rna.37
38	tRNA-Arg-CCG	fig|6666666.28483.rna.38	tRNA-Arg-ACG	fig|6666666.28480.rna.38
39	tRNA-Pro-GGG	fig|6666666.28483.rna.39	tRNA-Thr-GGT	fig|6666666.28480.rna.39
40	tRNA-Leu-TAG	fig|6666666.28483.rna.40	tRNA-Met-CAT	fig|6666666.28480.rna.40
41	tRNA-Pro-CGG	fig|6666666.28483.rna.41	tRNA-Trp-CCA	fig|6666666.28480.rna.41
42	tRNA-Val-CAC	fig|6666666.28483.rna.42	tRNA-Val-TAC	fig|6666666.28480.rna.42
43	tRNA-Gly-GCC	fig|6666666.28483.rna.43	tRNA-Ala-CGC	fig|6666666.28480.rna.43
44	tRNA-Cys-GCA	fig|6666666.28483.rna.44	tRNA-Glu-CTC	fig|6666666.28480.rna.44
45	tRNA-Val-GAC	fig|6666666.28483.rna.45	tRNA-Gln-CTG	fig|6666666.28480.rna.45
46	tRNA-Arg-TCT	fig|6666666.28483.rna.46	tRNA-Pro-CGG	fig|6666666.28480.rna.46
47	tRNA-Lys-TTT	fig|6666666.28483.rna.47	tRNA-Gly-CCC	fig|6666666.28480.rna.47

(*Continued*)

Table 1. (Continued)

	UM_NYF	RAST Feature ID	UM_RHS	RAST Feature ID
48	tRNA-Glu-TTC	fig|6666666.28483.rna.48	tRNA-Gln-TTG	fig|6666666.28480.rna.48
49	tRNA-Asp-GTC	fig|6666666.28483.rna.49	tRNA-Leu-CAG	fig|6666666.28480.rna.49
50	tRNA-Phe-GAA	fig|6666666.28483.rna.50		

https://doi.org/10.1371/journal.pone.0214663.t001

At the gene level, we clustered all RAST-predicted genes of the three strains using BLASTClust (http://ftp.ncbi.nih.gov/blast/documents/blastclust.html) which resulted in non-redundant 6,957 orthologous gene families. Our data clearly showed that the three strains shared a high number of common gene families (5,657), which accounted for 81.3% of the total gene families suggesting that they are considerably conserved among them (Fig 6). Interestingly, we found a relatively higher number of strain-specific gene families (333) in DSM 44829 compared to UM_RHS (87) and UM_NYF (33). Of these 333 genes, 89 (26.7%) were believed to be inserted into the *M. cosmeticum* DSM 44829 genome as these genes were found inside the predicted horizontally transferred genomic islands in the *M. cosmeticum* DSM 44829. In addition, we also found two putative *M. cosmeticum* DSM 44829 specific-genes, Type I restriction modification enzymes that may play important roles in the defense mechanism of the bacteria [23].

Another interesting observation is that 558 gene families were present only in both UM_RHS and UM_NYF but not in the *M. cosmeticum* DSM 44829. The high number of

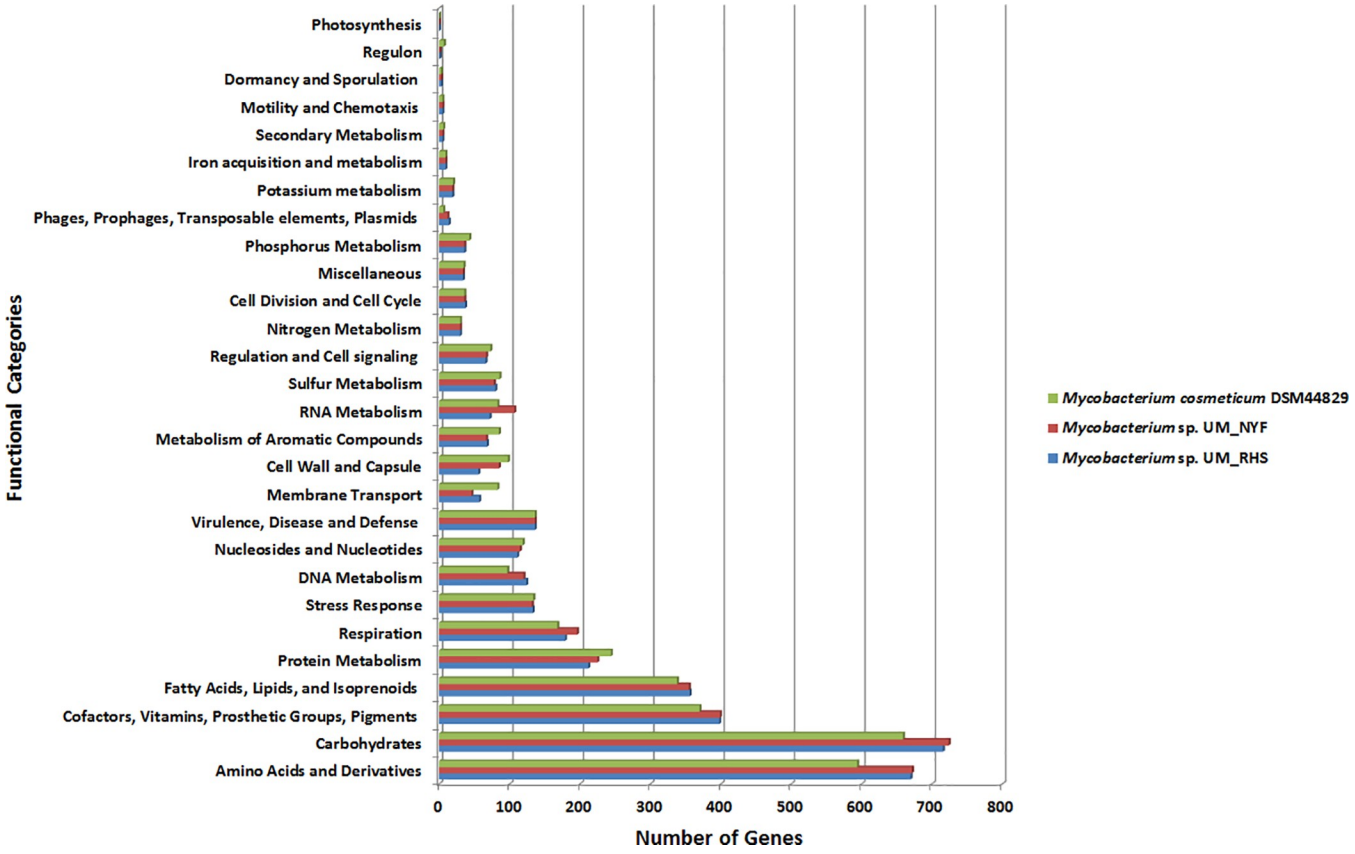

Fig 4. RAST functional categories of UM_RHS and UM_NYF genes. Number of genes in UM_RHS in comparison to UM_NYF which belong to specific RAST functional categories with *M. cosmeticum* DSM 44829 as the reference genome.

https://doi.org/10.1371/journal.pone.0214663.g004

Table 2. RAST predicted genes related to gene transfer in the genomes of UM_NYF and UM_RHS.

Types of genes encoding	UM_RHS	UM_NYF
Integrase	4	4
Mobile element proteins	37	34
Phage like proteins	18	18
Transposases like proteins	6	3

https://doi.org/10.1371/journal.pone.0214663.t002

shared genes between UM_RHS and UM_NYF is probably due to the horizontal gene transfer supported by the observation that at least 154 of these genes are (28%) found within the predicted genomic islands. These specific genes may have arisen from the ancestors of UM_RHS and UM_NYF and probably contribute to the unique traits/phenotypes of these strains which are absent in the previously reported *M. cosmeticum* DSM 44829. Among these specific genes, five specific genes together with other four non-specific genes formed a cluster of nine genes (~12 Kbp) encoding enzymes which metabolize acetone and acetoacetate to acetyl-CoA (Acetone utilization pathway) [24]. Similar gene cluster has also been found in *Helicobacter pylori* and *Helicobacter acinonychus* strains [24]. However, the *Helicobacter pylori* strains contain only a cluster of eight genes lacking one gene (*acxD*) compared to the UM_RHS and UM_NYF, harboring four *acx* genes (*acxABCD*), *scoA*, *scoB*, *fadA* and two hypothetical proteins. The presence of the gene cluster in the sequenced genomes of UM_RHS and UM_NYF may indicate the capability of these strains to metabolize acetone to acetyl-CoA and feed into the TCA cycle, thus providing them with energy [24].

Furthermore, we wanted to identify genes that are specific to *M. cosmeticum*, but not in other *Mycobacterium* species. To examine this, we further clustered the 5,657 common gene families of UM_RHS and UM_NYF with genes from other known mycobacterial genomes belonging to 27 different mycobacterial species. We successfully identified 552 gene families specific to *M. cosmeticum* but not present in other mycobacterial species that we examined. Among these specific genes, we found a gene operon involved in the sorbitol (glucitol) specific phosphoenolpyruvate-dependent sugar phosphotransferase (PTS) system, which has three components (EIIA, EIIB and EIIC) [25, 26].

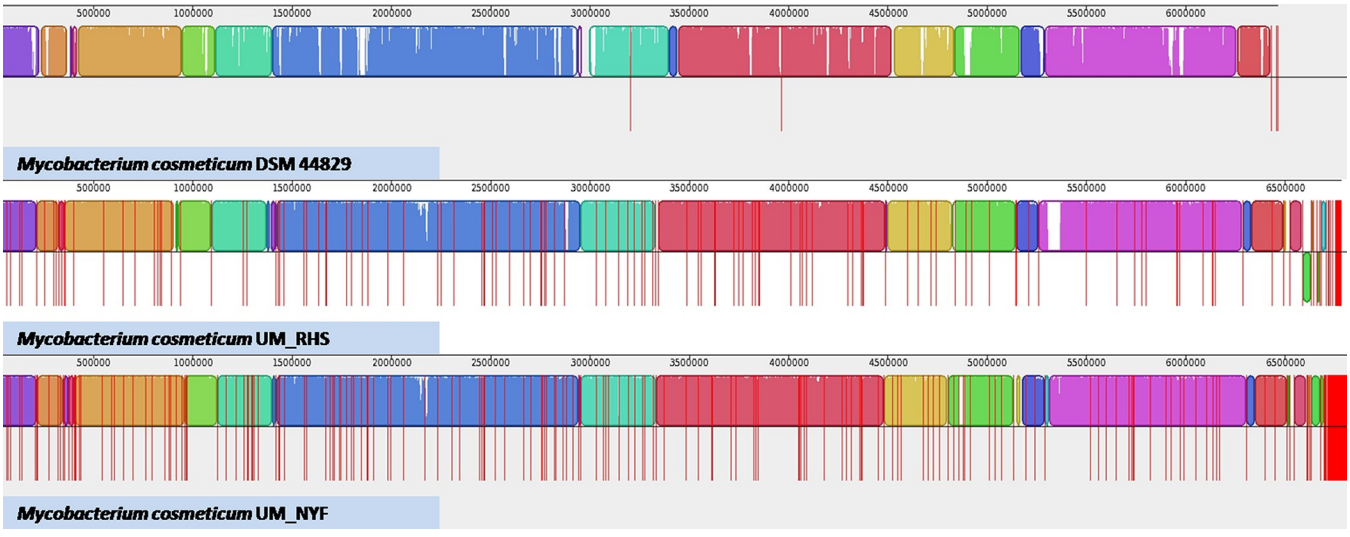

Fig 5. Genomic structure of *M. cosmeticum* genomes. The genome structures are generally conserved among three studied genomes (UM_RHS, UM_NYF and DSM44829).

https://doi.org/10.1371/journal.pone.0214663.g005

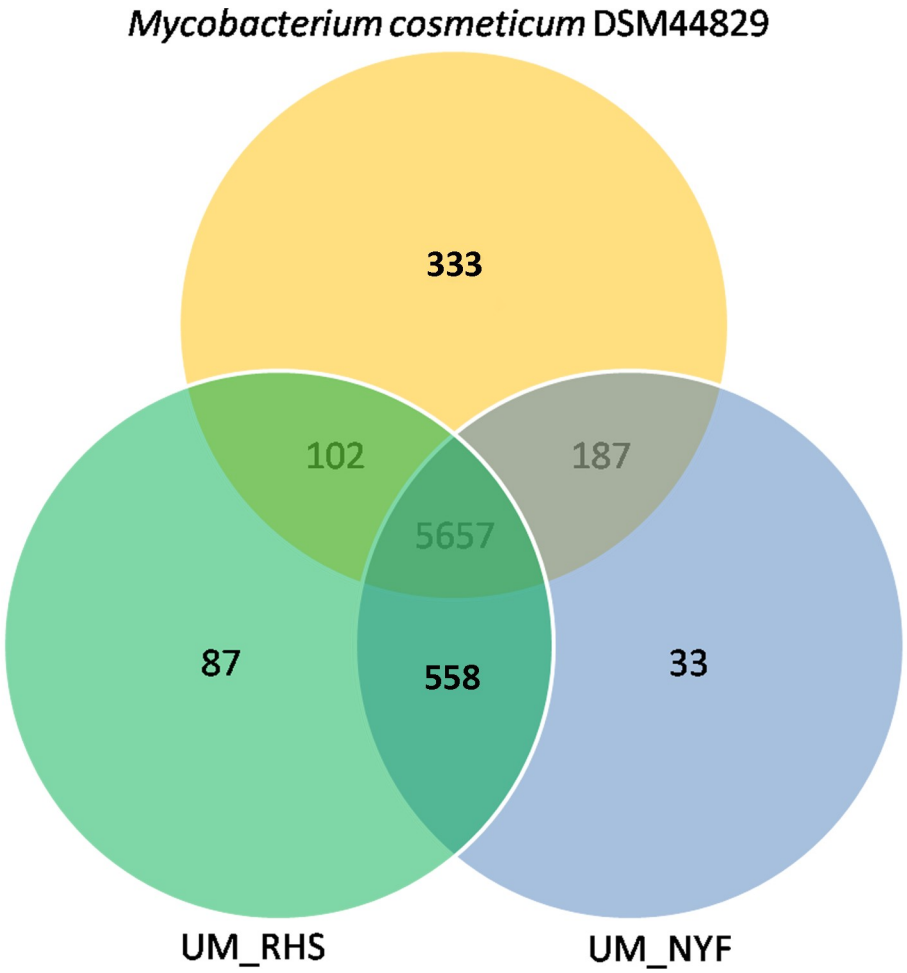

Fig 6. Gene family distribution. The *M. cosmeticum* DSM 44829, UM_RHS and UM_NYF have generally shared a high number of common gene families.

https://doi.org/10.1371/journal.pone.0214663.g006

Comparative genomic islands analysis

To predict horizontally transferred genomic islands, all genome sequences used in this analysis were uploaded to the IslandViewer server [27, 28]. A total of 41 putative genomic islands were found in the two genomes with genomic sizes ranging from 4k to 32kbp (S3 Table). Both strains shared 34 common genomic islands by which five are specific to UM_RHS and two other specific to UM_NYF. A high number of putative genomic islands were present in the two genomes indicating that horizontal gene transfer events might have played significant roles in reshaping the genomes of UM_RHS and UM_NYF throughout the evolutionary period. Among these common genomic islands, some have harbored putative virulence genes which could contribute to the virulence of the two strains. For instance, one of the genomic islands (GI23) harbored two putative virulence genes, *espR* and *phoP*. Another genomic island, GI28, has a virulence gene *sigH*.

Furthermore, we found a genomic island which is believed to have originated from *Mycobacterium goodie* X7B being inserted into both UM_RHS and UM_NYF during the evolutionary period. This genomic island contains the *dsz* (dibenzothiophene biodesulfurization) operon harboring three desulfurization genes (*dszA*, *dszB*, *dszC)*. Combustion of sulfur-containing compounds can cause adverse effects on health and environment [29]. Benzothiophene

(BTH) and dibenzothiophene (DBT) account for more than 50% of the sulfur content of diesel. Current industries usually apply hydro-desulfurization which requires high temperature and pressure. However, biodesulfurization is an environmental friendly method to eliminate sulfur from the refractory organic compound. The major pathway of DBT desulfurization has been reported as the "4S pathways". The "4S pathway" includes four steps which will catalyze DBT into sulfoxide (DBTO), sulfone ($DBTO_2$), sulfinate (HPBSi) and hydroxybiphenyl (HBP). These catalytic reactions are carried out by three Dsz enzymes namely DszA, DszB and DszC encoded by the *dsz* operon. Thus, we postulate that the acquisition of this genomic island possibly from *M. goodie* X7B might have given both UM_RHS and UM_NYF the capability to catalyze the DBT desulfurization.

Besides, three large common genomic islands associated with prophages were identified by screening the genomic regions using PHAST software, a software to predict prophage sequences in bacterial genomes or plasmids [30]. As predicted by PHAST, two are intact prophages and the remaining one is an incomplete prophage in the genomes of UM_RHS and UM_NYF. However, the origin of these prophages are yet to be identified as no hit could be found from BLAST search of these sequences against the NCBI databases.

Virulence gene analyses

Our analysis revealed that both UM_RHS and UM_NYF share 117 similar putative virulence genes with UM_RHS having one extra virulence gene (*secA*) resulting in a total of 118 putative virulence genes in UM_RHS and 117 in UM_NYF (Fig 7). Furthermore, most of the virulence genes found in both genomes were orthologs to well-known human pathogens like *M. tuberculosis*. Gey van Pittius and colleagues have shown that the *ESX-5* genes cluster, a type VII secretion system, is able to separate the rapid and slow growing mycobacteria. They showed that this *ESX-5* genes cluster is only present in the slow growing mycobacteria [31]. Likewise, our strains UM_RHS and UM_NYF do not have the ESX-5 gene cluster, further supporting that the two strains are rapid growing mycobacteria and this is also consistent with the results from the phylogenetic analysis. The data showed that UM_RHS and UM_NYF contain three *ESX* gene clusters (ESX1, 3, 4) although a few genes were missing in these clusters probably due to gene deletion events. However, since both UM_RHS and UM_NYF are draft genomes, additional experiments e.g. PCR are crucial to further confirm the existence of these genes.

Apart from the *ESX* gene clusters, both UM_RHS and UM_NYF genomes showed presence of other putative virulence genes such as *ahpC*, *katG*, *sodC* responsible for enhancing resistance against host toxic compounds, whereas *nuoG* and *sodA* which are involved in evading apoptosis [32]. Some of these genes (*ahpC*, *katG* and *sodC*) encode enzymes important for detoxification of bacteria killing component like reactive oxygen species (ROS) and reactive nitrogen species (RNS), whereas the *nuoG* gene encodes for protein involving the inhibition of extrinsic TNF-α- dependent apoptosis pathway [33]. Furthermore, we also identified virulence genes (*fbp*, *erp*, *hbhA*, *mce*) which encode for cell envelope proteins [34]. These proteins are important in the mycobacterial cell wall maintenance, adhesion and transportation of materials and also survival of mycobacteria in the host cells [34]. Both UM_RHS and UM_NYF also harbored three *fbp* genes (*fbpA*, *fbpB* and *fbpC*) encoding for the antigen 85 complex [34]. This complex is known to be related to the pathogenesis of mycobacteria, for example, by promoting the entry of bacteria into host cells through the binding of fibronectin [35].

In order to survive in host cells, mycobacteria have to adapt to a wide range of environments, stressors and growth condition [34]. Two component regulatory signal transduction systems are important in allowing bacteria to adapt to a variety of environmental stimuli. Interestingly, four pairs of two-component systems were found (PhoP-PhoR, DevR-DevS,

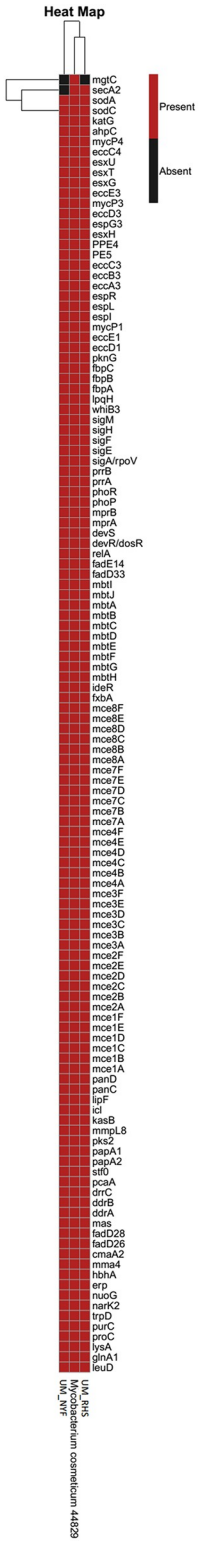

Fig 7. Predicted virulence genes in the genomes of *M. cosmeticum* DSM44829, UM_RHS and UM_NYF.

https://doi.org/10.1371/journal.pone.0214663.g007

MprA-MprB and PrrA-PrrB) in the genomes of UM_RHS and UM_NYF. Previous studies have demonstrated that disruption of the *phoP-phoR* operon can affect the replication of *M. tuberculosis* in the cellular and animal models [36, 37]. This gene operon is also involved in the regulation of genes with potential roles in mycobacterial virulence by regulating the expression of both *espB* and *espR* genes in the ESX-1 secretion system [38].

Overall, both genomes are structurally conserved although differences in the genomic islands and virulence genes of these genomes can be observed. We did not notice any large genome inversions in any of the *M. cosmeticum* strains. The number of shared genes between UM_RHS and UM_NYF were higher compared to that of UM_RHS and *M. cosmeticum* DSM44829 or UM_NYF and *M. cosmeticum* DSM44829, probably reflecting the fact that both UM_RHS and UM_NYF are highly similar because they were isolated from very close geographical regions (Indonesia and Malaysia respectively), whereas the *M. cosmeticum* DSM44829 was isolated from a granulomatous lesion of a female patient in Venezuela [12] [https://www.dsmz.de/catalogues/details/culture/DSM-44829.html].

Both the genomes of UM_RHS and UM_NYF showed presence of sorbitol (glucitol) specific phosphoenolpyruvate-dependent sugar phosphotransferase (PTS) system [25, 26]. Sorbitol is a sugar substitute used when caries formation occurs in the presence of readily fermentable carbohydrate like sucroses. We hypothesize that in the presence of these genes, UM_RHS and UM_NYF could be capable of utilizing sugars using the PTS system.

One interesting finding in this study is the presence of *dsz* (dibenzothiophene biodesulfurization) operon in the horizontally transferred genomic island in the genomes of UM_RHS and UM_NYF. Our data suggested that this genomic region might have originated from *M. goodie* X7B, which has the capability to desulfurize benzothiophene (BTH) and dibenzothiophene (DBT) through the BTH degradation pathway [39]. Combustion of fossil fuel such as petroleum releases sulfur oxides causing air pollution [40, 41]. Deep desulfurization of gasoline can reduce sulfur content, but the conventional hydrodesulfurization (HDS) technology of gasoline results in a significant reduction of the octane number [42]. The biodesulfization of gasoline can serve as an alternative method of sources which not only avoids octane degradation but is also less expensive as compared to the HDS method [39]. Therefore, we postulate that the UM_RHS and UM_NYF may be able to be used for performing biodesulfization of gasoline or with its additionally potential to reduce the expenses and air pollution during fuel combustion. However, further experiments are needed to confirm the usability of this bacterial species for these purposes.

In comparison to the reference genome, *M. cosmeticum* DSM44829, we found relatively higher number of genes and genomic islands in the genomes of UM_RHS and UM_NYF. Intriguingly, our comparative analysis revealed two genes encoding for Type I restriction modification enzymes in *M. cosmeticum* DSM44829, whereas the UM_RHS and UM_NYF lack these genes that are important for bacterial defense. Therefore, the presence of high number of genomic islands in our strains may be partially explained by the fact that they might be more prone to invasion by foreign DNA compared to the reference strain.

In summary, the phylogenetic and ANI analysis of these two investigated strains UM_RHS and UM_NYF showed that they most likely belong to the species *M. cosmeticum*. The addition of these genome sequences may be an important avenue for comparative analyses and functional studies of *M. cosmeticum* in future.

Methods

Library construction and next-generation sequencing

The DNA of UM_RHS and UM_NYF were sequenced using Illumina HiSeq 2000 PE technology at about 1,000X coverage. Covaris S2 was used to fragment the DNA samples for 120

seconds at a temperature of 5.5–6.0 degree celcius. The quantity and quality of fragmented materials were examined using 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 purpose. Agilent BioAnalyzer 2100 was used to validate the constructed genomic library and a pool of 8pM was loaded onto 1 lane of Illumina HiSeq2000 flow cell v3 for sequencing using a 100bp PE sequencing strategy.

Read preprocessing and genome assembly

PRINSEQ lite version 0.20 [16] was used to filter exact duplicates and reverse complement duplicate reads. Reads were trimmed at Phred quality score < 20. The final reads were *de novo* assembled using CLC Genomic Workbench version 5.1 (CLC bio, Aarhus, Denmark). The assembly of the preprocessed reads were performed using the following criteria: length fraction of 0.7, similarity fraction of 0.9, and any contigs with size lesser than 500bp were discarded.

Phylogenetic inferences

The *16S rRNA*-based phylogenetic tree was constructed using Hasegawa-Kishino-Yano DNA substitution model with a bootstrap value of 500. Five selected bacterial classification marker genes, *hsp65*, *rpoB*, *tuf*, *sodA* and *16S rRNA* from the closest species were extracted and concatenated for construction of supermatrix-based tree using the same approach as the 16S rRNA-based tree.

Genome annotation

The genome of UM_RHS and UM_NYF were annotated using the RAST annotation pipeline [20]. To predict the genomic islands, the assembled sequences of UM_RHS and UM_NYF were submitted to IslandViewer [27, 28]. The generated output results were further filtered by eliminating the genomic islands situated within two different contigs [43]. The RAST-predicted protein coding genes found in the genomes of UM_RHS and UM_NYF were used for virulence genes prediction. BLAST search was performed with the RAST-predicted protein sequences against the Virulence Factors Database (VFDB) [44–46] with e-value of 10 and orthologous genes that have at least 50% sequence identity and 50% sequence completeness and with known virulence genes were considered as putative virulence genes.

Gene family clustering

The *M. cosmeticum* strain DSM44829 (accession no. GCA_000613185.1) isolated from a granulomatous lesion of a female patient in Venezuela [12] [https://www.dsmz.de/catalogues/details/culture/DSM-44829.html], was used as the reference genome sequence for the gene family clustering study. The RAST-predicted protein sequences of UM_RHS, UM_NYF and *M. cosmeticum* DSM44829 were clustered into orthologous gene families using BLASTClust with maximal e-value of $1e^{-10}$ and minimum score of 40 (http://ftp.ncbi.nih.gov/blast/documents/blastclust.html). Protein sequences with at least 50% sequence identity and 50% sequence coverage between each other were clustered into the same orthologous gene family.

Genomic islands and virulence gene prediction

Genomic islands in UM_RHS and UM_NYF were predicted using IslandViewer [27, 28] with the integration of few approaches such as the sequence composition based SIGI-HMM [47] and IslandPath-DIMOB [48] and the comparative genomics approach IslandPick [28]. The generated results from IslandViewer were further filtered by eliminating the islands situated

within 2 contigs. The RAST-predicted protein sequences in UM_RHS and UM_NYF were further BLAST searched against the Virulence Factors Database (VFDB). The BLAST results were filtered using in-house Perl scripts to select orthologous genes that are at least 50% sequence identity and 50% sequence completeness.

Supporting information

S1 Table. RNAs identified by RAST in the genomes of UM_NYF and UM_RHS.
(DOCX)

S2 Table. RAST predicted genes related to gene transfer in the genomes of UM_NYF and UM_RHS.
(DOCX)

S3 Table. Genomic islands present in UM_NYF and UM_RHS.
(DOCX)

S1 Fig. The graphical representation of GI35.
(PDF)

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 materials and inputs in this study.

Author Contributions

Conceptualization: Wei Yee Wee, Siew Woh Choo.

Data curation: Wei Yee Wee, Siew Woh Choo.

Formal analysis: Wei Yee Wee.

Funding acquisition: Siew Woh Choo.

Investigation: Wei Yee Wee.

Methodology: Wei Yee Wee.

Resources: Wei Yee Wee, Siew Woh Choo.

Software: Wei Yee Wee.

Supervision: Wei Yee Wee, Siew Woh Choo.

Validation: Wei Yee Wee.

Writing – original draft: Wei Yee Wee, Avirup Dutta, Jayasyaliny Jayaraj, Siew Woh Choo.

Writing – review & editing: Wei Yee Wee, Avirup Dutta, Jayasyaliny Jayaraj, Siew Woh Choo.

References

1. Smith I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. Clinical microbiology reviews. 2003 Jul 1; 16(3):463–96. https://doi.org/10.1128/CMR.16.3.463-496.2003 PMID: 12857778

2. Pinheiro RO, de Souza Salles J, Sarno EN, Sampaio EP. Mycobacterium leprae–host-cell interactions and genetic determinants in leprosy: an overview. Future microbiology. 2011 Feb; 6(2):217–30. https://doi.org/10.2217/fmb.10.173 PMID: 21366421
3. KJRaCG R. Sherris medical microbiology: an introduction to infectious diseases.
4. Lustgarten S. The Bacillus of Syphilis. The Lancet. 1885 Apr 4; 125(3214):609–10.
5. Khan K., Wang J. and Marras T.K., 2007. Nontuberculous mycobacterial sensitization in the United States: national trends over three decades. *American journal of respiratory and critical care medicine*, 176(3), pp.306–313. https://doi.org/10.1164/rccm.200702-201OC PMID: 17507546
6. Lai CC, Tan CK, Lin SH, Liu WL, Liao CH, Huang YT, Hsueh PR. Clinical significance of nontuberculous mycobacteria isolates in elderly Taiwanese patients. European Journal of Clinical Microbiology & Infectious Diseases. 2011 Jun 1; 30(6):779–83.
7. Falkinham J. O. III in *Seminars in respiratory and critical care medicine*. 095–102 (Thieme Medical Publishers).
8. Wallace Jr RJ, Brown BA, Griffith DE. Nosocomial outbreaks/pseudo outbreaks caused by nontuberculous mycobacteria. Annual Reviews in Microbiology. 1998 Oct; 52(1):453–90.
9. Tanaka E, Kimoto T, Matsumoto H, Tsuyuguchi K, Suzuki K, Nagai S, Shimadzu M, Ishibatake H, Murayama T, Amitani R. Familial pulmonary Mycobacterium avium complex disease. American journal of respiratory and critical care medicine. 2000 May 1; 161(5):1643–7. https://doi.org/10.1164/ajrccm.161.5.9907144 PMID: 10806169
10. Von Reyn CF, Arbeit RD, Horsburgh CR, Ristola MA, Waddell RD, Tvaroha SM, Samore M, Hirschhorn LR, Lumio J, Lein AD, Grove MR. Sources of disseminated Mycobacterium avium infection in AIDS. Journal of Infection. 2002 Apr 1; 44(3):166–70. https://doi.org/10.1053/jinf.2001.0950 PMID: 12099743
11. Perez-Martinez I, Aguilar-Ayala DA, Fernandez-Rendon E, Carrillo-Sanchez AK, Helguera-Repetto AC, Rivera-Gutierrez S, Estrada-Garcia T, Cerna-Cortes JF, Gonzalez-y-Merchand JA. Occurrence of potentially pathogenic nontuberculous mycobacteria in Mexican household potable water: a pilot study. BMC research notes. 2013 Dec; 6(1):531.
12. Cooksey RC, de Waard JH, Yakrus MA, Rivera I, Chopite M, Toney SR, Morlock GP, Butler WR. Mycobacterium cosmeticum sp. nov., a novel rapidly growing species isolated from a cosmetic infection and from a nail salon. International journal of systematic and evolutionary microbiology. 2004 Nov 1; 54 (6):2385–91.
13. Zhang L, Zhang C, Cheng Z, Yao Y, Chen J. Biodegradation of benzene, toluene, ethylbenzene, and o-xylene by the bacterium Mycobacterium cosmeticum byf-4. Chemosphere. 2013 Jan 1; 90(4):1340–7. https://doi.org/10.1016/j.chemosphere.2012.06.043 PMID: 22960059
14. Addley J, McKeagney P, Turner G, Kelly M. Other full case: Mycobacterium cosmeticum as an unusual cause of ascites. BMJ case reports. 2010;2010.
15. Boschetti G, Cotte E, Moussata D, Chauvenet M, Breysse F, Chomarat M, Isaac S, Berger F, Kaiserlian D, Nancey S, Flourie B. Identification of Mycobacterium cosmeticum sp. as a novel colitogenic infectious agent in a nonimmunocompromised patient. Inflammatory bowel diseases. 2011 Oct; 17(10): E128–30. https://doi.org/10.1002/ibd.21804 PMID: 21739530
16. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011 Jan 28; 27(6):863–4. https://doi.org/10.1093/bioinformatics/btr026 PMID: 21278185
17. Stahl DA, Urbance JW. The division between fast-and slow-growing species corresponds to natural relationships among the mycobacteria. Journal of Bacteriology. 1990 Jan 1; 172(1):116–24. PMID: 1688423
18. Cunningham CW. Can three incongruence tests predict when data should be combined?. Molecular Biology and Evolution. 1997 Jul 1; 14(7):733–40. https://doi.org/10.1093/oxfordjournals.molbev.a025813 PMID: 9214746
19. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences. 2005 Feb 15; 102(7):2567–72.
20. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F. The RAST Server: rapid annotations using subsystems technology. BMC genomics. 2008 Dec; 9(1):75.
21. Kimura S, Suzuki T. Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA. Nucleic acids research. 2009 Dec 3; 38(4):1341–52. https://doi.org/10.1093/nar/gkp1073 PMID: 19965768
22. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PloS one. 2010 Jun 25; 5(6):e11147. https://doi.org/10.1371/journal.pone.0011147 PMID: 20593022

23. Murray NE. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiology and molecular biology reviews. 2000 Jun 1; 64(2):412–34. PMID: 10839821

24. Brahmachary P, Wang G, Benoit SL, Weinberg MV, Maier RJ, Hoover TR. The human gastric pathogen Helicobacter pylori has a potential acetone carboxylase that enhances its ability to colonize mice. BMC microbiology. 2008 Dec; 8(1):14.

25. Saler Jr MH, Reizer J. The bacterial phosphotransferase system: new frontiers 30 years later. Molecular microbiology. 1994 Sep; 13(5):755–64. PMID: 7815935

26. Gorke B, Stulke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nature Reviews Microbiology. 2008 Aug; 6(8):613. https://doi.org/10.1038/nrmicro1932 PMID: 18628769

27. Langille MG, Brinkman FS. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics. 2009 Jan 16; 25(5):664–5. https://doi.org/10.1093/bioinformatics/btp030 PMID: 19151094

28. Langille MG, Hsiao WW, Brinkman FS. Evaluation of genomic island predictors using a comparative genomics approach. BMC bioinformatics. 2008 Dec; 9(1):329.

29. Zhong J. J., Bai F. W. & Zhang W. in *Advances in Biochemical Engineering/Biotechnology* Vol. 113 (ed Scheper T.) (Springer, 2009).

30. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic acids research. 2011 Jun 14; 39(suppl_2):W347–52.

31. Van Pittius NC, Sampson SL, Lee H, Kim Y, Van Helden PD, Warren RM. Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC evolutionary biology. 2006 Dec; 6(1):95.

32. Li W, Zhao Q, Deng W, Chen T, Liu M, Xie J. Mycobacterium tuberculosis Rv3402c enhances mycobacterial survival within macrophages and modulates the host pro-inflammatory cytokines production via NF-kappa B/ERK/p38 signaling. PloS one. 2014 Apr 10; 9(4):e94418. https://doi.org/10.1371/journal.pone.0094418 PMID: 24722253

33. Miller JL, Velmurugan K, Cowan MJ, Briken V. The type I NADH dehydrogenase of Mycobacterium tuberculosis counters phagosomal NOX2 activity to inhibit TNF-α-mediated host cell apoptosis. PLoS pathogens. 2010 Apr 22; 6(4):e1000864. https://doi.org/10.1371/journal.ppat.1000864 PMID: 20421951

34. Forrellad MA, Klepp LI, Gioffré A, Sabio y Garcia J, Morbidoni HR, Santangelo MD, Cataldi AA, Bigi F. Virulence factors of the Mycobacterium tuberculosis complex. Virulence. 2013 Jan 1; 4(1):3–66. https://doi.org/10.4161/viru.22329 PMID: 23076359

35. Wiker HG, Harboe M. The antigen 85 complex: a major secretion product of Mycobacterium tuberculosis. Microbiological reviews. 1992 Dec 1; 56(4):648–61. PMID: 1480113

36. Cimino M, Thomas C, Namouchi A, Dubrac S, Gicquel B, Gopaul DN. Identification of DNA binding motifs of the Mycobacterium tuberculosis PhoP/PhoR two-component signal transduction system. PloS one. 2012 Aug 7; 7(8):e42876. https://doi.org/10.1371/journal.pone.0042876 PMID: 22880126

37. Pérez E, Samper S, Bordas Y, Guilhot C, Gicquel B, Martín C. An essential role for phoP in Mycobacterium tuberculosis virulence. Molecular microbiology. 2001 Jul 1; 41(1):179–87. PMID: 11454210

38. Li AH, Waddell SJ, Hinds J, Malloff CA, Bains M, Hancock RE, Lam WL, Butcher PD, Stokes RW. Contrasting transcriptional responses of a virulent and an attenuated strain of Mycobacterium tuberculosis infecting macrophages. PloS one. 2010 Jun 10; 5(6):e11066. https://doi.org/10.1371/journal.pone.0011066 PMID: 20548782

39. Li F, Xu P, Feng J, Meng L, Zheng Y, Luo L, Ma C. Microbial desulfurization of gasoline in a Mycobacterium goodii X7B immobilized-cell system. Applied and environmental microbiology. 2005 Jan 1; 71(1):276–81. https://doi.org/10.1128/AEM.71.1.276-281.2005 PMID: 15640198

40. Kilbane JJ. Desulfurization of coal: the microbial solution. Trends in Biotechnology. 1989 Apr 1; 7(4):97–101.

41. Monticello DJ. Biodesulfurization and the upgrading of petroleum distillates. Current Opinion in Biotechnology. 2000 Dec 1; 11(6):540–6. PMID: 11102787

42. Ma X, Sun L, Song C. A new approach to deep desulfurization of gasoline, diesel fuel and jet fuel by selective adsorption for ultra-clean fuels and for fuel cell applications. Catalysis today. 2002 Dec 1; 77(1–2):107–16.

43. Juhas M, Van Der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS microbiology reviews. 2009 Feb 13; 33(2):376–93. https://doi.org/10.1111/j.1574-6976.2008.00136.x PMID: 19178566

44. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic acids research. 2011 Nov 8; 40(D1):D641–5.

45. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. Nucleic acids research. 2005 Jan 1; 33(suppl_1):D325–8.

46. Yang J, Chen L, Sun L, Yu J, Jin Q. VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. Nucleic acids research. 2007 Nov 4; 36(suppl_1):D539–42.

47. Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC bioinformatics. 2006 Dec; 7(1):142.

48. Hsiao W, Wan I, Jones SJ, Brinkman FS. IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics. 2003 Feb 12; 19(3):418–20. PMID: 12584130