

EXTENDED GENOME REPORT

Open Access



# Genomic insights into *Mycobacterium simiae* human colonization

José L. Steffani-Vallejo<sup>1</sup>, Marion E. Brunck<sup>3</sup>, Erika Y. Acosta-Cruz<sup>1,2,4</sup>, Rafael Montiel<sup>2</sup> and Francisco Barona-Gómez<sup>1\*</sup>

## Abstract

*Mycobacterium simiae* (Karassova V, Weissfeiler J, Kraszanay E, Acta Microbiol Acad Sci Hung 12:275-82, 1965) is a slow-growing nontuberculous *Mycobacterium* species found in environmental niches, and recently evidenced as an opportunistic Human pathogen. We report here the genome of a clinical isolate of *M. simiae* (MsiGto) obtained from a patient in Guanajuato, Mexico. With a size of 6,684,413 bp, the genomic sequence of strain MsiGto is the largest of the three *M. simiae* genomes reported to date. Gene prediction revealed 6409 CDSs in total, including 6354 protein-coding genes and 52 RNA genes. Comparative genomic analysis identified shared features between strain MsiGto and the other two reported *M. simiae* genomes, as well as unique genes. Our data reveals that *M. simiae* MsiGto harbors virulence-related genes, such as *arcD*, ESAT-6, and those belonging to the antigen 85 complex and *mce* clusters, which may explain its successful transition to the human host. We expect the genome information of strain MsiGto will provide a better understanding of infective mechanisms and virulence of this emergent pathogen.

**Keywords:** *Mycobacterium simiae*, Nontuberculous mycobacteria, Opportunistic pathogen

## Introduction

Contrasting with the declining incidence of *Mycobacterium tuberculosis*-caused tuberculosis, the increasing number of nontuberculous mycobacteria infections is concerning. Amongst NTM, the *Mycobacterium simiae* complex contains 19 species [1], including *M. simiae* [2], which is considered the most important species in terms of its clinical relevance [3]. *M. simiae* is a slow-growing saprophyte that has been isolated from several environments including water and soil [3, 4]. In addition to thriving in environmental niches, *M. simiae* has been associated to infections in both immunocompromised [4, 5] and immunocompetent patients [6], with clinical cases reported worldwide [5, 6]. Thus, *M. simiae* is considered an emergent pathogen [7–9]. Previous sequencing efforts have provided draft genome sequences for *M. simiae*, namely, strains MO323 (accession PRJNA276839) and DSM 44165 (accession PRJEB1560) [10], which were isolated in early years from the United States of America (1989) and India (1965), respectively.

We expand here on this genomic data by reporting the genome sequence of a clinical isolate from a Mexican patient (bronchial lavage) obtained in 2011, designated as MsiGto, which is approximately 745 Kbp and 901 Kbp larger than the previously published *M. simiae* genomes, respectively. Additionally, we perform a comparative analysis between the genomes of strain MsiGto and the previously reported DSM 44165 and MO323 strains to provide insights that could unearth the transition from environmental bacteria to pathogenic NTM organisms.

## Organism information

### Classification and features

To identify strain MsiGto as an *M. simiae* isolate, typical DNA markers for mycobacterial identification, such as the 16S rRNA, *rpoB* and *hsp65* genes [11–13] were localized on the sequenced MsiGto genome, and compared to other publicly available sequences using BLASTN on its default settings. The MsiGto 16S rRNA gene exhibited 100% identity with previously deposited *M. simiae* sequences (strains DS39, DS34, DS31, DS24, DS19, DS3, DS2, MO323 and ATCC 25275). The MsiGto *rpoB* gene sequence showed 99% identity with *M. simiae* MO323. Similarly, there was 100% identity

\* Correspondence: francisco.barona@cinvestav.mx

<sup>1</sup>Evolution of Metabolic Diversity Laboratory, Unidad de Genómica Avanzada (Langebio), Cinvestav-IPN, Irapuato, Mexico

Full list of author information is available at the end of the article

between the MsiGto and MO323 *hsp65* genes. For comparison, a BLAST of these markers from DSM 44165 and MO323 evidenced a similar trend with 99% identity for the 16S rRNA gene and 100% identity for *rpoB* and *hsp65*. Figure 1 shows the phylogenetic position of *M. simiae* MsiGto within the *M. simiae* complex, based on a concatenated gene tree including the sequences of the 16S rRNA, *rpoB* and *hsp65* genes (1507 bp in total).

Further analysis such as the Average Nucleotide Identity and Average Amino Acid Identity, performed as previously reported [14], confirmed the high resemblance between *M. simiae* MsiGto and *M. simiae* strains MO323 and DSM 44165 (Table 1). The ANI values were above 95% in all cases, indicating that the compared organisms belong to the same species. Similar results were found using Genome-to-Genome Distance Calculator [15, 16] with a calculated distance of 0.0292 and a DNA-DNA hybridization estimate of 77.40% using the formula for draft genomes. These results provide a taxonomic background in which the genome insights presented in the following sections can better accounted for, namely, large genotype differences of MsiGto, despite close taxonomic relationships with DSM 44165 and MO323 strains.

## Genome sequencing information

### Genome project history

Despite *M. simiae* being one of the most relevant NTM due to its emergence as a human pathogen, the genomic features and genetic potential of this species remain poorly described. In collaboration with the State Laboratory of Public Health of Guanajuat, located in central Mexico, we had access to a clinical isolate of *M. simiae*, termed MsiGto. The sample was isolated from a 90 years old female patient from León, the largest and most industrialized city of the State of Guanajuato, México, in 2011 (Table 2). This sample was selected for genomic

**Table 1** Average Nucleotide Identity (A) and Average Amino acid Identity (B) between *M. simiae* MsiGto and other *M. simiae* strains sequenced to date

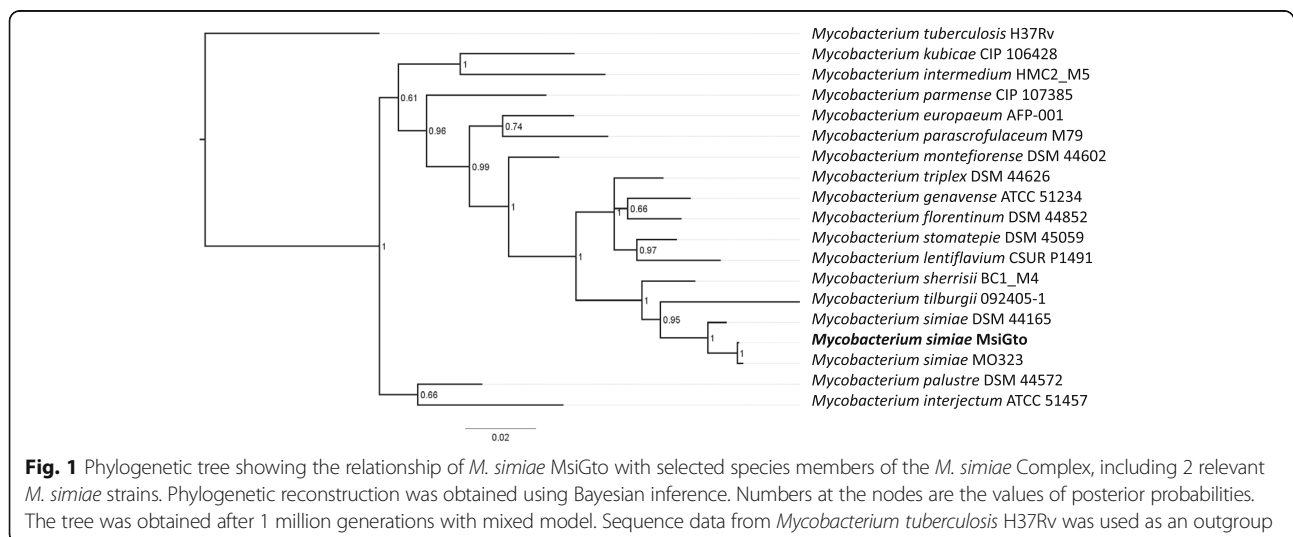
Organism	M. simiae MsiGto A/B	M. simiae DSM 44165	M. simiae MO323
M. simiae MsiGto	–	97.25/ 97.51	98.99/98.95
M. simiae DSM 44165	97.25/ 97.51	–	97.26/97.59
M. simiae MO323	98.99/98.95	97.26/97.59	–

sequencing due to the emerging clinical importance of strains from the *M. simiae* complex worldwide [8, 9, 17–19], combined with a lack of representative genomic sequences from Mexico to this date. The available genome sequences could allow comparative analyses in order to increase our understanding of this opportunistic bacterium.

### Growth conditions and genomic DNA preparation

MsiGto was isolated from a sputum specimen. Briefly, 2 mL sample were transferred to a sterile tube and decontaminated by adding an equal volume of 4% NaOH with phenol red. The mix was immediately vortexed and incubated at 37 °C for 15 min. The liquefied sample was centrifuged at 3000 g at 4 °C for 15 min, the supernatant was discarded and pH was adjusted between the 6.5 and 7.2 range using 1 N HCL. Finally, 0.2 mL of sample were inoculated into Lowenstein-Jensen slants (Difco) and incubated at 37 °C for three weeks.

Biomass was collected and suspended in phosphate buffered saline solution, pH 7.4, and inactivated by heating at 80 °C for 45 min. After centrifugation, genomic DNA was extracted from the pellet using the FastDNA SPIN Kit for Soil (MP Biomedicals) according to manufacturer's instructions. Extracted gDNA was assessed for quantity and quality using a NanoDrop Spectrophotometer (Thermo Fisher Scientific).



**Table 2** Classification and general features of *Mycobacterium simiae* MsiGto [51]

MIGS ID	Property	Term	Evidence code <sup>a</sup>	
	Classification	Domain <i>Bacteria</i>	TAS [52]	
		Phylum <i>Actinobacteria</i>	TAS [53]	
		Class <i>Actinobacteria</i>	TAS [54]	
		Order <i>Actinomycetales</i>	TAS [55, 56]	
		Family <i>Mycobacteriaceae</i>	[54, 56, 57]	
		Genus <i>Mycobacterium</i>	TAS [58, 59]	
		Species <i>Mycobacterium simiae</i>	IDA	
		Gram stain	Weakly Postive	IDA
		Cell shape	Irregular rods	IDA
	Motility	Non Motile	IDA	
	Sporulation	Nonsporulating	NAS	
	Temperature range	Mesophile	NAS	
	Optimum temperature	37 °C	NAS	
	pH range; Optimum	5.5–8; 7	IDA	
	Carbon source	Starch	IDA	
MIGS-6	Habitat	Human Associated	NAS	
MIGS-6.3	Salinity	Normal	NAS	
MIGS-22	Oxygen requirement	Aerobic	NAS	
MIGS-15	Biotic relationship	Parasitic	IDA	
MIGS-14	Pathogenicity	Pathogenic	NAS	
MIGS-4	Geographic location	Mexico/Guanajuato	NAS	
MIGS-5	Sample collection	2014	NAS	
MIGS-4.1	Latitude	Not Reported	NAS	
MIGS-4.2	Longitude	Not Reported	NAS	
MIGS-4.4	Altitude	Not Reported	NAS	

<sup>a</sup>Evidence codes - *IDA* Inferred from Direct Assay, *TAS* Traceable Author Statement (i.e., a direct report exists in the literature), *NAS* Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [60]

### Genome sequencing and assembly

The MsiGto genome was sequenced on a HiSeq 2000 platform with a 100 bp paired-end cycle according to standard Illumina protocols, at MacroGen facilities. Quality of the generated sequencing reads (13,676,836 reads, total read length: 1,381,360,436 bp) was checked with FastQC. Reads were filtered before assembly, such that both sequences in paired-end reads exhibited more than 90% bases of quality greater than or equal to Q20. Post-filtering Q20% was 97.69 and Q30% was 88.6 at the base level. The filtered reads were assembled using SOAP de novo aligner [20], yielding 50 contigs. Scaffolding was performed with SSPACE Standard [21] and resulted in 12 scaffolds, generating a genome size of 6.7 Mb with a coverage of 216X (Table 3).

### Genome annotation

Gene prediction and functional annotation were performed using the Rapid Annotation using Subsystem

**Table 3** Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Draft
MIGS-28	Libraries used	Paired End Illumina
MIGS 29	Sequencing platforms	Illumina HiSeq 2000
MIGS 31.2	Fold coverage	216
MIGS 30	Assemblers	SOAPdenovo
MIGS 32	Gene calling method	RAST
	Locus Tag	B5M45
	Genbank ID	MZZM000000000
	GenBank Date of Release	April 17, 2017
	GOLD ID	Ga0183212
	BIOPROJECT	PRJNA378996
MIGS 13	Source Material Identifier	<i>Mycobacterium simiae</i> MsiGto
	Project relevance	Medical, Evolutionary

Technology platform [22]. A prophage region prediction was completed using PHAST [23]. CRISPRs were searched using the CRISPR finder, and antibiotic resistance genes were investigated using Resistance Gene Identifier from the Comprehensive Antibiotic Resistance Database [24]. Genome-to-genome comparisons were performed using multiple approaches, including conservation analysis of protein families across genomes with the Protein Family Sorter from the PATRIC Platform [25], and comparing functionally related clusters using the function-based comparison of the RAST server. MsiGto and related genomes were also aligned using a genome-wide BLAST comparison, and visualized through the Artemis Comparative Tool [26] for manual inspection.

### Genome properties

Genomic assembly yielded a total length of 6,684,413 bp fragmented in 50 contigs, the largest *M. simiae* genome reported to date. The MsiGto genome consists of a unique chromosome, as no plasmid DNA was found. The GC content of the genome was 66.08%, consistent with other reported *M. simiae* strains. Gene prediction analysis documented 3 rRNAs, 49 tRNAs and 6409 coding sequences. Out of these, 4462 genes (69.62%) were assigned to putative functions, and 3669 genes (approximately 62.57%) were assigned to clusters of orthologous groups functional categories. Sequence searches using RGI evidenced resistance genes to aminoglycosides, ethambutol and beta lactams, which is a concern as to date standard antibiotic regimes for this species include the use of ethambutol, the aminoglycoside amikacin, as well as the macrolides azithromycin and clarithromycin [27]. The genome properties and statistics are summarized in Tables 4 and 5. Gene distribution among the COG functional categories is shown in Table 6. A circular map of MsiGto chromosome is provided in Fig. 2.

### Insights from the genome sequence

#### Extended insights

We report here the third genome sequence of *M. simiae*. MsiGto was isolated from a clinical sample obtained from an elderly woman, and is the largest *M. simiae* genome reported to date with an approximate 745 Kbp additional to the next largest genome (strain MO323). As pathogenic mycobacteria tend to undergo genome decay [28, 29], the large genomic size of MsiGto speaks out of an organism capable of thriving in both environmental conditions and in the human host. Indeed, by

**Table 4** Summary of genome: one chromosome, no plasmids

Label	Size (Mb)	Topology	INSDC identifier	RefSeq ID
Chromosome	6,684,413	Circular	GenBank	NZ_MZZM00000000.1

**Table 5** Genome statistics

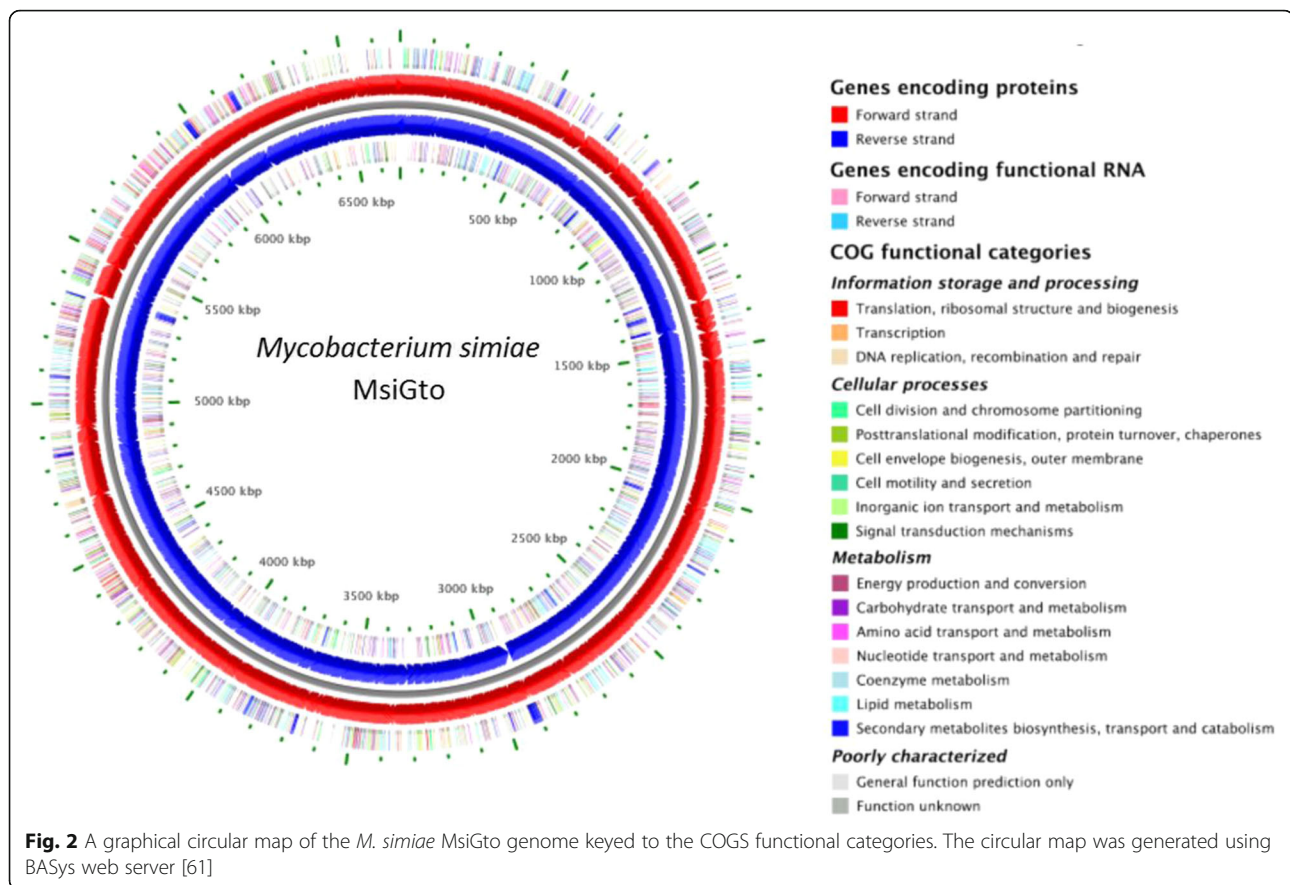
Attribute	Value	% of Total
Genome size (bp)	6,684,413	100
DNA coding (bp)	5,978,008	89.43
DNA G + C (bp)	4,416,391	66.07
DNA scaffolds	15	100
Total genes	6369	100
Protein coding genes	6299	99.90
RNA genes	70	1.10
Pseudo genes	160	2.51
Genes in internal clusters	579	9.09
Genes with function prediction	4713	74.00
Genes assigned to COGs	5272	82.29
Genes with Pfams domains	5009	78.19
Genes with signal peptides	260	4.08
Genes with transmembrane helices	1292	20.29
CRISPR repeats	15	

**Table 6** Number of genes associated with general COG functional categories

Code	Value	%age <sup>a</sup>	Description
J	161	2.53	Translation, ribosomal structure and biogenesis
A	21	0.33	RNA processing and modification
K	448	7.05	Transcription
L	192	3.02	Replication, recombination and repair
B	1	0.01	Chromatin structure and dynamics
D	55	0.86	Cell cycle control, Cell division, chromosome partitioning
V	47	0.73	Defense mechanisms
T	218	3.43	Signal transduction mechanisms
M	159	2.50	Cell wall/membrane biogenesis
N	57	0.89	Cell motility
U	36	0.56	Intracellular trafficking and secretion
O	153	2.41	Posttranslational modification, protein turnover, chaperones
C	447	7.03	Energy production and conversion
G	243	3.82	Carbohydrate transport and metabolism
E	298	4.69	Amino acid transport and metabolism
F	82	1.29	Nucleotide transport and metabolism
H	210	3.30	Coenzyme transport and metabolism
I	558	8.78	Lipid transport and metabolism
P	239	3.76	Inorganic ion transport and metabolism
Q	494	7.77	Secondary metabolites biosynthesis, transport and catabolism
R	795	12.51	General function prediction only
S	358	5.63	Function unknown
-	1082	17.03	Not in COGs

<sup>a</sup>The total is based on the total number of protein coding genes in the genome



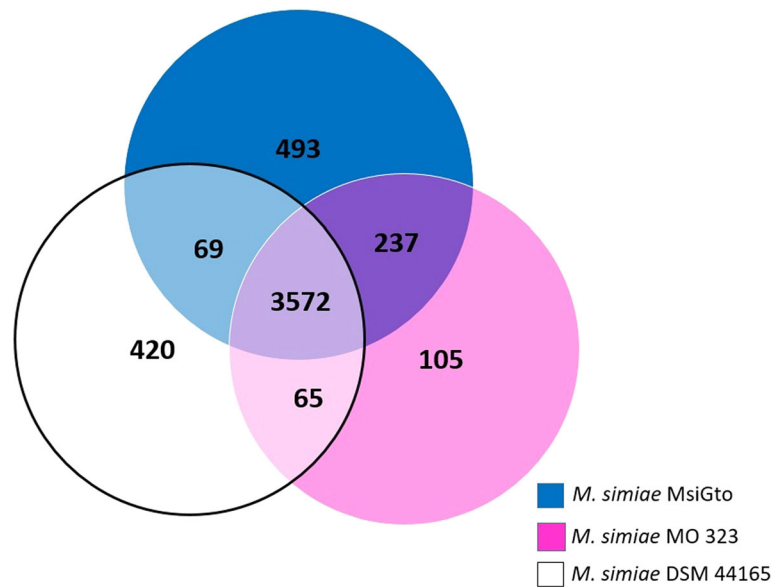


comparison, the genome of *M. tuberculosis*, an obligate pathogen, is significantly smaller with 4.41 Mb [30]. Meanwhile, with 6406 genes, the genome size of MsiGto is similar to that of the opportunistic pathogen *Pseudomonas aeruginosa* (5570 predicted ORFs), which thrives in a variety of environments including soil, water, as well as the multiple human tissues it infects [31]. Analysis of the *P. aeruginosa* genome evidenced its large size arose from genetic expansion to enhance functional diversity rather than from gene duplication [32]. Interestingly, investigating protein families in MsiGto, in comparison with MO323 and DSM 44165, evidenced a larger proportion of proteins uniquely found in MsiGto (1.17 times more than DSM 44165 and 4.7 times more than MO323), suggestive of a relatively increased versatility in accordance with the multiple niches it can thrive in (Fig. 3).

The genome of MsiGto is rich in virulence factors, genes conferring infective mechanisms, and host immune response evasion systems. Some of these proteins are common to other pathogenic mycobacteria, and shared by the two previously sequenced *M. simiae* genomes, such as ESAT-6, known to modulate host immune responses by affecting human T-cell responses [33, 34]. Interestingly, all *M. simiae* strains including MsiGto have three of the four antigen 85 complex genes

(*fbpA*, *fbpC*, and *fbpD*) responsible for cell wall synthesis. The enzyme products of these genes are responsible for the conversion of trehalose monomycolate (TMM) into the Cord Factor trehalose dimycolate (TDM) [35], which is considered one of the most important virulence factors of mycobacteria [36]. Other genera belonging to the *Mycobacteriaceae* family that include both pathogenic and environmental species, such as *Rhodococcus* [37], *Corynebacterium* and *Nocardia* [38], also produce TDM [39].

Mammalian Cell Entry proteins are cell surface exposed proteins that play a crucial role in *M. tuberculosis* virulence by permitting the bacteria to enter mammalian cells and survive inside the macrophage, modulating the immune response [40, 41]. Mce clusters consist of 4 homologous operons in *M. tuberculosis* (*mce1*, *mce2*, *mce3*, *mce4*) with a similar arrangement: two genes encoding integral membrane proteins followed by six *mce* genes (A, B, C, D, E and F) [40]. Mce proteins are also involved in lipid metabolism, acting as transporters and allowing cholesterol degradation to free carbon and energy for use by *M. tuberculosis* [41]. In the genome of MsiGto we found that the cluster *mce3* is overrepresented. While the DSM 44165 and MO323 genomes present none and a single copy of *mce3*, respectively,



**Fig. 3** Venn diagram analysis showing the number of unique and shared family proteins as evidenced using PATRIC, between the *M. simiae* strains MO323, DSM 44165 and MsiGto

two complete copies of the *mce3* cluster were found in MsiGto. As the transition from an environmental organism to a pathogen has been associated with the acquisition of Mce genes in actinobacteria [42], and Mce3 proteins are expressed by *M. tuberculosis* during the infection phase [43], it is tempting to speculate the increased number of Mce3 copies found in MsiGto provided the strain with human infection potential.

The presence of multiple copies of these potential lipid transporters in mycobacterial genomes is consistent with the finding that pathogenic mycobacteria switch from carbohydrates to lipids as their main carbon and energy source inside cells [44]. The evolution of this locus, through duplication and divergence, has almost certainly contributed to virulence in mycobacteria, and even in other distantly related actinobacteria, such as *Streptomyces* [42, 45]. Given that the ability to acquire cholesterol from the host is crucial to maintain a chronic infection, we postulate that cells having a large *mce* copy number, such as that found in MsiGto, may potentially evolve pathogenicity relatively faster when compared with other environmental mycobacteria.

A total of 493 protein families were found exclusively present in MsiGto when compared to the other two available *M. simiae* genomes (MO323 and DSM 44165). Consistent with findings in the *Pseudomonas aeruginosa* [46] genome [32], it seems that MsiGto has undergone genome expansion. Within the gene pool unique to MsiGto we found the arginine/ornithine antiporter gene *arcD*, which is involved in the persistence of the zoonotic pathogen *Streptococcus suis* [47] in host cells [48].

Additional unique genes found in the MsiGto genome participate in aromatic amino acid metabolic pathways. For instance, 2-oxo-hepta-3-ene-1,7-dioic acid hydratase (*hpcG* gene, EC 4.2.1.80) participates in the degradation of tyrosine and the 2-keto-3-deoxy-D-arabino-heptulosonate-7-phosphate synthase I alpha participates in the synthesis of chorismate (*aroG* gene, EC 2.5.1.54). The latter is an interesting observation as tyrosine is a key nutrient source during infectious growth within macrophages of some pathogenic fungus [49, 50]. At this stage, however, it would be risky to rule out the involvement of these specific genes in important environmental functions that allow MsiGto to survive outside the host, or during both lifestyles.

## Conclusions

*Mycobacterium simiae* is an organism of interest for genomic studies due to the scarce genomic data available, and its recent emergence as a human pathogen. Here we present the largest genome sequence of this species to date. The genome of *M. simiae* MsiGto presents characteristics in accordance with its adaptation to infect the Human host, with the presence of numerous virulence genes, plus some specific features that deserve further investigation. Additional *M. simiae* genomes, from both environmental and clinical isolates, should be sequenced to provide a wider evolutionary picture with functional implications. Indeed, our comparative analysis helps to better understand the evolution of host-pathogen interactions, and the molecular mechanisms of virulence, of this emergent human pathogen.

**Abbreviations**

ACT: Artemis Comparative Tool; CARD: Comprehensive Antibiotic Resistance Database; LaESaP: State Laboratory of Public Health of Guanajuato; MCE: Mammalian Cell Entry; NTM: Nontuberculous mycobacteria; RAST: Rapid Annotation using Subsystem Technology; RGI: Resistance Gene Identifier; TDM: Trehalose 6-dimicolate; TMM: Trehalose monomycolate

**Acknowledgements**

We thank the Secretaría de Salud del Estado de Guanajuato and María Guadalupe Hurtado Torres from the LaESaP for their invaluable contribution to obtain the biological material, and Karina Gutiérrez-García for help with bioinformatic analysis.

**Funding**

The authors thank Fomix-Gto (grant number GTO-2011-C04-165962) for funding of this work, and Conacyt for a postdoctoral scholarship to EA (CVU No. 248685).

**Authors' contributions**

JLS performed all bioinformatics analyses and contributed to the writing of the manuscript. MB participated in data analysis and writing of the manuscript. EA participated in sample processing for genomic sequencing and collaborate writing the first draft of the manuscript. RM and FBG conceived the study, supervised this project, and were responsible for completing the manuscript. All authors read and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Evolution of Metabolic Diversity Laboratory, Unidad de Genómica Avanzada (Langebio), Cinvestav-IPN, Irapuato, Mexico. <sup>2</sup>Paleogenomics Laboratory, Unidad de Genómica Avanzada (Langebio), Cinvestav-IPN, Irapuato, Mexico. <sup>3</sup>Centro de Biotecnología FEMSA, Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Monterrey, Mexico. <sup>4</sup>Present address: Laboratorio de Biología Molecular, Facultad de Ciencias Químicas, Universidad Autónoma de Coahuila, Saltillo, Mexico.

Received: 30 June 2017 Accepted: 24 November 2017

Published online: 08 January 2018

**References**

- Tortoli E. Microbiological features and clinical relevance of new species of the genus mycobacterium. *Clin Microbiol Rev.* 2014;27:727–52.
- Karassova V, Weissfeiler J, Kraszanay E. Occurrence of atypical mycobacteria in *Macacus rhesus*. *Acta Microbiol Acad Sci Hung.* 1965;12(3):275–82.
- Makovcova J, Slany M, Babak V, Slana I, Kralik P. The water environment as a source of potentially pathogenic mycobacteria. *J Water Health.* 2014;12:254–63.
- García Martos P, García Agudo L, González Moya E, Galán F, Rodríguez Iglesias M. Article in press. *Enferm Infecc Microbiol Clin.* 2014:1–9.
- Van Ingen J, Boeree MJ, Dekhuijzen PNR, Van Soolingen D. Clinical relevance of mycobacterium *simiae* in pulmonary samples. *Eur Respir J.* 2008;31:106–9.
- Cortés-Torres N, González-y-Merchand JA, González-Bonilla C, García-Elorriaga G. Molecular analysis of mycobacteria isolated in Mexican patients with different immunodeficiencies in a tertiary care hospital. *Arch Med Res.* 2013;44:562–9.
- Hashemi-Shahraki A, Darban-Sarokhalil D, Heidarieh P, Mehdi Feizabadi M, Deshmir-Salameh S, Khazaei S, et al. Mycobacterium *simiae*: a possible emerging pathogen in Iran. *Jpn J Infect Dis.* 2013;66:475–9.
- Qvist T, Katzenstein TL, Lillebaek T, Iversen M, Mared L, Andersen AB. First report of lung transplantation in a patient with active pulmonary Mycobacterium *simiae* infection. *Transplant. Proc. Elsevier Inc.* 2013;45, 803–845.
- Martínez-barricarte R, Megged O, Stepensky P, Casimir P, Averbuch D, Assous MV, et al. Mycobacterium *simiae* infection in two unrelated patients with different forms of inherited IFN- $\gamma$ R2 deficiency. 2015;34:904–9.
- Sassi M, Robert C, Raoult D, Drancourt M. Non-contiguous genome sequence of mycobacterium *simiae* strain DSM 44165(T). *Stand Genomic Sci.* 2013;8:306–17.
- Kim H, Kim SH, Shim TS, Kim MN, Bai GH, Park YG, et al. Differentiation of Mycobacterium species by analysis of the heat-shock protein 65 gene (hsp65). *Int J Syst Evol Microbiol.* 2005;55:1649–56.
- Adékambi T, Colson P, Drancourt M. rpoB-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. *J Clin Microbiol.* 2003;41:5699–708.
- Pauls RJ, Turenne CY, Wolfe JN, Kabani AA. High proportion of novel mycobacteria species identified by 16S rDNA analysis among slowly growing AccuProbe-negative strains in a clinical setting. *Am J Clin Pathol.* 2003;120:560–6.
- Rodríguez RL-M, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J Prepr.* 2016; <https://doi.org/10.7287/peerj.preprints.1900v1>.
- Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci.* 2010;2:117–34.
- Auch AF, Klenk H-P, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci.* 2010;2:142–8.
- Campos CED, Dias CF, Unis G, de Souza Caldas PC, Redner P, de Carvalho LD, et al. First detection of Mycobacterium Triplex in Latin America. *Int J Mycobacteriology.* 2016;5:89–91.
- Floyd MM, Guthertz LS, Silcox VA, Duffey PS, Jang Y, Desmond EP, et al. Characterization of an SAV organism and proposal of Mycobacterium Triplex sp. nov. *J Clin Microbiol.* 1996;34:2963–7.
- List E. Validation of publication of new names and new combinations previously effectively published outside the IJSB. *Int J Syst Bacteriol.* 1997;54:601–2.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience.* 2012;1:18.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* 2011;27:578–9.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. *BMC Genomics.* 2008;9:75.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DSPHAST, Fast Phage A. Search tool. *Nucleic Acids Res.* 2011;39.
- Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 2016;gkw1004.
- Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 2014;42.
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill JACT. The Artemis comparison tool. *Bioinformatics.* 2005;21:3422–3.
- Amsterdam D. *Antibiotics in Laboratory Medicine.* 6th Editio. Philadelphia, PA; 2014.
- Moran NA. Microbial minimalism: genome reduction in bacterial pathogens. *Cell.* 2002;583–6.
- Gómez-Valero L, Rocha EPC, Latorre A, Silva FJ. Reconstructing the ancestor of mycobacterium *leprae*: the dynamics of gene loss and genome reduction. *Genome Res.* 2007;17:1178–85.
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of mycobacterium tuberculosis from the incomplete genome sequence. *Nature.* 1998;393:537–44.
- Driscoll JA, Brody SL, Kollef MH. The epidemiology, pathogenesis and treatment of *Pseudomonas Aeruginosa* infections. *Drugs.* 2007;67:351–68.
- Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas Aeruginosa* PAO1, an opportunistic pathogen. *Nature.* 2000;406:959–64.
- Sreejit G, Ahmed A, Parveen N, Jha V, Valluri VL, Ghosh S, et al. The ESAT-6 protein of mycobacterium tuberculosis interacts with Beta-2-microglobulin ( $\beta$ 2M) affecting antigen presentation function of macrophage. *PLoS Pathog.* 2014;10

34. Samten B, Wang X, Barnes PF. Mycobacterium tuberculosis ESX-1 system-secreted protein ESAT-6 but not CFP10 inhibits human T-cell immune responses. *Tuberculosis*. 2009;89 (Suppl 1):S74-S76.
35. Backus KM, Dolan MA, Barry CS, Joe M, McPhie P, Boshoff HIM, et al. The three mycobacterium tuberculosis antigen 85 isoforms have unique substrates and activities determined by non-active site regions. *J Biol Chem*. 2014;289:25041–53.
36. Hunter RL, Olsen MR, Jagannath C, Actor JK. Multiple roles of cord factor in the pathogenesis of primary, secondary, and cavitary tuberculosis, including a revised description of the pathology of secondary disease. *Ann Clin Lab Sci*. 2006;36:371–86.
37. Zopf W. Ber ausscheidung von fettfarbstoffen (lipochromen) seitens gewisser spaltpilz. *Berichte der Dtsch Bot Gesellschaf*. 1891:22–8.
38. Trevisan V. I. generi e le specie delle batteriacee. Zanaboni and Gabuzzi, Milano. 1889;
39. Spargo BJ, Crowe LM, Ioneda T, Beaman BL, Crowe JH. Cord factor (alpha,alpha-trehalose 6,6'-dimycolate) inhibits fusion between phospholipid vesicles. *Proc Natl Acad Sci U S A*. 1991;88:737–40.
40. Forrellad MA, Klepp LI, Gioffré A. Sabio y García J, Morbidoni HR, Santangelo M de la P, et al. Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence*. 2013;4:3–66.
41. Pandey AK, Sassetti CM. Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A*. 2008;105:4376–80.
42. Clark LC, Seipke RF, Prieto P, Willemsse J, van Wezel GP, Hutchings MI, et al. Mammalian cell entry genes in *Streptomyces* may provide clues to the evolution of bacterial virulence. *Sci Rep*. 2013;3:1109.
43. Ahmad S, El-Shazly S, Mustafa AS, Al-Attayah R. The six mammalian cell entry proteins (Mce3A-F) encoded by the mce3 operon are expressed during in vitro growth of mycobacterium tuberculosis. *Scand J Immunol*. 2005;62:16–24.
44. Forrellad M, McNeil M. Role of the Mce1 transporter in the lipid homeostasis of mycobacterium tuberculosis. *Tuberc*. 2014;94:170–7.
45. Waksman SA, Henrici AT. The nomenclature and classification of the Actinomycetes. *J Bacteriol* 1943;46:337–41.
46. Migula W. *Pseudomonas aeruginosa* (schrter) mig. Gustav Fischer, Jena. 1900;2
47. Kilpper-Balz R, Schleifer KH. *Streptococcus Suis* sp. nov., nom. Rev. *Int J Syst Bacteriol*. 1987;37:160–2.
48. Fulde M, Willenborg J, Huber C, Hitzmann A, Willms D, Seitz M, et al. The arginine-ornithine antiporter ArcD contributes to biological fitness of *Streptococcus Suis*. *Front Cell Infect Microbiol*. 2014;4:107.
49. Boyce KJ, McLauchlan A, Schreider L, Andrianopoulos A. Intracellular Growth I. Dependent on tyrosine catabolism in the dimorphic fungal pathogen *Penicillium marneffeii*. *PLoS Pathog*. 2015;11:1–30.
50. Greene GH, McGary KL, Rokas A, Slot JC. Ecology drives the distribution of specialized tyrosine metabolism modules in fungi. *Genome Biol Evol*. 2014;6:121–32.
51. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequences (MIGS) specification. *Nat Biotechnol*. 2008;26:541–7.
52. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and Eucarya. *Proc Natl Acad Sci U S A*. 1990;87:4576–9.
53. Phylum GM, XXVI. Actinobacteria phyl. *Nov. Bergey's man. Syst. Bacteriol*. 2012;3:33–4. <https://doi.org/10.1601/nm.5711>
54. Stackebrandt E, Rainey F a., Ward-Rainey NL. Proposal for a new hierarchic classification system, Actinobacteria classis nov. *Int J Syst Bacteriol* 1997;47:479–491. <https://doi.org/10.1601/nm.5712>
55. Zhi XY, Li WJ, Stackebrandt E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int J Syst Evol Microbiol*. 2009;59:589–608.
56. Skerman V, McGowan V, Sneath P. Approved lists of bacterial names. *Int J Syst Bacteriol*. 1980;2:225–420.
57. Chester F. Report of mycologist: bacteriological work. *Delaware Agric Exp Stn Bull*. 1897:59–80.
58. Lehmann K, Neumann R. Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik. First Ed. J.F. Lehmann, München. 1896.
59. RE B, NE G. *Bergey's manual of determinative bacteriology*. Balt: Williams Wilkins Co.; 1974.
60. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25–9.
61. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo AC, Dong X, et al. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res*. 2005;33

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

